Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence

Inaugural-Dissertation

zur Erlangung der Doktorwürde

der

Philosophischen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

vorgelegt

von

Volker Dellwo

aus

Trier

Bonn 2010

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

Zusammensetzung der Prüfungskommission:

1. Vorsitzende: Prof. Dr. Caja Thimm
2. Erster Gutachter: Prof. Dr. Wolfgang Hess
3. Zweiter Gutachter: Prof. Dr. Bernhard Schröder
4. Weiteres Mitglied: Prof. Dr. Bernd Möbius

Tag der mündlichen Prüfung: 15.01.2009

TABLE of CONTENT

Acknowledgements

For the present work I received support and help from numerous friends and colleagues. First of all I would like to thank *Petra Wagner* for giving me the chance to work at Bonn University and for her constant support and collaboration during the project. Special thanks also go to my supervisors and examiners *Wolfgang Hess* and *Bernhard Schröder* for their support and helpful comments on several draft versions of this document.

Next in line is *Patti Adank* who helped me with important discussions, comments on early drafts, and detailed suggestions on statistical procedures. I would also like to thank my colleagues at UCL, *Stuart Rosen* and *Paul Iverson* for helpful discussions of experimental methods and statistical analysis, *Adrian Fourcin* for some ingenious ideas and the collaboration on the voice data, *Jana Dankovičová* for her collaboration with the Czech data (she recorded and segmented this data), *Bronwen Evans* for helpful comments on early drafts of the document, *Mark Huckvale* for numerous programming advice, *Andy Faulkner* for a review of earlier draft versions of parts of the document, and finally *Yi Xu* for insightful discussions about his radically different approach to speech rhythm.

Further thanks go to my colleagues at Bonn University, *Stefan Breuer* and *Christian Weiss* for programming advice on Perl and Praat scripts, *Bianca Aschenberner* and *Ingmar Steiner* for help with segmenting of the BonnTempo corpus.

Many thanks also to my collaborators at Lyon University, *Emmanuel Ferragne* and *Francoise Pellegrino* for carrying out the listening experiment in chapter 9 with French listeners at Lyon University. I also wish to thank *Judith Adrien* for her help on the French translation and on acquiring the French data in Bordeaux, *Stacy Dellwo* and *Franco Ruina*, for translating the reading material into their native languages (English and Italian respectively) and *Eva Orth* for proofreading and comments on early drafts.

Last but not least very special thanks go to *Jan Eaton* for introducing me to some powerful methods of practical time management and for making sure they are applied.

Deutsche Zusammenfassung (Summary in German)

Die vorliegende Untersuchung ist eine experimentalphonetische Studie zum Einfluss der Sprechgeschwindigkeit auf den Sprachrhythmus. Aufgrund auditiver Impressionen wurden Sprachen bezüglich rhythmischer Charakteristika traditionell in Klassen unterteilt, von denen die prominentesten die so genannten akzentzählenden und silbenzählenden Rhythmen beinhalten. Die Nomenklatur motiviert sich aus der Isochronie-Hypothese, die davon ausgeht, dass silbenzählende Sprachen annähernd gleiche silbische Intervalldauern aufweisen, während die Dauer der Intervalle zwischen zwei betonten Silben (Inter-Akzent-Intervall, oder Fuß) in akzentzählenden Sprachen ähnlich lang sind. Sollte die Isochronie-Hypothese auf akustischer Ebene Bestand haben, so ist davon auszugehen, dass der auditive Eindruck der unterschiedlichen Intervall-Isochronie in akzent- und silbenzählenden Sprachen durch entsprechende akustische Korrelate hervorgerufen wird. Wie in der Einleitung zur vorliegenden Arbeit (**Kapitel 1**) jedoch gezeigt wird, konnten in zahlreichen experimentellen Studien in den 1960er, 70er und 80er Jahren solche akustische Korrelate nicht ermittelt werden. Erst als Ende der 90er Jahre der Fokus von Silben und Füßen als rhythmische Grundeinheiten zu konsonantischen und vokalischen Intervallen verschoben wird, melden zwei unabhängig aber fast zeitgleich voneinander durchgeführte Studien positive Befunde akustischer Korrelate für Rhythmusklassen im Sprachsignal (Ramus et al., 1999, Grabe & Low, 2002). Konsonantische Intervalle werden in diesem Zusammenhang als eine Sequenz eines oder mehrerer Konsonanten zwischen zwei Vokalen, vokalische Intervalle als Sequenz eines oder mehrerer Vokale zwischen zwei Konsonanten definiert (unabhängig von Silben-, Wort- oder Satzgrenzen, jedoch nicht unabhängig von Pausen). Besagte Studien gehen auf die Hypothese zurück, dass sprachlicher Rhythmus eine Folgeerscheinung der phonotaktischen Struktur einer Sprache ist. Wahrgenommener Rhythmus wird demnach durch die Variabilität in den Dauern konsonatischer und/oder vokalischer Intervalle hervorgerufen, die wiederum durch die phonotaktische Komplexität der Silben verursacht wird. Ursache für unterschiedliche Rhythmusklassen ist daher die Tatsache, dass Sprachen, die traditionell als silbenzählend eingestuft wurden, eine geringere Silbenkomplexität aufweisen als akzentzählende Sprachen (trotz nun nicht mehr stimmiger

Motiviertheit wurde die Nomenklatur 'akzentzählend' und 'silbenzählend' bis heute beibehalten). Phonotaktische Komplexität, so wird weiter argumentiert, äußert sich auf vokalischer Ebene hauptsächlich im Vorkommen vokalischer Reduktionen (vorhanden in akzentzählenden, nicht vorhanden in silbenzählenden Sprachen) und auf konsonantischer Ebene in der Komplexität konsonantischer Sequenzen, d.h. in der Anzahl konsonantischer Segmente, die in einer Sequenz vorkommen können (niedrig in silbenzählenden, hoch in akzentzählenden Sprachen). Es wird daher vermutet, dass die Dauer vokalischer und konsonantischer Sequenzen in silbenzählenden stärker als in akzentzählenden Sprachen variiert.

Als akustische Korrelate werden in den beiden erwähnten Studien jeweils ein konsonantisches und ein vokalisches Variabilitätsmaß vorgeschlagen. In Ramus et al. (1999) sind dies die prozentuale Dauer, während der ein Sprachsignal vokalisch ist (%V), und die Standardabweichung konsonantischer Intervalle (ΔC). In Grabe und Low (2002) werden konsonantische und vokalische Variabilität mit einem Maß erfasst, das durchschnittliche Unterschiede zwischen zwei unmittelbar hintereinander auftretenden konsonantischen oder vokalischen Intervallen berechnet (Pairwise Variability Index; PVI). Da alle akustischen Rhythmuskorrelate auf Messungen der Dauer vokalischer und konsonatischer Intervalle basieren, ist davon auszugehen, dass diese Intervalldauern a) durch die Sprechgeschwindigkeit beeinflusst werden (schnelle Sprache = kürzere Intervalldauern, langsame Sprache = längere Intervalldauern) und dass b) eine solcher Einfluss nicht unbedingt linear ist.

Die vorliegende Arbeit motiviert sich daher zunächst hauptsächlich aus der Fragestellung, welchen Effekt das Sprechtempo auf die Variabilität konsonantischer und vokalischer Intervalle hat und wie sich eine eventuelle qualitative Abweichung der Dauervariabilität in Abhängigkeit von der Sprechgeschwindigkeit auf die rhythmische Wahrnehmung eines Sprachsignals auswirken kann.

Kapitel 2 macht einen Exkurs in die geschichtliche Entwicklung der wichtigsten Thesen auf dem Gebiet des Sprachrhythmus. Dieses Kapitel liefert relevante Hintergrundinformationen zu Sprachrhythmustheorien, die jedoch nicht essenziell für die Entwicklung des Hauptarguments der Arbeit sind (das Hauptargument wird im wesentlichen in den Kapiteln 1 und 3 bis 10 entwickelt). Die Herkunft und Ursprünge der Isochronie-Hypothese werden anhand von Studien des späten 19. und frühen 20. Jahrhunderts untersucht. Es wird festgestellt, dass die noch bis zum

gegenwärtigen Zeitpunkt häufig getätigte Annahme, dass sprachlicher Rhythmus in Silben und/oder Füßen manifestiert ist, wahrscheinlich eine direkte Übertragung von Erkenntnissen auf dem Gebiet des musikalischen Rhythmus auf sprachliche Zusammenhänge ist. Es wird argumentiert, dass eine solche Übertragung jedoch nur im Zusammenhang mit Studien legitim ist, die sprachlichen Rhythmus in Gedichts- oder Liedform untersuchen. Darüber hinaus werden in dem Kapitel unterschiedliche Motivationen von Studien im Bereich des Sprachrhythmus untersucht.

In **Kapitel 3** werden hauptsächlich methodische Aspekte der vorliegenden experimentellen Studie diskutiert. Kernstück der Studie ist eine Datenbank (BonnTempo Korpus) gelesener Sprache (ca. 90 Silben pro Sprache), die zur Analyse der rhythmischen Korrelate in Abhängigkeit mit Sprechgeschwindigkeit erstellt wurde. Die Hauptuntersuchungsprachen werden in diesem Kapitel festgelegt und ihre Auswahl begründet. Es sind Englisch und Deutsch als Beispiele für Sprachen, die traditionell als akzentzählend klassifiziert werden, Französisch und Italienisch als silbenzählende Beispiele und Tschechisch als Sprache, für die auditive Beobachtungen zu unterschiedlichen Klassifizierungen führten. Eine Analyse der Silbenkomplexität in den ausgewählten Texten der Untersuchungssprachen bestätigt die erwartete typische niedrigere Komplexität in silbenzählenden verglichen mit akzentzählenden Sprachen. Um Einflüsse der Sprechgeschwindigkeit auf die rhythmischen Korrelate besser untersuchen zu können, wurden Sprecher bei den Aufnahmen zu BonnTempo gebeten, ihr Sprechtempo stark zu verlangsamen bzw. zu beschleunigen. Diese Methode führte zur Erfassung von fünf Versionen eines unterschiedlich intentionierten Sprechtempos: normal (ein Sprecher [Genus bezieht sich auf Männer und Frauen] liest den Text in einem für ihn normalen Tempo), langsam (ein Sprecher versucht langsamer als normal zu lesen), sehr langsam (ein Sprecher wird gebeten, sein Sprechtempo wiederum zu verlangsamen), schnell (ein Sprecher versucht schneller als für ihn normal zu lesen), am schnellsten (ein Sprecher liest nach beliebig vielen Durchgängen in dem für ihn schnellst möglichen Tempo). Alle Versionen wurden mit der Software Praat (www.praat.org) bezüglich der Silbendauern sowie der Dauern konsonantischer und vokalischer Intervalle etikettiert. Zur Analyse der Intervall-Dauern wurden mit Praat umfangreiche Datenverarbeitungsskripte erstellt (BonnTempo-tools, siehe Kapitel 3). Die auf vokalischer oder konsonantischer Variabilität beruhenden rhythmischen Korrelate

werden in diesem Zusammenhang eingehend diskutiert.

Erstes Ziel in **Kapitel 4** ist ein Vergleich der Daten in BonnTempo mit bereits publizierten Ergebnissen zu Rhythmusmaßen. Bei den Maßen handelt es sich um die in Ramus et al. 1999 vorgeschlagenen %V (prozentualer vokalischer Anteil im Sprachsignal) und ΔC (Standardabweichung der Dauer konsonantischer Intervalle) sowie um den von Grabe und Low (2002) entwickelten Pairwise Variability Index (PVI). Letzteres Maß berechnet die durchschnittliche Differenz konsekutiver konsonantischer oder vokalischer intervalle. Es wird zwischen einem rohen konsonantischen rPVI (von Engl. 'raw' für 'roh') und einem sprechgeschwindigkeitsnormalisierten nPVI (von English 'normalized') unterschieden (vgl. Grabe und Low, 2002).
Es wird zunächst festgestellt, dass alle Intervalldauern eine stark positive Schiefverteilung sowie stark positive Kurtosis aufweisen. Diese Verteilung führt unter anderem zu für die Datenpopulation nicht repräsentativen Ergebnissen beim Berechnen von Mittelwerten und Standardabweichungen und ist daher besonders kritisch im Fall von ΔC. Eine logarithmische Transformation der Daten führt zu einer annähernd perfekten Normalverteilung der Intervalldauern, weshalb ΔC in jedem Zusammenhang arithmetisch und geometrisch berechnet wird. In Kapitel 4 wird demonstriert, dass Ergebnisse aus anderen Studien perfekt repliziert werden können. Absolut können identische Werte zwischen entsprechender Intervallvariabilität in BonnTempo und vergleichbaren Studien ermittelt werden. Darüber hinaus zeigen statistische Signifikanztests (ANOVA), dass Rhythmusklassen von allen rhythmischen Korrelaten unterschieden werden. Nichtsdestotrotz können teilweise signifikante Unterschiede zwischen Sprachen innerhalb einer Rhythmusklasse ermittelt werden (z.B. Englisch und Deutsch für ΔC und rPVI) und keine signifikanten Unterschiede zwischen Sprachen unterschiedlicher Rhythmusklassen (z.B. zwischen Englisch und Italienisch bezüglich nPVI). Aus diesem Grund wird geschlussfolgert, dass die rhythmischen Korrelate dazu geeignet sind, sprachtypische prosodische Unterschiede zu ermitteln, jedoch nur bedingt traditionelle Rhythmusklassen voneinander trennen. Für das Tschechische wird ermittelt, dass es bezüglich unterschiedlicher Maße in unterschiedliche Rhythmusklassen eingeteilt wird. Dies zeigt an, dass die rhythmischen Korrelate Sprachen nicht einheitlich einer bestimmten Klasse zuordnen können. Demnach scheinen verschiedene Korrelate auf

unterschiedliche Aspekte konsonantischer und vokalischer Intervalldauervariabilität unterschiedlich zu reagieren.

In **Kapitel 5** werden die Sprechgeschwindigkeitscharakteristika der Versuchssprachen näher untersucht. Zu diesem Zweck wird zunächst ein Sprechgeschwindigkeitsmaß ermittelt. Mehrere Anzeichen sprechen dafür, dass die Rate der Summe konsonantischer und vokalischer Intervalle pro Sekunde (kv-Rate) ein adäquateres Maß als die traditionell verwendete Silbenrate darstellt. Bezüglich der kv-Rate wird ermittelt, dass diese deutlich mit dem intendierten Sprechtempo der Versuchssprecher korreliert. Es wird weiterhin ermittelt, dass sich die Versuchssprachen (abgesehen von Französisch und Italienisch) weitgehend in ihren spezifischen kv-Raten unterscheiden. Der wichtigste Befund in Kapitel 5 ist die Tatsache, dass Sprechgeschwindigkeit einen ebenso verlässlichen Parameter für die Klassifizierung von Sprachen in Rhythmusklassen darstellt, wie die meisten Korrelate des Sprachrhythmus aus Kapitel 4. Daraus wird gefolgert, dass Sprechgeschwindigkeit auch auditiv einen wichtigen Beitrag zur Rhythmusklassenunterscheidung leisten könnte. Diese Annahme wird in Kapitel 8 und 9 weiter untersucht.

Kapitel 6 analysiert den Einfluss der Sprechgeschwindigkeit auf die Korrelate des Rhythmus. Lineare Regressionsanalysen zwischen den Parametern Sprechgeschwindigkeit und den unterschiedlischen Sprachrhythmus Korrelaten zeigen, dass die beiden konsonantischen Maße ΔC und rPVI stark mit Sprechgeschwindigkeit korrelieren. Aus diesem Grund werden für die konsonantischen Maße Verfahren entwickelt, die den Einfluss der Sprechgeschwindigkeit normalisieren. Für ΔC wird festgestellt, dass die logarithmische Transformation der Dauerwerte konsonantischer Intervalle ausreichend ist, um Sprechgeschwindikeitsartefakte auszuschließen. Der konsonantische rPVI wird bezüglich der Sprechgeschwindigkeit durch in der Literatur für andere Maße vorgeschlagene Verfahren (nPVI, vgl. Grabe & Low, 2002) erfolgreich normalisiert. Eine Untersuchung des Einflusses der Sprechgeschwindigkeit auf die Rhythmus Korrelate in den einzelnen Untersuchungssprachen (für die konsonantischen Korrelate vor und nach der Sprechgeschwindigkeitsnormalisierung) ergibt, dass %V und der auf

Sprechgeschwindigkeit normalisierte rPVI die Sprachen am robustesten in Rhythmusklassen unterteilt.

Kapitel 7 ist ein kurzer Exkurs in andere Einflussfaktoren. Am Beispiel des %V Maßes, das im Vorhergehenden als sprechgeschwindigkeitsunabhängig befunden wurde, wird festgestellt, dass Korrelate des Rhythmus zwischen unterschiedlichen Satzphrasen innerhalb einer Sprache stark und statistisch signifikant variieren. Diese Unterschiede sind so stark, dass manche Phrasen in bestimmten Fällen deutlich in durchschnittliche Bereiche anderer Rhythmusklassen/Sprachen fallen. Auch in Abhängigkeit vom Sprecher kann eine Variabilität von %V ermittelt werden. Dies ist jedoch nicht in jeder Sprache der Fall. Die Beobachtung in Kapitel 7 sind von besonderer Bedeutung für die Methode des Perzeptionsexperiments in Kapitel 8, wo eine Auswahl an Sätzen mit maximal hoher und geringer konsonantischer und vokalischer Intervallvariabilität benötigt wird, um unter anderem zu testen, ob Veränderungen in der Intervallvariabilität auch innerhalb einer Sprache zu Abweichungen des wahrgenommenen Rhythmus führen können.

In **Kapitel 8** wird in einem Perzeptionsexperiment getestet, ob Hörer eher die für sprachliche Äußerungen typische Rate oder Variabilität konsonantischer oder vokalischer Intervalle verwenden, um rhythmische Aussagen über eine sprachliche Äußerung zu treffen. In diesem Experiment werden daher Sätze des BonnTempo Korpus so delexikalisiert, dass nur noch Dauerinformation konsonantischer und vokalischer Intervalle verbleiben (vokalische Intervalle werden in komplex periodische Signale mit konstanter Grundperiode, konsonantische Intervalle in weißes Rauschen umgewandelt). Versuchspersonen werden dann damit beauftragt, die in den delexikalisierten Stimuli vorkommenden Tonintervalle bezüglich ihrer Regularität zu beurteilen. Die Interpretation von Regularität wird dabei dem Hörer überlassen. Es wird angenommen, dass Stimuli, deren Intervalle weniger Dauervariabilität enthalten, als weniger regulär beurteilt werden. Die Ergebnisse des Experiments bestätigen diese Vermutung jedoch nicht. Alle Rhythmuskorrelate korrelieren, wenn überhaupt, dann nur unmerklich mit den Hörereinschätzungen (maximaler r^2 Wert 0.2). Im Gegensatz dazu weist Sprechgeschwindigkeit jedoch eine überaus deutliche Übereinstimmung mit den Hörereinschätzungen auf. Aus den Ergebnissen wird geschlussfolgert, dass wenn Hörer die Möglichkeit besitzen, für

Sprachrhythmusklassen typische rhythmische Variabilität anhand der Korrelate Sprechgeschwindigkeit und Intervallvariabiliät zu beurteilen, die Präferenz deutlich dem Parameter Sprechgeschwindigkeit zugemessen wird.

Primäres Ziel von **Kapitel 9** ist eine weitere Untersuchung der perzeptiven Relevanz der Intervallvariabilität im Vergleich mit Sprechgeschwindigkeit. Das dazu durchgeführte Perzeptionsexperiment motiviert sich aus Beobachtungen auf dem Gebiet des Zweitsprachenerwerbs, die im Laufe der experimentellen Untersuchungen auffällig wurden. Für deutsche Muttersprachler, die die englischen und französischen Versionen des BonnTempo Textes lesen, zeigte sich, dass sich die nicht-muttersprachlichen von den muttersprachlichen Produktionen nur in einem statistisch signifikant höheren %V unterscheiden. Die Sprechgeschwindigkeit des Englischen wird von den Deutschen muttersprachlich wiedergegeben, das Französische wird hingegen deutlich langsamer produziert, als dies für französische Muttersprachler üblich ist. Die Ergebnisse der akustischen Studie lassen methodologisch eine Perzeptionsstudie zu, die keine delexikalisierten Sprachstimuli verlangt. Da der sprachliche Rhythmus gerade beim Fremdsprachenerwerb als grundlegendes prosodisches Merkmal für Aussprachedefizite angesehen wird, sollten Sprecher, die rhythmisch messbar näher mit den Mutterprachlern übereinstimmen, auch als die Sprecher mit der besseren Aussprache wahrgenommen werden. Diese Hypothese wird in einem Experiment getestet, in dem muttersprachliche Hörer des Englischen und Französischen die entsprechenden fremdsprachlichen Produktionen der deutschen Sprecher auf einer Skala von 1 bis 11 bezüglich ihrer Aussprache bewerten. Die Ergebnisse zeigen eindeutig, dass keines der rhythmischen Korrelate mit den Hörereinschätzungen korreliert. Für das Französische hingegen, in dem die Nicht-Muttersprachler deutliche Variabilität bezüglich des muttersprachlichen Sprechtempos aufweisen, korreliert die Sprechgeschwindigkeit stark mit den Hörereinschätzungen. Der Befund wird als weiteres Indiz dafür betrachtet, dass Sprechgeschwindigkeit ein bevorzugtes Korrelat für perzeptive rhythmische Variabilität im Sprachsignal ist.

Kapitel 10 fasst die in den experimentellen Teilen ermittelten Ergebnisse im Hinblick mit den in der Einleitung aufgestellten Untersuchungszielen zusammen und führt eine abschließende Diskussion durch. Aus den akustischen Messungen und aus

den Perzeptionsexperimenten geht hervor, dass Sprechgeschwindigkeit ein akustisches Korrelat für Sprachrhythmus ist, das bislang noch keine Beachtung gefunden hat. Vorausgehende Perzpetionsexperimente haben bislang nur gezeigt, dass Hörer Sprachen aufgrund ihrer unterschiedlichen konsonantischen und vokalischen Intervallvariabilität unterscheiden können, wenn Sprechgeschwindigkeit nicht variiert und wenn Hörer auf die entsprechende Variabilität trainiert werden (Ramus et al. 1999). Wenn Hörer jedoch die Wahl zwischen Intervallrate (Sprechgeschwindigkeit) und Intervallvariabilität haben, ist die Intervallrate der perzeptiv stärker hervortretende Parameter. Es wird daher argumentiert, dass Sprechgeschwindigkeit für die Perzeption von Sprachrhythmus eine besondere Rolle spielt und dass ein Großteil des Eindrucks, dass silbenzählende Sprachen einen Maschinengewehr ähnlichen Rhythmus besitzen (Lloyd James, 1929, Pike, 1946, Aberchrombie, 1967) an der Tatsache liegt, dass sprechsprachliche Sequenzen wie konsonantische und vokalische Intervalle in silbenzählenden Sprachen schneller produziert werden als in akzentzählenden Sprachen.

1. Introduction

1.1. The general idea

It has often been argued that the languages of the world can be classified into distinct rhythmic types of which the two most prominent are the stress-timed and syllable-timed rhythm classes (James, 1929, Pike, 1945, and Abercrombie, 1967, Bolinger, 1981, Roach, 1982, Ramus et al., 1999, Grabe & Low, 2002). Behavioral experiments have shown that adult human listeners (Ramus & Mehler, 1999), as well as newborns (Nazzi et al., 1998, Ramus, 2002), monkeys (Ramus et al., 2000, Rincoff et al., 2005), and rats (Toro et al., 2003) can distinguish between languages from different rhythmic classes.

What are the acoustic cues that enable listeners to distinguish between stress- and syllable-timed languages? The rhythm of syllable-timed languages has metaphorically been likened to a 'machine gun' sound, stress-timed languages to a 'Morse-code' signal (see discussion in chapter 2 and James, 1929). These metaphors have been motivated by the apparent percept of rhythmic regularity and irregularity in syllable- and stress-timed languages respectively. Contemporary theories of speech rhythm imply that this regularity percept is mainly triggered by the variability of consonantal- (C) and vocalic- (V) intervals in connected speech (see definition below, 1.2., and also in Ramus et al., 1999, Grabe & Low, 2002). In terms of acoustic measurements, Ramus et al. (1999) demonstrated that the standard deviation of C-intervals (ΔC) and the percentage over which speech is vocalic (%V) correlate best with listeners' perception of rhythm class. Typically %V is higher and ΔC lower in syllable- than in stress-timed languages (Ramus et al., 1999, Dellwo and Wagner, 2003), which reflects that C-interval durations in syllable-timing are relatively shorter and durationally more equal. Also V-intervals have been demonstrated to be more regular in duration in syllable-timed languages, e.g. the average difference between consecutive V-intervals in connected speech is smaller (vocalic nPVI; Grabe & Low, 2002) as is the rate-normalized standard deviation of V-interval durations (VarcoV; White and Mattys, in press). An in depth explanation of the

parameters will be given below.

In summary, numerous studies agree that C- and V-interval durations are less variable in syllable- than in stress-timed languages. It therefore seems plausible that the 'machine gun' and 'Morse code' metaphors were evoked by such characteristics. However, these metaphors contain a hitherto neglected detail, which is the main motivation of the present study. Machine gun and Morse code signals are not only distinguished by variability in their respective interval durations but also by a second parameter: *rate*.

Evidence that the perception of interval variability can be dependent on their rate goes as far back as Weber's law which states that the ratio between the just noticeable difference (jnd) and the magnitude of a physical event is constant (which means: the relative jnd is constant). Psychoacoustic research (see Friberg & Sundberg, 1995, for a literature review) has repeatedly demonstrated that this is true for the perception of jnds in truly isochronous acoustic events (at least between certain ranges of rates). Given this evidence it is conceivable that irregularities in rhythmic intervals become reduced perceptually with higher rates. For speech this would mean that the rate of rhythmic units (i.e. C- or V-intervals) could influence how listeners perceive their durational regularity, hence how they perceive speech rhythm.

Do units of speech rhythm vary in speech? So far, the rate of C- or V-intervals has not received much attention in speech rhythm research. In fact, all of the studies (Nazzi et al., 1998, Ramus et al., 1999, Ramus et al., 2000, Ramus, 2002, Toro et al., 2003, Rincoff et al., 2005) that have found behavioral effects of speech rhythm in both humans and non-humans use stimuli selected from a corpus by Nazzi et al. (1998) in which sentences were of roughly equal number of syllables and durations across all languages under investigation. This, of course, has the effect of reducing speech rate variability within and between languages. By controlling for speech rate in this way, however, these studies may have overlooked that languages of different rhythmic class can probably be distinguished in the acoustic domain on the basis of rate alone. The rationale for C- or V-interval durations being less variable in syllable-timed languages is that these languages typically have phonologically less complex syllable structures (Bolinger, 1981, Roach, 1982, Dauer, 1983, 1987, Dankovičová

and Dellwo, in press, Ramus et al., 1999, Grabe & Low, 2002). Subsequently it is possible that mean C- or V-interval durations vary between stress- and syllable-timed languages. Listeners could thus use rate information to distinguish between languages of different rhythmic class. Even more, if, as hypothesized above, the rate of intervals should have an effect on listeners' perception of interval regularity, it is possible that rate differences between languages of different rhythmic classes are used to make judgments about rhythmic differences (e.g. in terms of whether a speech signal sounds more or less regular). The general aims of the present work are therefore:

(a) to analyze how speech rate varies naturally between stress- and syllable-timed languages (chapter 5).

(b) to analyze how correlates of speech rhythm based on durational characteristics of C- and V-intervals are affected by rate (chapter 6).

(c) to find procedures that normalize for rate artifacts in correlates of speech rhythm (chapter 6).

(b) to test the influence of naturally occurring rate variability between languages of different rhythm classes on the perception of regularity in C- and V-interval durations (chapters 8 and 9).

1.2. A new unit for measuring speech rhythm: consonantal & vocalic intervals

Consonantal and vocalic intervals (henceforth: C- and V-intervals) have already been mentioned above. Both Ramus et al. (1999) and Grabe & Low (2002) used these intervals as fundamental units for rhythmic measurements in speech. Ramus et al. argue that these units are probably most salient for our percept of rhythm. For this reason C- and V- intervals are, next to syllables, the most central and most often used units of speech in the present work. In the following they will be defined accurately:

1.2.1. A definition of consonantal intervals

A consonantal interval (C-interval) is an interval in speech consisting of one or more consonants preceded and followed by a vowel (or by a pause). The bi-syllabic utterance 'Christmas', for example, when produced in isolation, consists of three C-

intervals. In a typical standard southern British English realization of the utterance, the first C-interval would consists of two consonantal segments (/k/ and /r/) which is preceded by the initial pause and followed by the first vowel (/ɪ/). The second C-interval consists again of two consonantal segments (/s/ and /m/) and is preceded by the first vowel (/ɪ/) and followed by the second vowel (/ə/). This vowel also precedes the third and last C-interval of the utterance which consists of a single consonantal segment (/s/).

C-intervals can be structurally simple, which means that they consist of only one consonantal segment (e.g. the third C-interval in 'Christmas') or they can be complex, which means that they consist of at least two consonantal segments (typically more).

As can be seen in the example above, C-intervals can stretch across syllable boundaries, i.e. the syllable boundary between the /s/ and the /m/ in the second C-interval of 'Christmas' does not separate the interval. In utterances consisting of more than one word C-intervals also stretch across any grammatical boundary such as word or sentence boundaries, as long as these boundaries are not marked by a pause. So the utterance 'Christmas cake' would turn the last C-interval in 'Christmas' from a simple (single /s/ interval) to a complex C-interval consisting of two consonantal segments (/s/ and /k/). This consonantal segment is then followed by the diphthong in 'cake'.

1.2.2. A definition of vocalic intervals

A vocalic interval (V-interval) is an interval in speech consisting of one or more vowels (or vocalic segments like diphthongs, triphthongs, etc.) preceded and followed by a consonant (or by a pause). The above single word utterance 'Christmas' contains two V-intervals, the first one, /ɪ/, preceded by /r/ and followed by /s/ and the second one, /ə/, preceded and followed by the consonants /m/ and /s/ respectively.

Vocalic segments typically form the syllable nucleus and are typically preceded and followed by a consonant. As such V-intervals typically consist of one vowel or diphthong only. It is, however, possible that two vowels clash together at word boundaries in multi word utterances in which one word finishes with a vowel and the following word starts with another vowel (e.g. 'filthy artist'). In such case the V-

interval would stretch across the word boundary (here: from the onset of the vowel preceded by /θ/ in 'filthy' to the offset of the first vowel in 'artist').

1.2.3. A note on abbreviations

In many cases in the present work it makes sense to talk about the combined number of C- or V-intervals; this is to say: the sum of the number of C- and V-intervals a phrase consists of. Such a point of view will be of particular importance for the calculation of speech rate. It will be demonstrated that under numerous circumstances the total number of C- and V-intervals per second (see e.g. 3.3.2) is more applicable than the more commonly used number of syllables per second (e.g. in delexicalized speech in which only cues to C- and V-interval durations remain). This rate measurement will then be referred to as CV-rate.

As introduced above, vocalic and consonantal intervals are abbreviated with capital V and C respectively. In rare cases in the present work the same letters will be used to refer to vocalic and consonantal segments, however, to avoid confusion they have been set to lower cases; thus 'v' will be used for a vocalic segment and 'c' for a consonantal segment. A C-interval can thus, for example, consist of one c-segment (c) or of three c-segments (ccc).

1.3. The structure of the present work

Five languages were selected to analyze the effects of speech rate on rhythmic correlates: two languages that have traditionally been classified as stress-timed, English (see e.g. Lloyd James, 1929, Pike, 1946, Abercrombie, 1967, Grabe & Low, 2002, Ramus et al., 1999) and German (see Kohler, 1982 and 1983), another two languages that have been classified as syllable-timed, French (see e.g. Lloyd James, 1929, Abercrombie, 1967, Grabe & Low, 2002, Ramus et al., 1999) and Italian (see Grabe & Low, 2002, Barry et al., 2003). Czech was further selected as a language for which expert listeners diverge in auditory classification (Dankovičová & Dellwo, 2007).

In a first step of the experimental phonetic work it will be investigated (chapter 4) whether general findings of durational C- and V-interval variability for the languages under investigation can be replicated using a large database for the study of interval

variability in speech (BonnTempo corpus, see chapter 3). After this, general aspects of speech rate variability within and between languages will be analyzed in order to assess speech rate influences on rhythmic variability measures (chapter 5). This is followed by an analysis of how C- and V-interval variability is affected by changes in speech rate in general and in individual languages (chapter 6). Chapter 7 is a small excursion studying other influences on rhythm measures that turned out to be very prominent during the analysis of the data in BonnTempo. In chapters 8 and 9 the findings for rhythm and speech rate interaction will be analyzed in the perceptual domain. In these chapters, rate and C- and V-interval variability will be assessed by the listeners in order to investigate which of the two cues is the more salient one for listeners' perception of rhythmical differences. In chapter 10 the main findings of the work will be summarized and the relevance of speech rate in speech rhythm studies will be discussed in respect to psychoacoustic findings on isochronous rhythmic events (Friberg & Sundberg, 1995)

A small excursion into the history of rhythm measures is given in chapter 2. For this chapter, early work on rhythm and speech rhythm from C19th to C20th is reviewed in order to find why the syllable and the foot was, most probably wrongly, assumed to be the unit of rhythm in speech. While all other chapters build up on each other in terms of argumentation, chapter 2 can be regarded as substantial background information about rhythm classes but is not essential to understand the overall argument of the work.

For the remaining part of the introduction the concept of the rhythm class hypothesis will be explained in more detail (1.3). After that (1.4), measures of speech rhythm based on C- and V-interval variability will be discussed and in (1.5) previous work of speech rate on such rhythm measures will be reviewed. The introduction will finish (1.6) with a brief overview of motivations for the study of speech rhythm. These aspects are further discussed in chapter 2.

1.4. Speech rhythm and the rhythm class hypothesis

> "It is not easy to define rhythm, which is almost axiomatic."
> Lloyd James (no year) p. 11

Definitions of rhythm are numerous and so are definitions of speech rhythm (see Strangert, 1985, and Eriksson, 1991 for discussions of selected literature). Eriksson (1991) points out that all definitions of rhythm "are built on the regular occurrence of some event or events and some kind of structuring (grouping) of these events" (p.6). Eriksson further distinguishes two types of rhythm definitions, one emphasizing more on 'temporal regularity' and the other on 'temporal structuring' of rhythmical events. This means that rhythmical events can either be formed by time-aligning their underlying units in different ways or by giving these units different structures (e.g. in the frequency or amplitude domain).

Definitions of 'speech rhythm' are also based on the principle of 'regular occurrence' and 'structuring'. Strangert (1985) points out that definitions may vary from rather general views about the organization of stressed syllables (referring to Lehiste, 1979) to definitions in which speech rhythm is a property of structural aspects of its underlying language (referring to Dauer, 1983). So on the one hand rhythm can be studied in independence of a language system and on the other hand rhythm is inherent of and deeply connected with a language system. The latter statement divided the research community into believers and non-believers ever since it was proposed (first in early metrical research on poetry, see Sonnenschein, 1925, Thomson, 1904, 1916, 1923, and then for prose and spontaneous speech, see e.g. Lloyd James, 1929, Pike 1946 and Abercrombie, 1967). In this context Lloyd James, Pike, and Abercrombie proposed the rhythm class hypothesis according to which all languages of the world can be classified on the basis of the temporal rhythmical organization of their syllables. The two most prominent rhythm classes that have been studied by far the most in rhythm class research are the stress-timed and the syllable-timed rhythm class. The traditional view is that stress-timed languages are languages in which stressed syllables occur at durationally regular intervals (henceforth: isochronous intervals), irrespective of the number of syllables they contain. In contrast to this syllable-timed languages are characterized by syllables of approximately equal durations (isochronous syllables).

Figure 1-1 shows an idealized organization of syllable timing for syllable- (above) and stress-timed languages (below). Each example shows a hypothetical speech event on two tiers consisting of 11 syllables (upper tier: blue and red rectangles) and three feet (lower tier: black brackets) each. Syllables are illustrated as rectangles while red rectangles stand for stressed and blue rectangles for unstressed syllables. A foot is illustrated by a bracket with the duration from the onset of a stressed syllable (red) to the onset of the next stressed syllable or the coda of the phrase final unstressed syllable. The figure illustrates the original idea of stress- and syllable-timing in which syllables show quasi equal syllable but variable foot durations whereas stress-timed languages are characterised by equal foot durations but variable syllable durations.

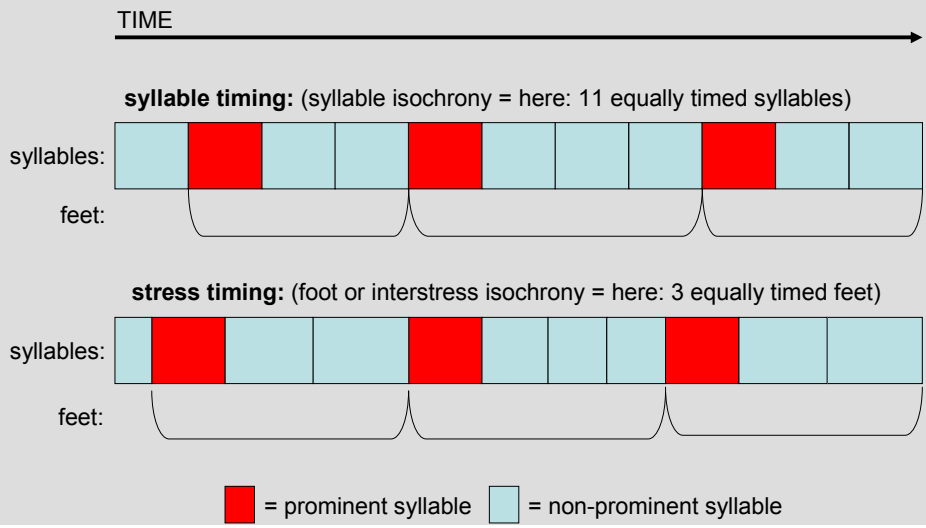

Figure 1-1: idealized timing of syllable- (above) and stress-timed languages (below).

Abercrombie (1967) tried to relate speech rhythm to the speech production level and claimed that stress- and syllable timing is a result of specific pneumonic processes which he defined as chest and stress-pulses. In his view chest pulses are puffs of air that are responsible for the production of a syllable and stress-pulses are reinforced chest-pulses leading to a production of syllables with a higher prominence (usually stressed syllables). Abercrombie's view was that languages differ on the timing of stress- and chest-pulses in that syllable-timed languages use equal inter-chest-pulse

and stressed-timed languages equal inter-stress-pulse intervals. With this view Abercrombie implicitly proposed that rhythm is independent of any phonological or phonetic properties but is solely property of the speech production level and there a result of a different use of the respiratory system. Ladefoged (1967) however demonstrated in a series of experiments measuring respiration patterns in speech that Abercrombie's views were incorrect.

A third rhythm class, mora-timed rhythm, was further proposed (Ladefoged, 1975). It is the rhythm of languages like Japanese which possesses the mora as a basic unit of speech. Mora-timing has often been considered as a sub-class of syllable-timed rhythm and is not part of the current work.

Contemporary views on speech no longer follow Eriksson's (1991) principle of 'regular occurrence of events'. Rhythm research has moved away from the isochrony concept to definitions in which speech rhythm is a result of the phonotactic arrangement of speech segments. In these views, rhythm is no longer a production based result of an individual timing target that the speaker wishes to reach (e.g. putting all stressed syllables equidistantly in stressed-timed languages) but simply the result of structural aspects of the language system. Xu & Wang (forthcoming) claim that this view is "no longer about rhythm but about why languages sound different in their timing characteristics." (p. 2). This view is inherent to a line of research that monitors the durational variability of consonantal and vocalic intervals in the speech signal which is the basis for a number of currently widely applied ways to measure rhythm in the speech signal (see Ramus et al., 1999, Grabe & Low, 2002, Barry et al., 2003, Dellwo, 2006). These rhythm measures stand at the centre of the current study and will be discussed further in more detail in the next section.

1.5. Speech Rhythm Class Measurements

1.5.1. Search for isochrony

The search for acoustic correlates of linguistic rhythm started in the late 1960s and had its first peak during the 70s and 80s (search for evidence for the isochrony-hypothesis) and its second peak in the present time since the late 90s (search for the effect of phonotactic structure of the language on perceived speech rhythm).

However, serious experimental research on speech rhythm was conducted long time before these phases. Wallin (1901) used a phonograph to play back short sections of speech and had subjects tapping the underlying rhythm with a telegraph key. The differences in length between the resulting dots on a paper strip were then used to calculate the durations between the perceived rhythmical landmarks (represented by a telegraph key stroke). Wallin's findings delivered strong support for the isochrony of stressed syllables in English measured on prose speech since he could demonstrate that equal intervals of time in prose may be filled with unequal numbers of unaccented syllables. His experimental equipment, however, was not very precise and what he measured was perceived rhythm rather than objective acoustic isochrony in the signal (which does not devaluate the study). With similar experiments on poetic speech Miner (1903) found that "an increase in the number of elements composing a group in a rhythmical series does not proportionally increase the apparent length of the groups" (p. 3), which is further support for isochrony of feet in English. On the same page the author writes: "Rhythm does not depend upon equality of successive time-intervals. Only an approximate equality is necessary." It needs to be pointed out here that this research has been performed long time before the rhythm class hypothesis was actually established e.g. by Pike (1946) and Abercrombie (1967).

From the time Lloyd James (no year, 1929, 1938), Pike (1946) and Abercrombie (1967) proposed their views on isochrony, speech researchers tried to verify the perceptual impression of different rhythm classes. However, from the 1970s to the late 1980s it has been demonstrated repeatedly and exhaustively that durationally isochronous (or at least quasi-isochronous) inter-stress intervals are not observable on an acoustic level in stress-timed languages nor can isochronous syllable durations be measured in syllable timed languages. Not even a proportionately higher amount of syllable isochrony could be found in syllable- compared to stress-timed languages, nor a proportionately higher amount of foot isochrony in stress-timed languages (see for example Roach, 1982). The experimental studies on this topic have been discussed and reviewed widely and in great detail (see Eriksson, 1991, for an in-depth review but also Bröggelwirth, 2003, Grabe & Low, 2002, or Laver, 1994).

1.5.2. Syllable complexity as the driving force in speech rhythm

In a rather recent promising attempt to find acoustic evidence of speech rhythm classes the focus was shifted from syllable and inter-stress intervals to consonantal- and vocalic-intervals (C- and V-intervals) as units of speech rhythm (see 1.2 for an in depth definition; also compare Ramus et al., 1999, henceforth: Ramus, Grabe & Low, 2002, henceforth: Grabe & Low). Ramus and Grabe & Low argue that speech rhythm classes can be distinguished on an acoustic level by monitoring the variability of C- and V-intervals. The rationale behind these measures is that a language's rhythm is reflected in its phonotactic structure. This argument goes back to earlier studies by Dasher and Bolinger (1982), Roach (1982), and Dauer (1983) who all argue that stress-timed languages contain a wide spectrum from structurally simple to complex syllables which causes the overall syllable duration to vary. Also these languages allow vowel reduction in the vocalic part of the syllables which in turn leads to a durationally variable syllable nucleus.

Ramus argues that because of their more complex syllable structure, languages which have traditionally been classified as stress-timed contain structurally more complex C-intervals than languages being classified as syllable-timed. On an acoustic level this should be reflected in the variability of C-interval durations. Languages containing a higher level of consonantal complexity should show this in a higher variability of C-interval durations. On a vocalic level the complexity of vocalic clusters is mainly characterised by whether languages allow vowel reduction which is said to be a feature of languages traditionally classified as stress-timed (in order to reduce the non-stressed syllables in inter-stress position) whereas vowel reduction is rather atypical for syllable-timed languages. Since vocalic reductions are characterised in the time domain by shorter durations, languages that require a variety of reduced and non-reduced vowels should reflect this on an acoustic level by highly variable V-interval durations. Based on these assumptions Ramus conducted an experiment in which they measured the standard deviation of V-intervals (ΔV) and of C-intervals (ΔC) as acoustic measures reflecting the overall durational variability of these intervals. Additionally they measured the percentage of time over which speech is vocalic (%V). The rationale to this measure is twofold: (a) languages allowing vowel reductions contain a greater number of small duration vocalic

intervals and thus the overall proportion of time during which speech is vocalic should be smaller. (b) Since C-intervals in stress-timed languages are typically more complex than in syllable-timed languages the proportional amount of time spent on C-intervals will be higher in stress-timed languages. To the present day it is unclear whether and if yes, to what extent, both factors contribute to %V.

After having tried various combinations of these measures, Ramus et al. (1999) found that the intuitive stress-timed/syllable-timed distinction is best reflected by the acoustic measures ΔC and %V. Along these dimensions, stress-timed languages show a high ΔC (reflecting high C-interval variability) and low %V (reflecting high V-interval variability), and syllable-timed languages show a low ΔC and high %V (see left graph in Figure 1-2).

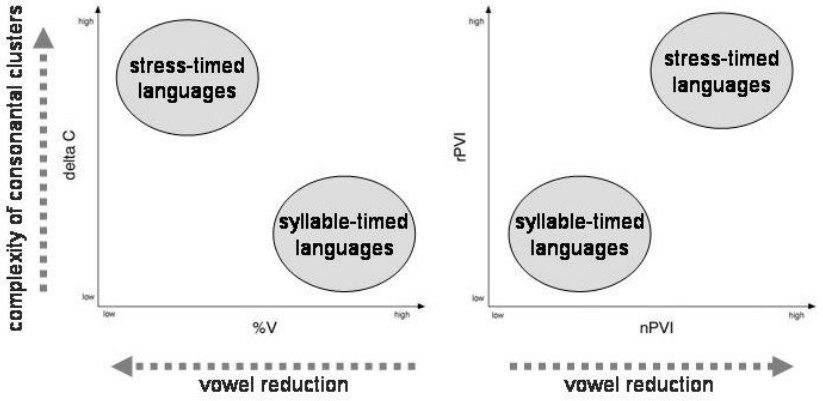

Figure 1-2: Stylized findings of Ramus et al. (1999), left graph, and Grabe & Low (2002), right graph. The left graph shows ΔC as a function of %V, the right graph rPVI as a function of nPVI. The dotted arrows indicate assumed correlations between phonological properties of the language and the C- and V-interval variability as expressed by the parameters on the x- and y-axes of the graphs.

With a similar argumentation Grabe & Low (2002) introduced a measure called the Pairwise Variability Index (PVI) which reflects the variability of C- or V-interval durations by pair wise comparison of consecutive C- or V-interval durations respectively. For vocalic interval variability the authors argue that a normalization process for varying speech rate is necessary for their measure (normalized Pairwise Variability index, henceforth: nPVI) while for consonantal intervals the non-normalized measure (raw Pairwise variability index, henceforth: rPVI) is more

applicable (see Grabe & Low, 2002, for the rationale). A more detailed explanation and the formula for both raw and normalized PVI can be found in the methodology section of this work (chapter 3). The rate normalization procedure will be thoroughly reviewed and experimentally tested in later chapters (3, 4, 6, and 7). In respect to the rhythm class concept, lower C- and V-interval variability, like in syllable-timed languages, is reflected by both a lower rPVI and nPVI, higher interval variability, like in stress-timed languages, by higher rPVI and nPVI variability. The results obtainable for these measures are stylized in Figure 1-2 (right graph) and show that stress-timed languages generally reflect a higher variability for both the nPVI and rPVI than syllable-timed languages.

1.5.3. The perception of durational C- and V-interval variability

On a perceptual level Ramus et al. (1999) tested a group of 20 listeners in a classification task on a set of 26 delexicalized sentences from English and Japanese. The delexicalization method was the same as in Ramus & Mehler (1999) in which vocalic intervals were turned into /a/-vowels (with constant f0) and consonantal intervals into /s/-sounds. This resulted in a reduction of all speech information to durational cues representing C- or V-interval variability. Ramus et al. (1999) found that listeners were well able to distinguish between the two languages. Furthermore in the case of English, listeners' performance showed a negative correlation of identification scores as a function of %V (see Figure 1-3). In a second step of this experiment it was shown that listeners could generally distinguish well between languages of different but not of the same rhythmic class. All results were taken as evidence for the psychological reality of %V, and it was argued that listeners use this value to distinguish between languages. Ramus et al. (1999) omit to evaluate ΔC perceptively because it showed more variability in the acoustic measurements compared to %V. However, concerning the mental reality of ΔC the authors assume a similar situation to be the case. Grabe & Low (2002) did not carry out a perceptual evaluation of their measures. It remains unclear whether nPVI and rPVI are perceptually salient.

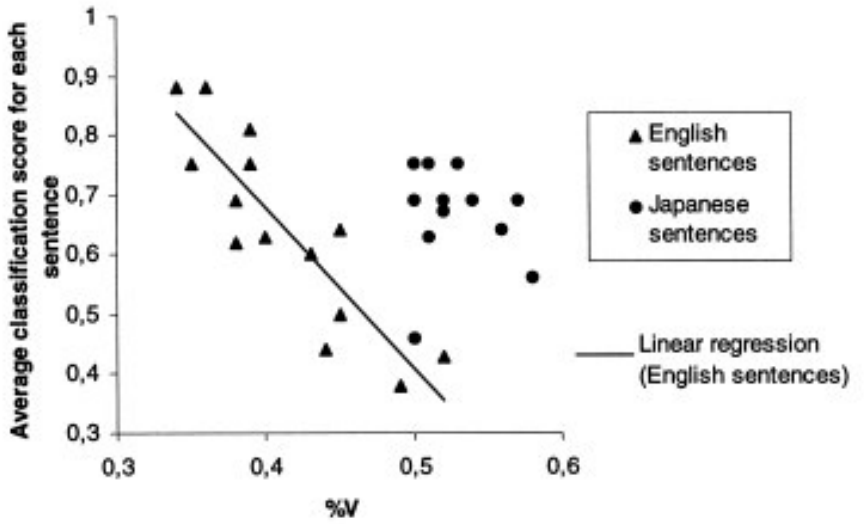

Figure 1-3: Perceptual findings of average classification score as a function of %V (reprinted from Ramus et al., 1999).

1.6. Speech Rhythm and Speech Rate

1.6.1. The idea of a relationship between rhythm and rate

Although the present work is on the influence of speech rate on rhythm as represented by C- or V-interval variability, the idea that rhythm and tempo can be related is much older than this. Thomson (1904) tries to characterize tempo influences on rhythm which leads him to formulate a 'law of tempo in rhythm-variation' in some later work (Thomson, 1923) in which he claims that qualitative/structural features of rhythmical units decrease with an increase in production speed.

> Law XV. The law of tempo in rhythm-variation.
> Increased rapidity beyond a certain medium tends, on the principle of economy or ease, to a leveling down of quantity and accent; diminished rapidity, on the same principle, to a wider discrimination of those elements. (p. 208)

This leads Thomson (1923) to conclude that

> It is a tacit fiction that quantities are unaffected by tempo. (p. 209)

From a merely intuitive point of view it seems plausible that rhythmic features and

rate may be interrelated. For instance, if one assumes that the hypothesis is correct that stress-timed languages show a pattern between stressed and unstressed syllables, then the basis for discriminating these types of syllables on a production level is by giving stressed syllables a higher articulatory effort than unstressed syllables. With increasing speed, these targets for a higher articulatory effort may no longer be reached, simply because of a lack of time to move articulators into the respective target positions or to vary air pressure in the lungs sufficiently.

Strangert (1985) however showed that this is not necessarily true for speech. She conducted an experimental study on the influences of speech rate on stressed and unstressed syllables in a phrase and on the total duration of the test phrase in order to find evidence concerning 'rhythmic grouping' of stressed and unstressed syllables. Methodologically she collected data by asking subjects to vary their speech between normal, slow and fast. Strangert intended to test two hypotheses, a) that the duration of a test interval changes proportionally with a change in speech rate, irrespective of the number of syllables the interval contains, and b) that with variable speech rate stressed and unstressed syllables are equally perceptually distinguishable on a temporal level. In an experiment with read speech she found that both hypotheses could be confirmed. Strangert concludes that the stressed/unstressed syllable distinction is maintained by the "different asymptotic values for the stressed syllable" (p. 80) at different speech rates. The important point in here is that evidence has been found that prominence structures do not change between different speech rates.

1.6.2. Rate normalization procedures

As described above in more recent views (Ramus et al., 1999, Grabe & Low, 2002) speech rhythm is regarded the product of the phonotactic organization of a language. The auditory impression of a language's rhythm is then assumed to be the result of durational characteristics of consonantal and vocalic intervals. The durations of these intervals are of course dependent on the rate of speech. On average fast speech must contain shorter and slow speech longer C- or V-intervals. It cannot be assumed, however, that speakers perform the process of interval shortening and lengthening across different rates on a linear basis. Gay (1978) for example found that with increasing speech rate vocalic rather than consonantal durations are affected in

English. It is rather likely that vowels and consonants do not change their durations linearly with rate.

Grabe & Low (2002) as well as Ramus et al. (1999) try to normalise for speech rate in different ways. Grabe & Low (2002) start from the assumption that vocalic rather than consonantal duration is affected by speech rate (partially based on findings by Gay, 1978). The authors argue that C-intervals are variable on multiple structural levels between languages (e.g. number of consonants per C-interval, qualitative structure of consonants, etc.); thus rate normalization of C-intervals would affect languages differently. For this reason rate normalization was not applied for the rPVI. This point will be reviewed again in later chapters (6) where it will be shown that rate normalization for rPVI is necessary and possible. For the vocalic nPVI Barry et al. (2003) criticise the vocalic rate normalization procedure in Grabe & Low as a purely local normalization (i.e. between two consecutive vocalic intervals) that does not account for global changes in rate within a phrase and even less between two or more different phrases. The data of this study, however, will demonstrate that Barry et al. (2003) were wrong (see chapter 6).

Ramus et al. (1999) normalize for speech rate variability by selecting sentences that are of roughly equal duration while they contain the same number of syllables. Here, however, the authors seem to confuse two measures: On the one hand, speech rhythm is monitored as the variability of C- and V-intervals; on the other hand, speech rate is monitored as the rate of syllables per second. Structurally these units may be very different. In syllable-timed languages syllables for example are structurally mainly composed of a single consonantal and a single vocalic segment (‘cv’ or ‘vc’) while in stress-timed languages structures consisting of a vocalic segment between two consonantal segments (cvc) are more common (see Dauer, 1983, or the results for the data used in the present study, chapter 3). With their normalization procedure, Ramus et al. make the implicit assumption that an increase in syllable rate leads to a proportionately equal increase in the rate of C- and/or V-intervals, or in other words: a constant syllable rate secures a constant rate of C- and V-intervals. Both assumptions, however, are not justified, as the data in chapter 5 of the present study will show. The Ramus et al. normalization procedure will then be reviewed in more detail.

In an attempt to normalize effects of speech rate on ΔC, Dellwo (2006) calculates the coefficient of variation for the standard deviation of C-intervals. This ensures that a possible dependency of ΔC on the mean durations of C-intervals is normalized for. However, for the current work this normalization procedure will be refrained from since it will be shown that C-interval duration populations are highly skewed and contain a fair amount of kurtosis. These data characteristics are as such unsuitable for the processing of means and standard deviations. For this reason, a different normalization procedure will be suggested (chapter 6).

1.6.3. Studies on speech rhythm and rate

The fact that C- or V-interval variability measures, as proposed by Ramus et al. (1999) and Grabe & Low (2002), are affected by speech rate has been pointed out in two experimental studies. Barry et al. (2003) studied a large database of Bulgarian, Italian and German. They found that the Ramus et al. (1999) measures ΔC and ΔV are both speech rate dependent. The values failed to separate the expected rhythm class differences but showed lower C- and V-interval variability (represented by lower ΔC and lower ΔV) as a function of rate. Barry et al. (2003) results, however, are rather impressionistic. They do not reveal the actual speech rates of their samples nor do they deliver results showing a statistical dependency between rhythm measures and speech rate (e.g. regression analysis) which makes it difficult to verify their claim.

For %V Barry et al. (2003) did not detect a speech rate dependency. This is why they claim that %V and another measure they derived from the PVI (the PVI-CV) differentiate rhythm classes in their data best. The PVI-CV does not treat C- and V-intervals separately but monitors the variability between a combination of consecutive C- and V-intervals (see Barry et al., 2003).

Dellwo & Wagner (2003) investigated speech rate influences on the rhythm measures ΔC and %V. For their highly variable speech rate data they find a) that ΔC is more speech rate dependent than %V but that b) the rhythm class distinction is still possible. In Figure 1-4 (taken from Dellwo & Wagner) these findings are illustrated. The figure shows ΔC plotted across %V for the languages German, English (both stress-timed) and French (syllable-timed). The dots in the graph represent average

values for 5 different intended speech tempi (speech intended to be produced very slow, slow, normal, fast and as fast as possible; see chapter 3 for an in detail description of the methodology) for each language.

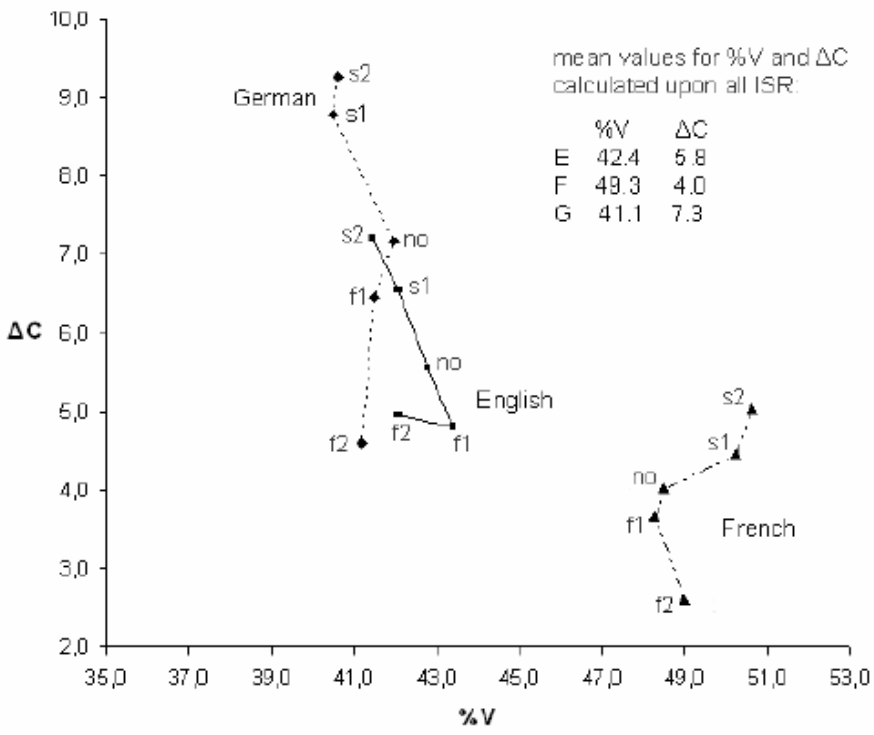

Figure 1-4: ΔC and %V cross plotted for the languages English (E), French (F), and German (G) (reprinted from Dellwo &Wagner, 2003). The different points in each language show variable intended speech tempo (s2 = very slow, s1 = slow, no = normal, f1 = fast, and f2 = fastest possible)

The data in Dellwo & Wagner is an earlier and smaller version of the dataset used for the current study. However, the study contains a number of deficiencies which were taken into account for the present work: In the Dellwo & Wagner study speech rate variability is only represented by the 5 different intended speech tempo classes. It will be shown, however, that actual laboratory measurable rates vary highly within these classes. Intended speech rate classes are therefore not necessarily a reliable rate predictor. (b) Only one language represents the syllable-timed class. This deficiency was taken account for in the present study by adding Italian to the dataset as another

example of classic syllable-timing. (c) The Dellwo & Wagner study lacks inferential data analysis. It will be shown that not all the differences pointed out in Dellwo & Wagner are actually statistically significant. (d) Dellwo & Wagner only analyzed rhythm correlates suggested by Ramus et al (1999). The present work also includes an analysis of the PVI measure.

1.7. Motivations for studying speech rhythm

Motivations for studying speech rhythm are manifold and have typically been in the context of language teaching (see Lloyd James, 1937, Milne, 1955, Adams, 1979, and the in-depth discussion in chapter 2). It was believed that speech rhythm contributes highly to the intelligibility of speech and that it is thus one of the main acquisition targets for non-native speakers.

Further applications can be found in the field of speech technology and there mainly in the area of enhancing the naturalness of speech synthesis systems (see Keller & Zellner, 1995, 1996, Zellner Keller & Keller, forthcoming b). Recent findings on C- or V-interval variability are motivated by the functions of speech rhythm in a language acquisition environment (Ramus & Mehler, 1999, Nazzi et al., 1998). The main argument is that infants use rhythmical cues (i.e. cues of durational C- or V-interval variability) in a bilingual environment to help them distinguishing between two languages (when they are members of different rhythm classes). The studies argue that the number of children growing up in a bilingual environment is constantly underestimated (it is the rule rather than the exception) and that one of the major tasks of infants in these particular environments is to separate between two languages during their language acquisition process. Observations show that bilingual children do generally not mix up the two different languages they grow up with. Rhythmical cues may well help performing this task. In an extension of this argument (Kim et al., forthcoming) it is claimed that listeners may use acquired rhythmic structures of a language in any speech perception situation to perform preliminary low level segmental segmentation of the signal.

2. History of rhythm studies in speech

2.1. The history of the concept of rhythm in speech

When the concept of rhythm got introduced into the field of speech research is unclear. It can be assumed that the terminology was established in music before it got used in speech. Early studies on speech rhythm implicitly assume that rhythm is only existent in poetical speech but not in prose, spontaneous, or read speech. In titles of translation work of the 19th century it is for example common to use the phrase 'rendered into English rhythm' to mean 'translated into the English verse'. This was typically the case for the translation of poetical Latin verse into English (see for example Singelton, 1855, in which the author entitles a translation of Latin verse by Vergil as "The works of Virgil, closely rendered into English rhythm […]")

It is difficult to name the first occurrence of the concept of rhythm in the literature. A survey of the 19th century literature shows that this concept seems equally present in studies on speech and music with speech being closely defined as poetic (see e.g. Roe, 1823, Edmonds, 1883, and Donovan, 1889). In combined studies of speech and music as for example by Richard Baillie Roe (1823) and in multiple studies of the rhythm of church psalms in the 19th century it is apparent that there is a common assumption to attribute a note or beat in music to a syllable in speech. It is probably here where the concept was born that the syllable is the basic unit of rhythm in speech and that the timing of syllables in return is responsible for our perception of rhythmic features in speech (an assumption that is most likely incorrect in respect to speech rhythm classes; see discussion in chapter 1). This view also finds support in early studies of the rhythm of church songs in which a syllable typically corresponds to a note. However, in such a situation the syllable's nucleus can be lengthened and shortened in ways untypical for spontaneous speech in order to make the syllable durations adapt to the rhythmical requirements of the corresponding melody.

At which point the concept of speech rhythm in poetry got extended to spontaneous or read (non-poetical) speech is again unclear. While the 1916 and 1923 works of William Thompson with the titles "Laws of speech rhythm" and "The rhythm of

speech" still treat speech rhythm as theories of verse and rhyme in poetry, it seems, however, that around the same time suggestions appear that rhythmical structures are also existent in prose. MacColl (1914) claims that:

> [...] the regular 'feet' of verse invade prose in all degrees till it is indistinguishable from verse [...]" (p. 49).

This claim is followed by the suggestion to combine rhythmical studies of prose and verse:

> I invite the reader, then to regard Prose and Verse not as sharply-divided entities under separate laws of rhythm, but as, in their characteristic forms, the extremities of a continuous chain, the variation being from freedom of syllable and emphasis towards strictness of foot and metrical pattern. (p. 49)

With this statement MacColl formulates early thoughts on the transferability of the metrical concept of rhythm to speech rhythm. It was probably statements like these that made speech researchers in the late 1930s draw direct analogies between poetical and prose speech rhythm.

Such analogies are, however, difficult and possibly misleading since the concept of rhythm in poetical speech or song needs to be treated differently from that of prose or spontaneous speech. The primary function of rhythm in poetry appears to be drastically different from that of rhythm in spontaneous speech: in poetry rhythm has solely a form function, i.e. it is responsible for creating a timing pattern that serves artistic rather than functional purposes (Yi Xu, personal communication). In spontaneous speech, however, artistic aspects are secondary if at all existent. These are again early sources of argumentation that probably lead to wrong assumptions in 20th century views on speech rhythm, above all the concept of syllable and foot isochrony in speech.

The discussion about the relation between produced and perceived rhythm is probably as old as the discussion about rhythm itself. Sonnenschein (1925) is one of the earlier authors of contemporary rhythm studies and he clearly distinguishes between a production and perceptual level:

> Rhythm is "that property of a sequence of events in time which produces on the mind of the observer the impression of proportion between the duration of the several events or groups of events of which the sequence is composed." (p. 16)

In the early literature it is apparent that authors on rhythm typically underestimated the complexity of the relationship between produced and perceived rhythm. It was often wrongly believed that there is some kind of linear relationship between these two levels (see Eriksson, 1991, who points out that this is particularly true for early authors on speech rhythm classes).

2.2. The History of the speech rhythm class hypothesis

2.2.1. The development of the concept of isochrony

Before the concept of isochrony entered the discussion on prose speech rhythm it was well-established in metrical speech rhythm. This becomes evident, for example, in Sonnenschein who notes in 1925 (well before the rhythm class was applied to prose speech):

> There is nothing novel in what I have said as to the proportioned duration of the groups of syllables in accentual verse. In one or another, and under various names, this doctrine has played a very important part in the history of prosody, and, indeed, has been often assumed as axiomatic. The ratio in which the durations of successive groups stand to one another has generally been supposed to be that of 1:1 - hence the terms 'isochronism,' 'Taktgleichheit' (equality of measures) [...].

Nevertheless, numerous authors before Sonnenschein (1925) do use the concept of successive metrical groups in poetry being in a 1:1 relationship but did not refer to it as 'isochrony'. In early 20th century work on speech rhythm the general idea of this concept in terms of equally long feet in English poetry irrespective of the number of syllables they contain was already well established. This concept was later transferred to a characteristic feature of the rhythm of stress-timed languages, in particular English. With the isochrony of English feet yet another concept of poetical rhythm studies entered the research on prose speech rhythm. However, no evidence can be found that early authors on speech rhythm in prose used isochrony to contrast different rhythms in different languages. Although these authors were well aware

that metrical rhythm in classical Latin and Greek poetry varies significantly in structure from English, no claim has been made that languages can systematically be grouped into different classes according to these rhythmical features. In the later literature on speech rhythm classes, isochrony of particular rhythmical units in different languages is seen as the most important concept for the earlier research in the field (see discussion on rhythm measurements in the introduction).

2.2.2. The development of the concept of the rhythm class

It has often been claimed that the origins of the rhythm class hypothesis are unclear and that it mainly goes back to work by Pike (1946), or Abercrombie (1967) (see e.g. Pompino-Marschall 1990). However, in-depths reviews of early rhythm literature, carried out in the context of the present study, reveal that the work of Arthur Lloyd James may be most responsible for the development of the idea. Moreover most of today's argumentation concerning the isochrony hypothesis and the rhythm class hypothesis seems to go back directly to his work. For this reason the work of Arthur Lloyd James will be discussed in reasonable detail in the following. It offers an insight into the ideas behind the rhythm class hypothesis that are to some degree still widely accepted.

The Welsh phonetician Arthur Lloyd James was an appointed lecturer in Phonetics at the University College London in 1920, one year before the legendary Daniel Jones became the first Professor in the field in the United Kingdom. Jones worked on numerous aspects of English pronunciation and Lloyd James's work is likely to have been influenced by this because of the same work environment they shared. In 1933 Lloyd James became professor for Phonetics at the School of African and Oriental Studies in London. Next to his academic career he was also the chief adviser to the BBC in issues regarding pronunciation. It seems that Arthur Lloyd James was a rather influential author at the time since numerous of his work in the 1930s was translated into other languages such as French and German (see e.g. Ogden, 1940, Heilmann, 1939, Hensel, 1940). Although Lloyd James' work lost the attention of the field at some point (for reasons we are not going to explore any further) it is evident that during his lifetime his work got read widely and thus had a major impact on the argumentation concerning speech rhythm and the rhythm class hypothesis.

Lloyd James (1929) published the (presumably) earliest study on cross-linguistic rhythm in speech. His scientific approach implicitly makes a distinction between stress- and syllable-timing, although he never used these terms explicitly (they go back to Pike, 1946). In Lloyd James (1929) quotes can be found which show a possible awareness of the concept of equal syllable durations in syllables but not in stress timed languages. The author writes (all following quotations are taken from Lloyd James, 1929):

> French students must not say **gud mɔniŋ**[1] but **gud mɔːniŋ […]** (p. 27)

The two transcribed versions of 'good morning' in the example are only distinguished by a length marker for the back half-open rounded vowel in the first syllable of 'morning' in the second version. This length marker in the vowel apparently signifies that the durational variability of syllables higher in the second than in the first version. The accompanying footnote confirms this assumption:

> [1]An example of even rhythm where the syllables are approximately of equal length.

Another example is given a little further below on the same page when Lloyd James corrects the pronunciation of English by Indian learners:

> And Indian students, practice a simple line like "This is the house that Jack built." Don't say what most of you so often do: **ðis iz ði haus ðaet dʒɑg bilt**.[3] (p. 27)

> And again in the accompanying footnote it says:

> [3] Another example of even rhythm […] (p. 27)

The description 'equal syllable length' presumably refers to the concept of 'even rhythm'. It seems obvious this incorporates the idea of the stress- and syllable-timing distinction. The argument that a French or Indian language influence changes the pronunciation of English syllables to 'an even rhythm' with 'equal syllable lengths' is still widely believed today (see the discussion in Grabe et al., 1999).

It is also interesting to note the languages he takes as examples (Indian, French, and English). While French is used in nearly all contemporary studies on speech rhythm

as a classic example of syllable-timed speech rhythm it can further be noted that rather influential work on currently established acoustic measurement techniques for speech rhythm are based on work using varieties of English pronunciation spoken on the Indian subcontinent (Low & Grabe, 1995).

In Lloyd James's 1940 study on pronunciation criteria for speech signals that are transmitted over the telephone, he distinguished clearly between the concepts of stress-timing and syllable-timing. This work was carried out with the aim to train users of telephone and radio signals in war time. Establishing such requirements was expected to lead to an improvement of the pronunciation of these users in order to enhance their intelligibility when the speech signal was distorted by typical signal degradations introduced by the telephone and radio technique. Lloyd James introduces the legendary terminology 'Morse-code' and 'machine-gun' rhythm; as metaphorical descriptions for stress-timing and syllable-timing respectively:

> The regular recurrence of accented syllables in our speech produces its characteristic rhythm.
> Speech rhythms are of two kinds.
> Morse code rhythm. English, Arabic, and Persian are good examples.
> Machine gun rhythm. French and Telugu are good examples.
>
> (Lloyd James, 1940, p. 25)

The usage of the Morse-code and machine-gun metaphors is taken rather literally as he continues to give examples for which he uses Morse-code illustrations to describe the rhythmic characteristics of British place names:

> London has the rhythm _ .
> Birmingham " " " _ . .
> Edinburgh " " " _ …
> Pontyridd " " " _ . _
> Ecclefechan " " " _ . _ .
> Dundee is . _ in Scotland, and often _ . in England
> Belfast is ._ in Northern Ireland, and often _ . in England.
> A signal such as
> **What's the time** has the rhythm
> _ . _
>
> (Lloyd James, 1940, p. 25)

The following quote from Lloyd James (1935) demonstrates that he was entirely aware of the concept of isochronous syllable durations in syllable-timed language

like French.

> The English language on an unfamiliar rhythm, e.g. French rhythm, sounds odd, and often quite unintelligible, as, for example, "Will you please to give me a ticket to Tottenham Court Road?" Try this yourself. Try to read a sentence of English putting the same accent on every syllable and making all the syllables exactly equally long. (p. 86)

It seems rather unlikely that Lloyd James (1940) used the terms machine-gun and Morse-code for the first time. Though he omits to quote a reference in any of his work he uses the terms in a way as if it was common knowledge, not giving any definition of the terms as to what the acoustic implications of this may be. For this reason it is assumed that at the time of Lloyd James the concepts of stress-timing and syllable-timing may have been a part of common knowledge amongst phoneticians and did not require any further explanations.

2.3. Original motivations for studying rhythm in speech

Motivations for studies of speech rhythm vary throughout the field. It is apparent from the previous discussion that speech rhythm has been of major interest in the field of phonetics. By now the rhythm class hypothesis has been revisited and discussed for over seven decades. However, surprisingly little effort has been made to argue for the importance of rhythm in speech communication.

At first sight speech rhythm could be regarded a rather minor feature of speech since after all it does have no direct linguistic/phonological function in the speech communication process. Unlike other prosodic features, as for example intonation, we cannot distinguish the meaning between two phrases or sentences or any other unit of speech on the basis of rhythm. Rhythm neither seems to have clear paralinguistic relevance since it is difficult if not impossible to mimic for example two different emotions on the basis of rhythm alone. So what is the importance of rhythm in speech?

In the context of pronunciation training to enhance speech intelligibility via telephone or radio transmissions, Lloyd James (1940) points out rhythmical characteristics of the speech signal as one of the key features to train for enhancing the intelligibility of speech:

> In Speech signals as in Morse signals, the better the rhythm the better the signal. Good rhythm means punching out the accented syllables, keeping them at equal intervals of time apart, and not smothering up the unstressed syllables. It is exactly like good rhythm in a dance band. (Lloyd James, 1940, p. 25/26)

And in another context Lloyd James (1935) argues:

> The easiest way to be unintelligible in a language is to speak it on a wrong rhythm, and incidentally to imagine that we can pick up a foreign rhythm or a foreign intonation easily, is a common mistake. (p.86)

With speaking in a 'wrong' rhythm Lloyd James generally gives the example of applying French syllable-timed structure (or in Lloyd James' terms: machine-gun rhythm) to English stress-timing (Morse-code rhythm) which, according to Lloyd James, has a major effect on intelligibility.

Therefore, considering the non-functionality on the one hand and the intelligibility aspects of speech rhythm on the other, one may argue that rhythm is an acoustic feature that is of no particular importance for the speech communication process as long as it is well performed according to the rhythmical norms set by the language. Violating rhythmical well-forming in connected speech, i.e. performing wrongly on durational aspects of the speech segments, may probably cause more significant intelligibility problems. This is for example the case when speech is transmitted in a very noisy environment like a public bar with strong background noise produced by other speakers and music from a stereo-system, or a traffic scene with loud interfering car noise. Under such adverse listening conditions, a great amount of the segmental information in speech is masked but even though mutual intelligibility is generally possible. One reason for this may be that listeners expect acoustic units of speech at predictable moments in time. These predictable moments could be present in the cognitive system of the listener because of knowledge of the rhythmical arrangement of units of speech.

Support for the theory of acoustic well-forming can be found in three major areas in which speech rhythm has been discussed predominantly, a) second language teaching/acquisition, b) speech technology (in particular speech synthesis, and c) speech pathology. All three fields are concerned with the speech signal being in an

acoustically well-formed state. In a) this must be acquired by the foreign learner, in b) it must be incorporated in a synthesis system and in c) this is what is distorted in the signal of pathological speech (among other features).

Numerous early linguistic works on prosodic parameters, in particular intonation, seem to be motivated by improving language teaching and learning (Yi Xu, personal communication). This situation is clearly true for speech rhythm. The application to the teaching environment has been one of the major driving factors for work in this field. The earliest evidence can be found again with Arthur Lloyd James. In one of his numerous Gramophone English Language teaching courses he claims that rhythm is the key feature to an acceptable pronunciation:

> (p.26) This language-rhythm seems to me to be at the root of many hitherto unexplored language problems and certainly is the very first and most striking feature in English speech. Unless and until you get it right you will never speak English well. (Lloyd James, 1937, p. 27)

The extraordinary importance of rhythm in Lloyd James' view is also nicely revealed by the following statement:

> My last word to all of you is to use your gramophone repeatedly; get this English rhythm fixed in your mind. Never learn a sentence, or indeed a word, without getting its rhythm right. This is one of the great secrets of pronunciation. (Lloyd James, 1937, p. 27)

Although Arthur Lloyd James has hardly received credit for his original thoughts, it seems obvious that these early statements have been highly influential in the field of second language acquisition. Numerous publications deal with speech rhythm and the rhythm class hypothesis (see e.g. Milne, 1955, Adams, 1979). Most of them share the general idea that an understanding of the rhythm of a language can enhance the pronunciation quality of learners in a second language. It is however typical for such studies that they are not particularly concerned with the acoustic reality of the isochrony hypothesis. This hypothesis, in its ideal form, is rather treated as being supportive for a learner to actively practice different rhythmical structures in a language even though acoustic correlates of syllable- and stress-timing are highly questionable (see chapter 1).

It can be assumed that there are numerous similarities between the problems that

learners of a foreign language and a speech synthesis device meet in producing acceptable rhythmic prosodic features in order to enhance the intelligibility of their speech output. And it can be assumed that this knowledge is also helpful in describing and analyzing rhythmical problems in distorted speech. Next to these fields a major part of the studies on speech rhythm have general phonetic interests rather than a specific application as their driving factor.

3. Data and data analysis procedures

3.1. Introduction

To study rhythm and speech rate measures in combination, a database was compiled for the present study that contains read speech material with highly varying speech rates (BonnTempo Corpus; henceforth: BTC). C- and V-intervals as well as syllables were labelled in the database to make the analysis of a wide variety of rhythm and rate measures possible. Labelling was carried out in the speech analysis software Praat using the annotation tools provided (see www.praat.org and Boersma, 2001). The analysis of the BTC with respect to a wide variety of rhythm and speech rate measures is facilitated by a collection of Praat based software tools (Praat scripts), the BonnTempo-Tools (henceforth: BTT). The present chapter gives an overview of content and construction of the BTC and explains the facilities of the BTT. The BTC and the BTTs are the basis for the experimental data analysis in all following chapters so the current chapter is the methodological kernel of the present work (this database is available from the author, see www.phonetiklabor.de for contact details).
In the next section of the introduction a characterisation of typical data sets used in a selection of other studies on speech rhythm is given. This is followed by a description on other research projects the BTC is currently being used for.

3.1.1. Databases for studying speech rhythm

Ramus et al. (1999) and Grabe & Low (2002) are an improvement over earlier studies in terms of the number of speakers and syllables used, although the amount of available material is still limited. In Ramus et al., eight languages are represented by four speakers each, while five sentences of 15 to 19 syllables (mean=17) have been recorded for each speaker. This adds up to 2720 syllables for the whole database (8*4*5*17) or 340 syllables per language. Grabe & Low examined 16 languages, each represented by one speaker reading the original or a respective translation of 'The North Wind and the Sun', which contains 141 syllables in the English version. Assuming the average number of syllables in each language version is around 150 syllables/text, the total number of syllables for the whole data analyzed sums up to

2256 syllables (16*150).

The use of a limited amount of data may lead to artefacts in the results, in particular in rhythm studies. For this reason studies based on Ramus et al. and Grabe & Lows's measures, such as Barry et al. (2003), Dellwo & Wagner (2003) and Dellwo et al. (2004) started using larger data collections. Dellwo & Wagner (2003) and Dellwo et al. (2004), who used an earlier form of the corpus presented here, base their study on three languages or 6420 syllables (about 2140 syllables/language). So far Barry et al. (2003) have apparently used the largest data set in rhythm research with more than 5000 inter-pause stretches for 3 languages. Since the authors do not specify the number of syllables in their database, a straightforward comparison is difficult. But considering the number of inter-pause stretches the number of syllables in Barry et al.'s database must sum up to several thousands. However Barry et al. do not reveal labelling techniques for their data so it is difficult to judge whether they used manual labelling or machine techniques. Regarding the large quantity of data from various different sources it seems likely they used already existent labelling on these data sets and transferred them either manually or automatically into the required formats to calculate respective C- or V-interval variability measures, which may introduce numerous artefacts (see Steiner, 2004).

3.1.2. Applications of BonnTempo

Ever since the BonnTempo database was used first by Dellwo & Wagner (2003) it became popular in many international research projects such as the ones listed in below in which the author of the present study is actively involved. An earlier version of the BTT and BTC was described by Dellwo et al. (2004).

One of the ideas behind BonnTempo is that it is easily extendable. New languages can be added using the existent format and can thus be compared with the existing data set. Following this idea the following research projects are currently known of using BonnTempo:

3.1.2.1 Czech speech rhythm (Jana Dankovičová, Volker Dellwo)

A group of researchers at University College London has now compiled a Czech section of the database. With this project a yet non classified language was added to

Bonn Tempo. Rhythmic classification of Slavic languages has been widely disputed, and Czech was included in the database as a good example of this. It has traditionally been classified as syllable-timed but later shown to have a tendency towards stress-isochrony, or towards either rhythmic patterns, depending on the type of measure (see discussion in Dankovičová & Dellwo, 2007).
Eight speakers in BonnTempo format have so far shown interesting results on C- and V-interval variability in Czech which seems to be syllable-timed according to V-variability but stress-timed according to C-variability. The Czech data was also used in order to test some hypotheses in the present work (see chapter 4, 5, and 6).

3.1.2.2 Development of L2 speech rhythm (Francisco Gutierrez-Diez, Volker Dellwo, Stuart Rosen)

In a grant-funded co-operation program (Seneca research fund, Spain) between Murcia University/Spain and University College London a group of researchers has started compiling Spanish L1 and Spanish speakers of English L2 data in a longitudinal study. The aim of this project is to monitor whether L2 English speech rhythm of Spanish speakers changes linearly as a function of expertise in English.

3.1.2.3 Rhythmical structure of glottal signals (Adrian Fourcin, Evelyn Abberton, Volker Dellwo)

This project uses the BonnTempo data as a basis and compiles new data based on laryngographic recordings (waveform of vocal fold contact area). The aim of the project is to investigate the relationship between vocalic and voiced intervals. First results demonstrated that languages of different rhythm classes are also distinguished by varying durational characteristics of voiced and voiceless intervals as opposed to vocalic and consonantal intervals (Dellwo et al., 2007). The advantages of such a technique are in a first instance of methodological nature since the automatic segmentation of voiced and voiceless stretches in speech can very reliably be performed automatically (in particular when a laryngographic waveform, i.e. a waveform of the vocal fold contact area during phonation, is available). The technique may also have implications about infant rhythm perception (see discussion

in the introduction). Since it is argued that infants are familiarized with speech rhythm in the mother's womb (Ramus et al., 1999) the highly low pass filtered signal the infant perceives is probably very similar to the fundamental frequency contour of speech.

3.2. The BonnTempo corpus

The BTC currently consists of 24 070 syllables and 43 227 C- and V-intervals (C = 22 705; V = 21 522) for 5 languages and 4 L2 conditions, while the absolute number of speakers (and thus syllables) per language still varies considerably (cf. 2.2).

3.2.1. Speech material

The speech material in BonnTempo consists only of read speech. The text is a short passage from a novel by Bernhard Schlink ('Selbs Betrug') with 76 syllables in the German version. This text was translated into the other languages under investigation by philologically educated native speakers of the target languages English (77 syllables), French (93 syllables), Italian (106 syllables). Czech (93 syllables) and Spanish (120 syllables) versions of the text have been recorded and a number of speakers was added for these languages (see above), however these languages are not part of the current project and will not be elaborated on. All reading texts including the syllabic segmentation are available in APPENDIX II.

The languages were selected to represent both traditional rhythmic classes. 'Stress-timing' is represented by English and German, 'syllable-timing' by French and Italian. It was decided to include read speech only to reduce the variability of the speech samples to a minimum. Significant differences between read and spontaneous speech have been reported by Barry et al. (2003). It seems plausible that the variability in spontaneous samples can also be high because of numerous different types of hesitations, which may lead to a wide variability of vocalic or consonantal lengthening. This, in return, may have large effects on their durational variability. In short, it seems that read speech is the most appropriate way to reduce this variability as much as possible.

3.2.2. Phonotactic complexity in the languages under investigation

The general argument underlying rhythm measures based on C- and V-interval variability is that there is higher variability due to higher phonotactic complexity on both the consonantal and the vocalic level (see Introduction). In this respect stress-timed languages are supposed to show a higher durational variability of both C- and V-interval durations than syllable-timed languages. On a consonantal level this higher variability is assumed to be caused by the complexity of C-clusters that is higher in languages traditionally classified as stress-timed. On a vocalic level this complexity is assumed to be mainly driven by vowel reductions which are a feature of most stress-timed but not of syllable-timed languages. Vocalic reduction leads to a great number of vowels in the speech signal to be reduced in quality as well as duration which in return leads to a higher variability in vocalic duration.

V-interval complexity is rather straightforward to assess, since languages either allow vowel reduction or not. This means that only two classes are possible here. The languages under investigation representing stress-timing (English and German) are languages that do allow vocalic reductions while the languages representing syllable-timing (French and Italian) do not. Czech as an example for not yet classified languages does not allow vocalic reduction but it does have a very prominent phonological length distinction that may introduce considerable durational V-interval variability (see Dankovičová and Dellwo, 2007).

To check whether the BonnTempo data is representative on the basis of consonantal complexity the syllabic and C-interval complexity for the languages under investigation was tested. The results are presented in the following and are based on an analysis of an assumed standard pronunciation of the reading material. It must be pointed out that individuals may vary in the actual realization (mainly through the elision of consonants) of these patterns, in particular in respect to the different intended speech tempo versions. It must be assumed that considerably more consonantal elisions take place when speech tempo is accelerated. However, for the current purposes it has been considered sufficient to study whether the language material is generally representative for stress- and syllable-timing in respect to the syllabic complexity rationale. For this reason only an estimation of the complexity under normal conditions was performed. A study of changes in syllabic complexity

as an effect of speech rate would be interesting but would lead to far away from the central topic of this work. Additionally, it would require information about the segmental structure of each individual C- and V-interval in the corpus which is not available in BonnTempo.

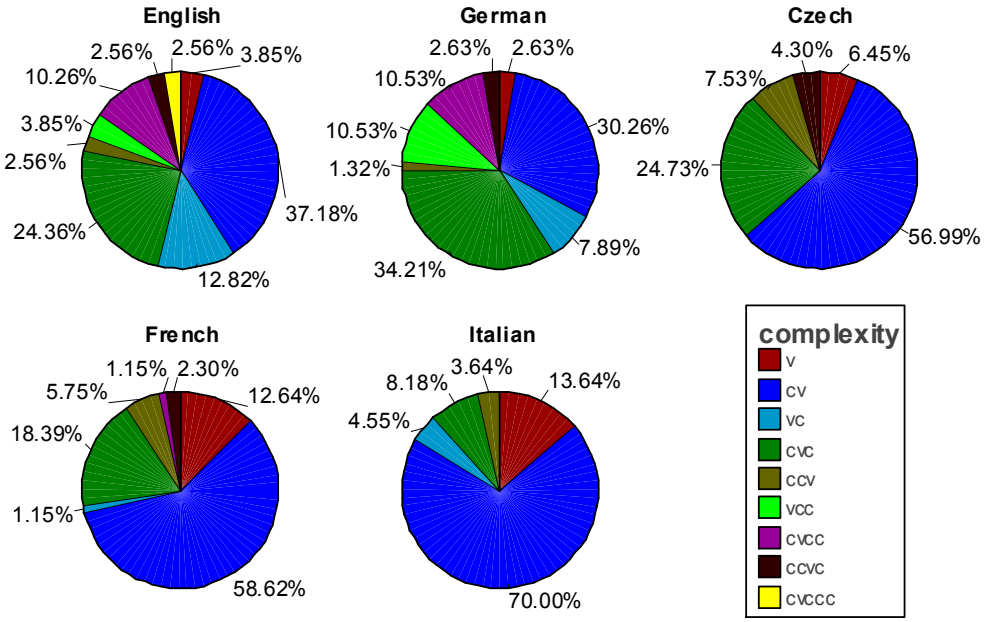

Figure 3-1: Pie charts for the languages German, English, Czech, French, and Italian showing the complexity of syllables.

3.2.2.1 Syllable complexity

The phonotactic syllable complexity for the languages German, English, French, Italian and Czech is summarized in Figure 3-1. For the purpose of complexity attribution, syllables were categorized in complexity classes. The least complex class is characterized by a single vowel syllables (v) and the most complex class by a syllable consisting of a consonant in the syllable onset and three consonants in the offset (cvccc). All other complexity patterns, organized by increasing complexity, can be viewed in the legend to the figure at the bottom right (small character ‘c’ and

'v' abbreviations stand for consonantal and vocalic segments respectively). From the pie charts it is obvious that the BonnTempo text for syllable-timed languages is characterized by a much higher amount of syllables of the segmental structure v, cv and vc (around 73% for French and 90% for Italian) then stress-timed languages (around 40% for German and 53% for English). The remaining percentages for each language are filled with more complex syllables while the segmental cvc pattern is most prominent (however, a complexity up to cvccc is observable for English). Czech, a yet not classified language, lies in between stress- and syllable timing patterns with about 63% of single v and cv segmental patterns (no vc-segment combination is allowed in Czech).

While phonotactic syllable complexity can be observed to be different between stress- and syllable-timed clusters it needs to be pointed out that there is considerable variability within the clusters, especially within the stress-timed cluster. This is to say, although English is phonotactically less complex than French it remains unclear whether it is actually more similar to German with which it is said to share the same rhythmic class. It can probably safely be concluded that English is somewhere in between German and French in terms of phonotactic complexity. However, an exact assessment of complexity is difficult here and will be further discussed below.

3.2.2.2 C-interval complexity

More important than syllable complexity is probably the C-interval complexity, i.e. the number of c-segments a C-interval consists of. After all it may be that languages do allow a wider variety of complex syllable clusters but they may arrange them in a way that C-interval complexity is kept low. If a language has segmental patterns of the type v, cv, and cvc they may for example be aligned in the form cvc+v+cv so that the actual C-interval complexity remains at one consonant for each interval. However, the pattern may be aligned v+cvc+cv, thus creating a double consonant between the second and third syllables of the example. This tells us that if we want to ensure that intervocalic consonantal complexity is different in the languages under investigation, we need to have a look at the number of consonants a language allows in between two vowels.

The results of such an analysis for the reading material in BonnTempo can be viewed

in Figure 3-2. The highest C-interval complexity that could be observed in any of the languages was a pattern of four consecutive consonants (cccc). This pattern can only be observed in stress-timed English and German, but occurs in rare cases (about 3% in each language). The more distinctive class between stress- and syllable timing is a cluster type consisting of one consonant only (c). This cluster is most prominent in French and Italian (75% and 80% respectively) but less prominent in English and German (51% and 37%). Czech again is somewhere in the middle with about 60%. Triple consonant clusters (ccc) do occur in the stress-timed languages at about 13% and in the syllable-timed languages (only in French) at a negligible amount of about 2.5%.

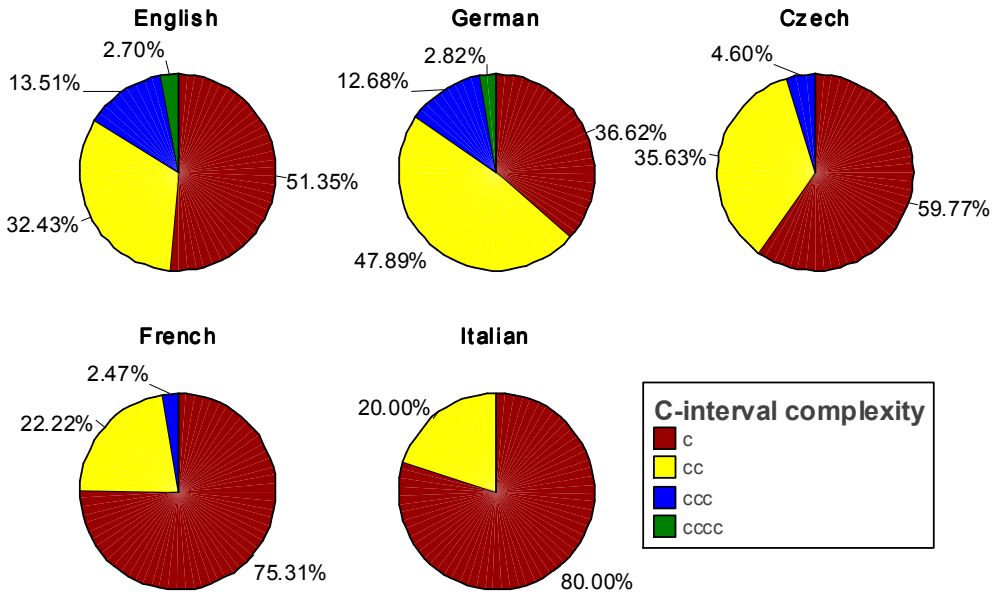

Figure 3-2: Pie charts for the languages German, English, Czech, French, and Italian illustrating the number of consonants in a consonantal interval.

Like with syllabic complexity, again, it can be argued that variability within the stress- and syllable-timed classes is considerable. English, although far lower than French in single c-segment C-intervals (51% as opposed to 75% respectively), shows

a considerably higher amount of single consonant C-intervals than German. An assessment of absolute complexity is problematic here because of the variability of higher complex clusters. Reasons for this are discussed below.

3.2.2.3 Discussion and conclusion about phonotactic complexity

The analysis of syllabic and C-interval complexity has demonstrated that the reading material in BonnTempo adequately represents the phonotactic characteristics assumed for stress- and syllable-timed languages. Stress-timed languages show a considerably higher amount of complex syllables and complex C-clusters than syllable timed languages. It has to be pointed out, however, that although stress-timing and syllable-timing seem to be well-separated in terms of phonotactic complexity, there is high within rhythm class variability. German appears in many respects more complex than English and French appears slightly more complex than Italian. An exact assessment of complexity is difficult because it is unclear to what degree individual structures contribute to complexity and how they can be compared. It must also be stressed that the quality of consonants in a C-interval may contribute highly to its complexity. After all, the argument states that C-interval complexity attributes to the overall duration of a C-interval. However, a C-interval consisting of three c-segments (e.g. nasals, approximants and/or laterals) is probably much easier to produce than the same type of interval consisting of a combination of plosives and fricatives. Thus a further analysis of the consonant types in C-intervals would be necessary to assess the overall complexity. Important findings have been gained here by Steiner (2004) who demonstrated that nasal consonants contribute most to the overall durational variability of C-intervals. However, such investigations would be part of a different study analyzing the influence of C-interval quality on C-interval duration.

The aim of the present analysis was to test if the BonnTempo reading material is likely to trigger speech which fulfills the general assumption that stress-timed languages are characterized by phonotactically more complex syllables (or C-intervals) than syllable-timed languages. It was demonstrated that this assumption is fulfilled by the reading texts. As such the reading material is considered suitable for the present study.

3.2.3. Speakers

The BTC 1.0 contains examples from speakers of the languages English, French, German, Italian, and Czech. Table 3-1 shows the number of speakers, syllables, C- and V-intervals, and pauses for all languages under investigation. The abbreviation in brackets behind the language names is made up from the language the text was read in capital letter (e.g. E for English) and the native language of the speaker(s) in small letter (e.g. e = for English); see below for a more detailed explanation.

language	speakers	syllables	C-intervals	V-intervals	pauses
English	7	2684	2475	2444	261
French	6	2734	2420	2455	250
German	15	5698	5028	4832	468
Italian	3	1619	1335	1380	95
Czech	8	3720	3608	3653	392
Total	**39**	**16455**	**14866**	**14764**	**1466**

Table 3-1: Table showing the number of languages, speakers, syllables, C-intervals, V-intervals, and pauses for native speakers of the languages English, French, German, Italian, and Czech in the BTC.

During the recording process it became apparent that one of the French speakers had a significantly different pronunciation (auditive judgement) compared to all other speakers in this group. It turned out that this speaker was from Cameroon and spoke the local variation of French typical in this country. Since there was consent among a group of phonetic expert listeners that this speaker revealed atypical rhythmical features for French (the speech sounded much more strongly syllable timed than any other French speaker) these speech recordings were treated separately for the data analysis.

- French as spoken in Cameroon (n = 1, 443, C = 389, V = 408, P = 74)

The database further contains the text read by non-native speakers of the respective languages. Table 3-2 shows the content of non-native read speech in BonnTempo. In the abbreviation the capital letter again refers to the text language (e.g. E for English) and the small letter to the native language of the speaker (e.g. f for French).

native language	text language	speakers	syllables	C-intervals	V-intervals	pauses
German	English	8	3087	2873	2833	247
German	French	8	3503	3119	3151	333
French	English	2	776	748	740	78
French	German	1	380	318	304	22
English	German	3	1140	1070	1012	135
Total		**22**	**8886**	**8128**	**8040**	**815**

Table 3-2: Table showing the non-native speaker content of the BTC by displaying the number of speakers, syllables, C-intervals, V-intervals and pauses for German speaking English (row 1) and French (row 2), French speaking English (row 3) and German (row 4) and English speaking German (row 5).

Further the database contains two subjects speaking German in syllable-timed manner, i.e. trying to produce syllables isochronously from a speaker's point of view by attributing equal prominence each syllable. Only a normal and a fast intended speech tempo are available here.

- Germans speaking German syllable-timed (2, S = 304, C = 303, V= 301, P = 10)

Recordings of the following languages have been made and are presently being labelled for future studies:

- Brazilian Portuguese
- European Portuguese
- Polish
- Russian
- Spanish
- Spanish speaking English (before and after a year-long university training in English)

We intend to extend the database to include more languages in the near future,

especially languages of the third recognised rhythm type - mora-timed languages (e.g. Japanese).

3.2.4. Recording procedure

During the recording process speakers first familiarized themselves with the text by reading it aloud while the recording levels were set. Speakers were allowed to practice the text as many times as they wanted (on average speakers did read through the text about four times) before the actual recording started. For the first recording they were asked to read the text in a way they considered 'normal reading'. It was left to the intuition of the speakers what 'normal' would be, assuming that there is some common sense native speaker intuition about normal reading in a language (see Dellwo et al., 2006, for a perceptual study of the psychological realities of intended speech tempi). After the task to read the text normally, speakers were asked to read the same text at different intended speech tempi: First speakers were asked to read the text slowly and then even more slowly. Following the recordings at slow rates, speakers were asked to read the text fast after which they consecutively had to increase their reading speed until they found themselves unable to speak any faster, or until reading quality became so poor that labelling would be impossible and recording was stopped. Speakers varied between three and eight attempts of the fast versions.

Recordings were carried out mainly in the sound proof booth of the Institut für Kommunikationsforschung und Phonetik (IKP, now Institut für Kommunikations-wissenschaften, IfK) at Bonn University with a large membrane condenser microphone directly on PC in Windows wav-file format (sampling rate = 44100 Hz, 16 bit quantisation). Most of the recordings for French were carried out in Bordeaux in private homes on mini-disc. The mini-disc signal was played out analogously to a PC sound card and then sampled in wav-file format (as above). All recordings for Czech were carried out in private homes in the Prague region on DAT and were then resampled to wav-file format in the same way as it was done for the mini-disc recordings (see above). Since the basic interest in the database is in segment durations differences in different recording techniques and places only play a minor role. While possible artefacts arising by the use of compression algorithms like

ATRAC (used by Sony Mini-Disc) are constantly under debate, it has been shown that the typical distortion types of this signal do not affect speech analysis (Van Son, 2002). It is however recommended to treat possible root mean square measurements or intensity in the mini-disc and DAT recordings with care since recording levels were automatically controlled by the respective devices. This inevitably leads to non-linear adjustments of consecutive amplitudes.

3.2.5. Labeling

Labelling of syllables as well as C- and V-intervals was carried out by human labellers to the normal version (normal), the two slow versions (very slow & slow), the first fast version (fast) and the fastest version (very fast). As specified above, to produce the fastest version speakers were given an indefinite number of trials. The version that was chosen to be 'very fast' was the version with the highest articulation rate (syllables/second) and the lowest amount of syllable elisions (typically no more than three syllables were permitted). Syllables have been labelled as phonological syllables unless no acoustic trace of the syllable could be found (elision). C-intervals were labelled from the offset of the preceding vowel (or pause) to the onset of the following vowel (or pause). V-intervals were labelled from the offset of the preceding consonant (or pause) to the onset of the following consonant (or pause). For the definitions of C- and V-intervals please refer to section 1.2.

Figure 3-3 shows a screenshot of a typical labelled section of a BonnTempo file, here the German phrase ‘Hinter Gießen werden die Berge und Wälder eintönig, [...]’. Four labelled tiers are visible at the bottom of the figure, a) the syllable tier, b) the tier containing all C- and V-intervals, c) the inter-pause-interval tier, and d) a random interval tier (control tier). Tiers a) and b) have been labelled manually, tiers c) and d) automatically (deriving from a) and b)). All pauses in the database are tagged as ‘p’, C- and V-intervals are tagged as ‘c’ and ‘v’ respectively; syllables are tagged with the respective syllable morpheme.

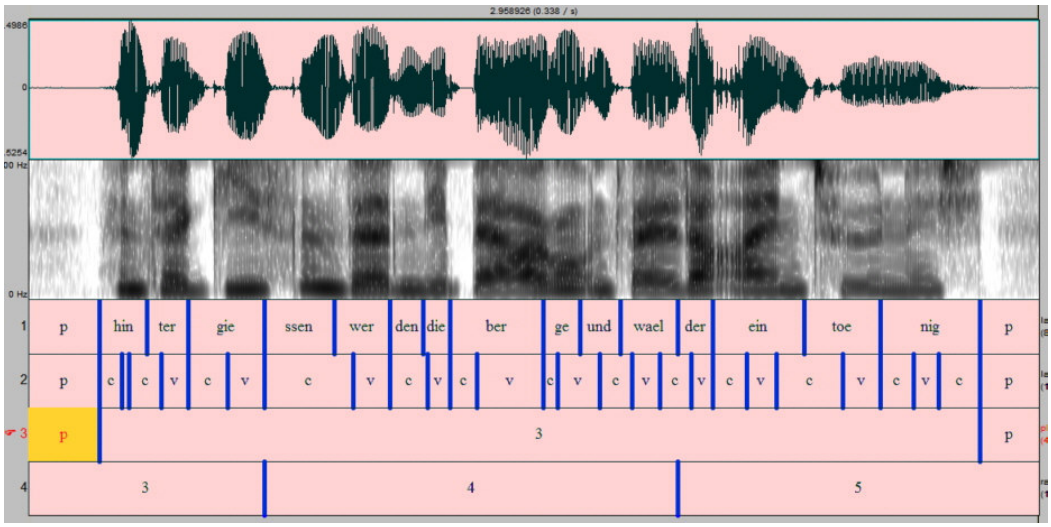

Figure 3-3: Screenshot of a typical labeled stretch of speech (here the German phrase 'Hinter Giessen warden die Berge und Wälder eintönig' between two pauses). The four tiers at the bottom of the screen represent 1) syllables, 2) C- and V-intervals, 3) inter-pause intervals, and 4) random segmentation.

Manual labelling of the data is an important issue in the construction of databases of the type presented. Currently available tools for automatic labelling have produced very poor results (especially for the fast versions) in the BTC and were therefore not considered (cf. Steiner, 2004, for a detailed account). Alternative tools are currently under observation and may be considered to assist future labelling work.

3.3. Data analysis

3.3.1. Analysis tools

The main tool for the analysis of the duration in speech was the popular speech processing package PRAAT (downloadable under a GUI license from www.praat.org). The statistical analysis software SPSS has in addition been used for further descriptive and inferential data analysis. In order to facilitate access and analysis of the BTC a collection of Praat based software (Praat scripts) was written by the author of the present study, referred to as the BonnTempo-Tools (BTT). The tools are independent of the corpus and are not necessarily needed to perform the analyses; however, they are required to export the durational information from the

Praat label files into a standard tab-separated table format (which can then be used by standard statistics software packages like SPSS). APPENDIX III provides an overview and manual instructions of the major functions of the BTT. The source code for all analysis tools can be obtained from the author (see www.phonetiklabor.de for contact details).

The BTT are a very useful set of tools to obtain an overview tailored to standard acoustic rhythm and speech rate correlates along with the type of data occurring in the BTC. They have been used widely in teaching environments (numerous universities have requested copies from the author) and also for parts of the data analysis of the present work. However, for a more in-depth and inferential statistic analysis of the data standard statistics processing programs like SPSS or R are more suitable. For this purpose, a command was created in the BTT which allows the export of relevant BTC data into standard tab separated spreadsheets to facilitate the transfer of the data into standard analysis tools (see APPENDIX III).

3.3.2. Analysis parameters

Amongst the analysis parameters are basically all mean values (mean), standard deviation of mean values (stdev) and the coefficient of variation (standard deviation proportional to the mean; henceforth: varco) of respective segment durations (syllables, consonantal and vocalic intervals, and also pauses). In addition to theses values laboratory measurable speech rate (lsr) or articulation rate (syllables/second without pauses) as well as the nPVI and rPVI are available. The following list gives an overview of the calculation for the major analysis parameters that have been used in the present work. This will serve as a reference list for the experimental chapters 4 to 9 in which these measures will be analyzed and discussed.

3.3.2.1 Laboratory measurable speech rate

Laboratory measurable speech rate (lsr) is calculated in two different ways in the present work, (a) as syllable-rate (Equation 3-1) or (b) as the CV-rate (Equation 3-2). In the following these rates will be explained further.

Syllable rate, is the rate of syllables per second, excluding any pauses or non-speech content. This rate is a parameter that is widely chosen for acoustic measurements of

speech rate in many contexts. Syllable-rate is assumed to be a good indicator for perceived speech tempo, however, for small utterances also phone and word rates can be important (see Pfitzinger, 1999, for an alternative approach). Syllable rate is calculated according to following formula.

Equation 3-1: Syllable-rate

$$\text{syllable-rate} = \frac{N_{syll}}{d}$$

N_{syll} = total number of Syllables in utterance
d = total duration of utterance in seconds (excluding pauses)

The CV rate is the sum of the number of C- and V-intervals per second. This rate is particularly important for the calculation of speech rate in the perception experiments when delexicalized speech is used in which only cues to C- and V-intervals are maintained (chapter 8). Whether CV-rate is a better predictor for perceived speech tempo than syllable-rate in real speech utterances is unclear and remains to be tested. Interesting cases for such a test would be utterances in which the ratio between the total number of C- and V-intervals in an utterance and the total number of syllables is not 2:1. Such situations emerge, for example, when V-intervals consist of multiple v-segments, e.g. as a result of single v-segment syllable being preceded by a syllable with a v-segment in syllable final position and followed by a syllable with a v-segment in syllable initial position (see 5.1.1 for a detailed example). In such cases the V-interval will be composed of v-segments from up to three syllables. Should the V-interval rate have an influence on perceptual tempo, this should be perceived as slower under such circumstances. The results of such a test, however, would not contribute to the main argument of the present work and are therefore not obtained. The following shows the formula for the calculation of CV-rate.

Equation 3-2: CV-rate

$$\text{CV-rate} = \frac{N_C + N_V}{d}$$

N_C = total number of C-intervals in utterance
N_V = total number of V-intervals in utterance
d = total duration of utterance in seconds (excluding pauses)

3.3.2.2 Standard deviation of C-intervals

The standard deviation of C-intervals (ΔC) was found to be a reliable predictor for rhythm class in Ramus et al. (1999). The formula for calculating ΔC is given below (the formula can be used to calculate standard deviations of other units too). To avoid fractional values the numbers for presentation are typically multiplied by 100 (see Ramus et al., 1999 and Grabe & Low, 2002).

Equation 3-3: Standard deviation of C-interval durations (ΔC)

$$\Delta C = \sqrt{\frac{N_C \cdot \sum_{i=1}^{N_C} C_i^{\,2} - \left(\sum_{i=1}^{N_C} C_i \right)^2}{N_C \cdot (N_C - 1)}}$$

N_C = number of C-intervals in utterance
C_i = duration of the i^{th} C-interval

3.3.2.3 Coefficient of variation of ΔC

As a measure for rate normalization (see chapter 6) the coefficient of variation (varco) has frequently been applied to measures based on standard deviations such as ΔC (varcoC; see Dellwo, 2006). This was implemented using the following formula (the formula can be used to calculate the coefficient of variation for other units too):

Equation 3-4: Coefficient of Variation of C-interval durations (varcoC)

$$\text{varcoC} = \frac{\Delta C \cdot 100}{((\sum_{i=1}^{N_c} C_i) / N_C)}$$

C_i = duration of the i[th] C-interval
N_C = number of C-intervals in utterance
ΔC = standard deviation of C-intervals (see equation 3-3)

3.3.2.4 ΔC based on ln-transformed C-interval durations

In section 6.3 it will be demonstrated that the above varcoC has disadvantages as a speech rate normalization measure. For this reason a new rate normalized measure will be introduced. For this measure ΔC will be calculated on the basis of C-interval durations which are expressed as logarithms to the base e (ln-transformed). This measure will be referred to as ΔCln. The formula for this measure is given in the following

Equation 3-5: ΔC based on logarithmically transformed C-interval durations.

$$\Delta C \ln = \sqrt{\frac{N_C \cdot \sum_{i=1}^{N_C} (\ln C_i)^2 - \left(\sum_{i=1}^{N_C} (\ln C_i) \right)^2}{N_C \cdot (N_C - 1)}}$$

N_C = number of C-intervals in utterance
C_i = duration of the i[th] C-interval

3.3.2.5 Percentage over which speech is vocalic (%V)

The percentage over which speech is vocalic is (next to ΔC) the second major rhythmic correlate proposed by Ramus et al. (1999). It was calculated according to the following formula.

Equation 3-6: Percentage over which speech is vocalic

$$\%\mathrm{V} = \frac{\left(\sum_{i=1}^{N_V} V_i\right) \cdot 100}{\sum_{i=1}^{N_C} C_i + \sum_{i=1}^{N_V} V_i}$$

N_V = total number of sampled V-intervals
N_C = total number of sampled C-intervals
V_i = duration of the i^{th} V-interval
C_i = duration of the i^{th} C-interval

3.3.2.6 Pairwise variability index PVI

The Pairwise Variability Index (PVI) was calculated according to Grabe & Low (2002) as the raw index (rPVI) for C-interval variability and the normalized index (nPVI) for V-interval variability. The rPVI calculates the mean consecutive C-interval differences by summing up the difference between each consecutive C-interval (pairwise variability) and dividing this sum by the numbers of C-interval samples. The formula for this is given in the following:

Equation 3-7: Raw Pairwise Variability Index for C-intervals (rPVI)

$$\mathrm{rPVI} = \frac{\sum_{i=1}^{N_C - 1} \left| C_i - C_{i+1} \right|}{N_C - 1}$$

N_C = number of C-intervals sampled
C_i = duration of the i^{th} C-interval

The nPVI follows a similar algorithm like the rPVI but it was normalized for speech rate variation within and across sentences by dividing each pairwise V-interval difference additionally by the average duration of the V-interval pair (sum of a V-interval pair divided by 2).

Equation 3-8: Normalized Pairwise Variability Index for V-intervals

$$\text{nPVI} = 100 \cdot \frac{\sum_{i=1}^{N_V - 1} \left| \frac{V_i - V_{i+1}}{(V_i + V_{i+1})/2} \right|}{N_V - 1}$$

N_V = number of V-intervals sampled
V_i = duration of the i^{th} V-interval

It will be demonstrated that the rPVI strongly correlates with CV-rate (chapter 6). For this reason the same normalization procedure applied in case of the nPVI will also be applied for the rPVI. This measure will then be referred to as the nPVI-C (nPVI applied to C-intervals). The formula for the calculation of the nPVI-C is given in the following

Equation 3-9: Normalized PVI for C-intervals (nPVI-C)

$$\text{nPVI-C} = 100 \cdot \frac{\sum_{i=1}^{N_C - 1} \left| \frac{C_i - C_{i+1}}{(C_i + C_{i+1})/2} \right|}{N_C - 1}$$

N_C = number of C-intervals sampled
C_i = duration of the i^{th} C-interval

3.3.3. Units of analysis

Ramus et al. (1999) calculated ΔC and %V for each sentence and then took the average for each speaker and each language. Grabe & Low (2002) had their subjects read a longer text of a standard phonetic recording text (The North Wind and the Sun) and calculated nPVI and rPVI for the entire text. However, to test within-speaker variability of their results they broke the discourse of each speaker randomly into three parts of equal number of intervals (they did not detect within-speaker variability that would have affected the distinction of languages on the basis of their measures).

For the present study it was therefore questioned whether rhythmical and speech rate correlates should be calculated for the all intended tempo versions of each speaker

and should then be averaged for each language (Grabe & Low style) or whether each version should be broken up in underlying sentences first (Ramus et al. style). The Ramus et al. style has advantages for a number of statistical analysis procedures that will be carried out in the present study since it produces more values per speaker and thus leads to more reliable inferential tests.

Both versions (averaging per sentence and averaging per language) were calculated for the current data to see whether results would actually differ. Since sentences in the current data often contain a number of sub-clauses the sentences where split up into so called intonation phrases (henceforth simply: phrases). This basically means that a stretch of speech between two major discourse pauses was taken. Speakers may obviously vary in their production of intonation phrases; some speakers may produce a discourse pause when others do not. In addition to that speakers may not be consistent with placing intonation phrase boundary pauses especially when they speed up their reading speed. At higher speeds speakers typically produced the entire reading discourse with only a single or no pause at all. For this reason typical intonation phrases have been identified for each language at normal and slow speeds by calculating which inter-discourse-pause interval appears most frequently at normal intended reading speed. It was found that there is generally high between-speaker agreement for placing discourse pauses (roughly between 70% and 90%). It turned out that exactly 7 phrases could be identified per language which can be viewed in the listing of the reading texts for each language in APPENDIX II.

Coming back to the question about which stretch of speech the average should be calculated for, it was found that the average results for each language were close to being the exact same when they were calculated for each phrase first or when they were averaged for the entire productions (however, chapter 7 will show that there can be highly significant differences between individual phrases in each language). Because of the lack of difference, rhythm measures have been calculated for each of the identified phrases of each speaker in each language to increase the amount of detail for assessing within-language and within-subject variability.

For each speaker in the database the number of phrases is 35 (5 intended speech tempo versions * 7 phrases. Thus the number of units of analysis in each language (native languages only) is summarized in the following table:

Language	speakers	phrases
English	7	245
French	6	210
German	15	525
Italian	3	105
Czech	8	280
Total	**39**	**1365**

Table 3-3: Number of phrases used for the analysis of rhythm and speech rate correlates in the experimental chapters (4 to 9).

The numbers of phrases in Table 3-3 are the numbers of analysis units in each language for which (unless stated otherwise) correlates of speech rhythm (ΔC, %V, nPVI, rPVI, etc.) and speech rate (syllable rate or CV-rate; see 3.3.2) have been calculated in the experimental chapters to follow (4-9).

4. Rhythm measurements in normally produced speech

4.1. Introduction

One of the first steps which need to be undertaken is to view whether the data gathered in BonnTempo shows similar tendencies to the data obtained by Ramus et al. (1999) and Grabe & Low (2002) (henceforth often referred to as: comparison studies). Since these authors did not obtain any measurements under varying intended speech tempo, only the 'normal' intended tempo version in BonnTempo will be compared to findings of Ramus and Grabe & Low. This is the intended speech tempo version in BonnTempo that presumably comes closest in rate to the material of both studies. In the following the data will be cross-plotted in the same fashion as in Ramus et al. (1999) and Grabe & Low (2002). In analogy to Ramus et al. (1999) ANOVAS will be computed to investigate whether there are significant between rhythm-class differences.

In addition Ramus et al.' and Grabe & Low's analysis language differences will also be checked within and between rhythm classes to investigate how uniformly correlates of rhythm are within rhythmic classes. If features of linguistic rhythm are robust then these features should not vary dramatically between languages of the same rhythm class. In return the features should vary between individual languages of each rhythm class.

The languages under investigation for this chapter are languages that have traditionally been classified as stress-timed (English, German) or syllable-timed (French, Italian). Czech is further investigated as a not yet classified language. If rhythm classes can be viewed as two categories then Czech should be attributable to a single rhythm class on the basis of rhythmic acoustic measurements.

4.2. Method

A first step in the analysis of the data will be a comparison of the results for %V and

ΔC with findings by Ramus et al. (1999) and for nPVI and rPVI with Grabe & Low (2002). Methodologically this will be done descriptively by plotting the %V and ΔC as well as nPVI and rPVI along the two dimensions in the same manner as it has been done by Ramus et al. and Grabe & Low. The data plots will first be interpreted descriptively. It will be expected that the traditionally stress–timed languages English and German have a higher ΔC and a lower %V than the syllable-timed French and Italian (see rationale in the introduction, chapter 1, where this expectation can be viewed in Figure 1-2). For the Grabe & Low paradigm, nPVI and rPVI will be expected to be higher for stress-timed languages than for syllable-timed languages. For Czech, no expectation can be formulated yet. However, if Czech can be attributed to one particular rhythm class then different rhythm measures should consistently reveal similar results between Czech and the other languages in this class. After the descriptive analysis the results will be tested in the inferential domain according to the following procedures:

- [1] ANOVAS will be carried out with rhythm class as a two class factor and each of the rhythmic correlates (%V, ΔC, nPVI, rPVI) as dependent variable. This type of analysis is carried out by Ramus et al. (1999) in order to test whether the label 'rhythm class' successfully distinguishes the data between languages.
  The languages English and German will be grouped together to rhythm class 'stress-timed' (1), French and Italian to syllable-timed (2) and Czech will form a third rhythm class (3) in order to see which class this language is attributed to.
- [2] ANOVAS will further be carried with language as a five class factor and each of the rhythmic correlates as dependent variables.
  Two hypotheses will be tested with this analysis:
  (a) The hypothesis of 'within rhythm class similarity' of languages: This hypothesis assumes that comparisons between languages within a rhythmic class (English/German, French/Italia) should lead to non-significant results since these languages share the same rhythmic features.

(b) The hypothesis of 'between rhythm class differences' of languages:
This hypothesis assumes that language comparisons across rhythmic classes (English/French, English/Italian, German/French, and German/Italian) are statistically significant since the contrasted languages do not share the same rhythmic features.
For Czech it should be shown again that it consistently groups with the one or the other rhythmic class.

In the following the data in BonnTempo will first be compared to the comparison studies with the descriptive methods and inferential method [1]. This will lead to statements about whether the results of the comparison studies can be replicated with the data in BonnTempo. After that within and between rhythm class variability of languages will be assessed with inferential method [2].

4.3. Comparison of BonnTempo data and previous findings

4.3.1. Descriptive analysis

Figure 4-1 shows the results for ΔC and %V (left) as well as nPVI and rPVI (right) cross-plotted against each other. A comparison of this data with Ramus et al. (1999; henceforth: Ramus) shows that the general pattern of the ΔC/%V graph is replicated: Stress-timed languages (English and German) have higher ΔC and lower %V values than syllable-timed languages (French and Italian) (see Introduction for a detailed discussion of the rationale). Moreover, out of the five languages displayed here, three have been used by Ramus; they are English, French and Italian. For all languages Ramus received nearly identical values on the ΔC scale (note that ΔC has been multiplied by 100 in order to avoid fractional values so all values displayed here are higher by a factor of 100 compared to Ramus).

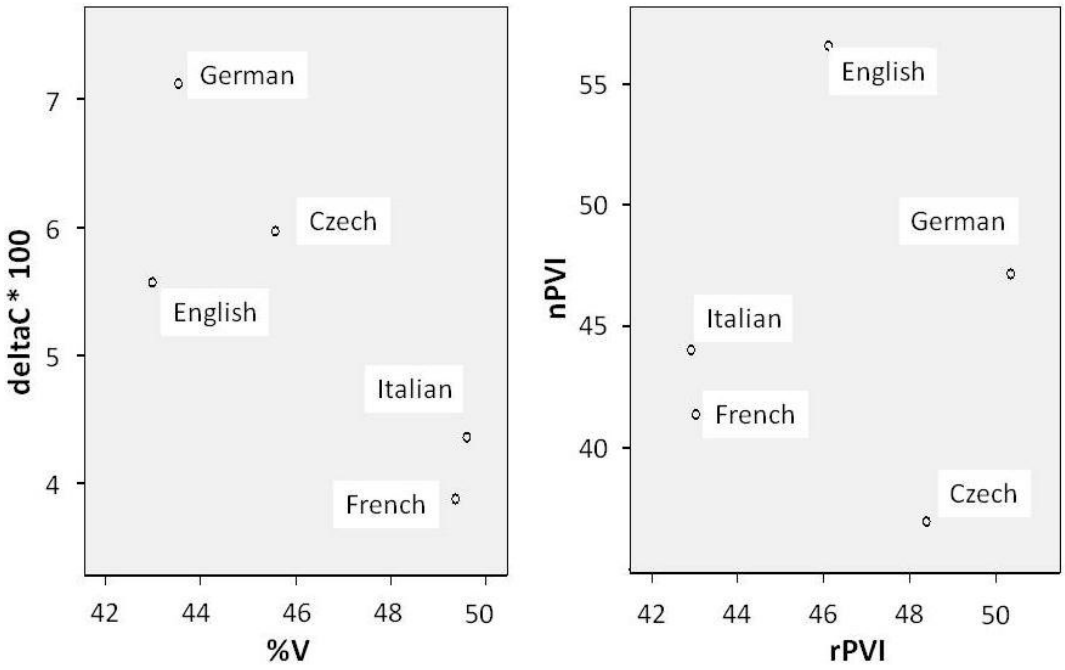

Figure 4-1: Scatter plots according to Ramus et al. (1999) measures ΔC and %V (left) and Grabe & Low (2002) measures nPVI and rPVI (right). The graphs display the native speaker data in BonnTempo for 'normal' intended speech tempo.

In the %V dimension all values are slightly higher here by about 3% compared to Ramus. It could be argued that this has to do with Ramus' method to control for speech rate in which sentences of similar duration and roughly equal numbers of syllables were used. However, it will be demonstrated in chapter 6 that %V is one of the most robust measures towards speech rate, so this assumption should not be followed any longer. The issue will be addressed in chapter 7 again where an analysis of factors other than speech rate will show that %V varies significantly according to phrase at different speech rates and it will be hypothesized that it was the choice of phrase that made the Ramus results vary from the ones presented here. All in all it can be concluded that there is overall a very high agreement between the Ramus data and the data in BonnTempo.

According to the dimensions nPVI and rPVI the languages under investigation also replicate the descriptive results obtained by Grabe & Low (2002; henceforth: Grabe & Low): the syllable-timed languages Italian and French both have a lower nPVI and

rPVI than stress-timed German and English. Note, however, that the languages German and Italian are very close together in the nPVI dimension. In fact, German is closer to Italian than it is to its stress-timed counterpart English.

In Grabe & Low only two languages under investigation are the same ones as used in the present study, English and French. Furthermore, the authors only used Southern Standard British English while the English in BonnTempo consists of different varieties of British and American English. It could thus be assumed that this pronunciation variability in the data between the two studies may lead to variations in the results. However, this assumption is not supported: both for French and English the values obtained by Grabe & Low match almost exactly the values from BonnTempo.

Czech, a not yet classified language, shows mixed results according to different rhythmic parameters. Along the ΔC axis it is between the two stress-timed languages. According to %V it is somewhere in the middle between stress- and syllable-timing. In the nPVI and rPVI dimensions, Czech is clearly below the syllable timed cluster according to nPVI and right within the stress-timed languages according to rPVI. In conclusion results show that rhythm measures do not agree on how to classify Czech and this is even true for rhythm measures belonging to the same paradigm.

4.3.2. Inferential tests for rhythm class differences

Rhythm class variability was tested with a one-way ANOVA with rhythm class as a fixed factor and the rhythm measures %V, ΔC, nPVI and rPVI as dependent variables. All F values (Fisher predictor) have a degrees of freedom ratio of 2:279 (numerator: 3 groups – 1; denominator: 280 phrases under normal intended speech tempo – 1). For all rhythm measures the p values for the ANOVA are smaller 0.001, i.e. the variation between the groups is statistically highly significant (F values: %V=28.7, ΔC=47.5, nPVI=24.8, rPVI=40.8).

The post-hoc test reveals that there is in general highly significant difference ($p<.001$) for group comparisons apart from the following instances: nPVI between Czech and syllable-timed languages shows a p value of .63; stress-timed class and Czech for %V, ΔC are at a slightly lower significance level ($p<.05$).

In conclusion, the results show that all rhythmic correlates are good indicators for

stress- and syllable-timing. The effect is highly significant. Czech, however, is not consistently grouped to either of the classes. While Czech is syllable-timed according to the nPVI, it cannot clearly be attributed to a rhythm class according to all other measures.

4.3.3. Conclusions about the data comparison

The comparison between the data from BonnTempo and Ramus as well as Grabe & Low has shown that there is very high agreement for the measures %V, ΔC, nPVI and rPVI for the languages that are shared between the present and the comparison studies. An ANOVA with rhythm class as a fixed factor (as carried out by Ramus) also replicates the finding that rhythmic classes vary with statistical significance for all four parameters under investigation. This is generally a good basis for the further analysis since it shows that the results are replicable for different datasets, even under slight methodological deviations.

4.4. Within- and between rhythm class variability

It is unlikely if not impossible that an ANOVA for a summary of languages into rhythm classes shows a statistically significant difference between the classes but not between any of the subgroups within the classes. An ANOVA with language as a fixed factor and the rhythm measures %V, ΔC, nPVI and rPVI as dependent variables confirms this and reveals p values smaller than .001 for each variable. (F[4-239] between 14.4 and 33.7, depending on dependent variable used). A post-hoc test (Tukey) shows more interesting results which can be viewed in Table 4-1. If languages were equally well contributing to rhythm class it should be found that within rhythm class contrasts between languages should lead to insignificant differences between the data distributions, while between rhythm class contrasts should turn out to be significant. The table shows that this assumption is only fully true for %V. Here contrasts between languages within a rhythm class show p-values close to 1 and between rhythm classes all significant levels are smaller than .001.

	within-class variability		between-class variability			
	English-German	French-Italian	English-French	English-Italian	German-French	German-Italian
%V	.98	1.0	<.001	<.001	<.001	<.001
deltaC	<.001	.83	<.001	.052	<.001	<.001
nPVI	<.001	.936	<.001	.002	.097	.843
rPVI	<.001	1.0	.007	.038	<.001	<.001

Table 4-1: P-values for an ANOVA post-hoc test (Tukey) with language as a four class fixed factor (English, German, French, Italian) and the Ramus and Grabe & Low rhythm measures as dependent variables.

ΔC distinguishes English from German which violates the hypothesis of 'within rhythm class equality' (see method above) and it also, although only marginally, does not distinguish between English and Italian which violates 'between rhythm class differences'. Both the nPVI and rPVI fail with a within-class separation of English and German. While the rPVI reaches at least significance levels with $p<.05$ for all between-class comparisons, the nPVI does not distinguish between German/French and German/Italian.

Results for a post-hoc test contrasting Czech with all other languages can be viewed in Table 4-2. The table reveals that there are no consistent significant differences between Czech and another language under investigation that would be supported the fact that Czech could be attributed uniformly to a single rhythm class. In most cases Czech turns out to be significantly different from other languages with a significance level of $p<.05$. No significant differences were found between Czech and English for %V and ΔC. According to %V Czech is not significantly different from German, and according to the nPVI it is not significantly different from French and Italian.

	Czech-English	Czech-German	Czech-French	Czech-Italian
%V	.09	.13	.005	.027
deltaC	.729	<.001	<.001	.002
nPVI	<.001	<.001	.414	.184
rPVI	.047	.037	<.001	<.001

Table 4-2: P-values for post-hoc comparison of an ANOVA with language as a five class fixed factor and the Ramus and Grabe & Low rhythm measures as dependent variables (%V, ΔC, nPVI and rPVI). The columns show the results for a comparison with Czech and all remaining languages under investigation (English, German, French, and Italian).

4.5. Discussion

In the present chapter it was demonstrated that the results for Ramus and Grabe & Low are well replicable with BonnTempo. It was demonstrated that absolute measurements as well as an inferential statistics investigation (ANOVA with rhythm class as a fixed factor) has lead to the same results (apart from %V where minimal variations were found). Results drawn from the BonnTempo corpus are therefore regarded as being well comparable for datasets from which the actual models for rhythmic measures were drawn. Thus the BonnTempo corpus serves as a solid data set for further tests on the relationship between rhythmic correlates and speech rate as they will be performed in the following experimental chapters 5 to 7.

In general it could be shown that once the languages English, German, French and Italian are grouped into respective stress and syllable-timed rhythmic classes, a statistically significant difference between the classes is observable. However, when languages are treated individually the rhythm class difference is not supported in a consistent way. Dependent on the measure, languages from the stress-timed class do then not necessarily vary any more from languages from the syllable-timed class. According to individual language variation there seems to be some random way according to which languages can sometimes be stress, and sometimes be syllable-timed. This notion of random classification is also supported when a not yet classified language, Czech, is being added. Czech is not clearly attributed to any of the rhythm classes (in a comparison between Czech and the rhythm classes),

however, one measures, the nPVI, puts Czech in with the syllable-timed languages. Comparing Czech to individual languages it was found that Czech can be similar to almost any of the languages depending on which rhythmic measure is chosen.

The results about Czech and the disagreement between rhythm measures are not surprising. Grabe & Low (2002) and Ramus (2003) have pointed out that there are inconsistencies in the way rhythm measures attribute individual languages to rhythmic classes and that not yet classified languages (Grabe & Low e.g. investigated Luxemburgish, Catalan and others) were not put into the same classes according to different measures. For the results provided in Grabe & Low and Ramus it can be well expected that equal within- and between-rhythmic class variability for individual languages is present. Although the authors do not provide a comparable analysis the descriptive data projected by the respective box-plots provides reason for the assumption that significance tests for languages from different rhythm classes would maybe not attribute them to different classes and that differences between languages within a class could be observed.

Another way to look at things may be to compare similarities in the outcome of measures that claim to monitor equal types of variability. In this respect the V-interval measures %V and nPVI claim to study durational variation of V-intervals resulting mainly from vocalic reduction. C-interval measures on the other hand are supposed to monitor durational C-interval variability caused by consonantal complexity within a syllable. Thus it should be worth to look at similarities between %V and nPVI as well as ΔC and rPVI. But also here we find that the measures do not agree. While %V fulfills the expected within- and between-class differentiations between languages the nPVI does not and also ΔC and rPVI lead to different classifications.

The only measure that fulfills the expectations according to within and between-class variability for languages is %V. Within-class %V does not distinguish between the pairs English/German and French/Italian while it clearly distinguishes in all between-class comparisons. Does this mean that %V is necessarily the best indicator for rhythm class? This question is difficult to answer, in particular in respect to the current data situation. Rhythm classes are here only represented by two languages each and it may be that the two languages that were picked for each class happen to

be languages that %V can well monitor. In the descriptive data provided by Ramus et al (1999) English is for example the language with the lowest %V value. If, for the present study, stress-timed languages at the higher end of the %V scale had been included (e.g. Dutch, according to Ramus' data) and syllable-timed languages at the lower end of the scale (e.g. Spanish) we would probably find that even %V is not able to distinguish these languages any more.

The discussion shows that the way rhythmic correlates reveal rhythmic features of different languages does not support the concept of rhythm class well. The fact that languages within a class may reveal highly significant differences between each other leads to the assumption that different auditory groupings could have been supported in the similar ways. Since German was shown to be significantly different from English in respect to nPVI, rPVI and ΔC, a grouping of presumably stress-timed English along with presumably syllable-timed French and Italian would probably have led to a significant distinction between the group English/French/Italian and the group German.

However, Dauer (1987) already proposed a model that moves away from a strict categorical distinction to a continuous variability of rhythmical features between a stress- and a syllable-timed extreme. The data from the above analyses fits much better into such a model. In case of the consonantal rhythmic correlates, ΔC and rPVI, German would form an extreme stress-timed and French an extreme syllable-timed point. For German and English this point is clearly reversed in the case of nPVI in which case English forms the most extreme point. So the data presented here is more in accordance with a model suggesting the continuous variability of rhythmic features between languages while measures may vary in the way they attribute languages on this scale.

One reason for the variability of the data, however, that has not been taken into consideration is the fact that some of the rhythm measures may interact with speech rate. After all, the basic units of rhythm measures are the durations of C- and V-intervals, and the durations of these intervals vary as an effect of an increase or decrease in rate. It is not necessarily the case that there is a linear proportional increase or decrease according to rate (Dauer, 1987). It may well be that vocalic units underlie de/compression mechanisms that are widely different from consonantal

units (Dauer, 1987, Grabe & Low, 2002). It may further be that consonantal units in themselves show a wide variability since consonantal units vary widely in quantity within the consonantal class (Steiner, 2004). Another point that has been put forward in Dellwo (2006) is the fact that standard deviations may be mean value dependent. Thus values like ΔC that are the absolute standard deviation of C-intervals may well be different for slow compared to fast speech. Although 'normal' intended speech tempo was analyzed in this chapter it may be that speech rate varies considerably between speakers and between different phrases. The way speech rate varies across normal and other intended speech tempi will therefore be investigated in the next chapter (5). In chapter 6 we will look at how rhythm measures are affected by speech rate influences.

A second point that will shed light on the question which of the variability measures most reliably points out rhythmical differences between languages is by testing the measures perceptively. Since perception tests in the rhythm domain are not unproblematic (reasons for this will be discussed in the respective chapters) we will gather evidence from delexicalized speech (chapter 7) and from the second language acquisition domain (chapter 8) in order to see which measures can be distinguished in the perceptual domain.

4.6. Conclusions about rhythm measures in normal speech

The present chapter has revealed that the findings of the two key studies in contemporary rhythm research, Ramus et al. (1999) and Grabe & Low (2002), can be replicated well with the BonnTempo data set. For this reason it can be assumed that the present data set, BonnTempo-Corpus, well represents the nature of the data in previous studies and is thus suitable for studying the effects of speech rate on speech rhythm measures in the following chapters.

It was further found in the chapter that the rhythm measures are reliable indicators to point out language individual differences rather than rhythmic class differences. In terms of rhythm class the data suggests that a model of continuous differences between the rhythmic extremes stress-timed/syllable-timed is more supported by the data than a dichotomous rhythm class model.

5. Speech Rate, language, and rhythm class

Speech rate has an effect on the segmental (segment duration, segment quality) and suprasegmental (intonation, intensity, etc.) domain. On a suprasegmental level the effects of speech rate on duration have been the main focus of research (Bartkova, 1991; Keller and Zellner, 1995). Investigations on the effect of speech rate on intonation (for Dutch: Caspers & van Heuven, 1995; for French: Vaissière, 1983; Fougeron & Jun, 1998; for German: Kohler, 1983, Trouvain & Grice, 1999) have shown that a higher speech rate leads to an elimination of phrase boundaries. Further it has been shown that variation of speech rate has an effect on fundamental frequency maxima and minima as well as the fundamental frequency contour. However, these parameters vary in experimental studies to such an extent that global conclusions are not possible at the current stage.

A general argument that underlies recent rhythm class measures as proposed by Ramus et al. (1999) and Grabe & Low (2002) is the observation that languages traditionally classified as stress-timed show a more complex syllable structure than languages claimed to be syllable-timed (see chapter 1 as well as Dasher & Bolinger, 1982, Bolinger, 1975, 1981, Dauer, 1983, 1987, Ramus et al., 1999, Grabe & Low, 2002). For the languages under investigation in the present study it was found (chapter 3) that they represent these general characteristics well. The experimental speech material for French and Italian was demonstrated to be phonotactically less complex on different levels (syllable complexity, C-interval complexity) than English and German. This difference, however, gives rise to the assumption that languages also differ in terms of the rates at which they produce syllabic elements (or in the current context: C- and V-intervals).

Since correlates of linguistic rhythm in the speech signal as analyzed in the previous chapter are based on the durational variability of C- and V-intervals it can be assumed that speech rate has a large effect on these durations and that this effect is a) not necessarily linear and b) may vary between languages (see ‘rhythm class hypothesis’, 1.3). Further, in order to receive a wide range of different speech rates in

each language, listeners were asked to produce speech under different intended speech tempi (very slow, slow, normal, fast and as fast as possible). So far it is unclear what effect this intended tempo variability actually had on measurable speech rate. Again, these effects may have varied between languages thus it may be that intended speech tempo variability had bigger effects on one language than it had on others.

For the stated reasons within- and between-language speech rate characteristics will be analyzed further in the present chapter. The languages under investigation are the same ones as in the previous chapter: English, French, German, Italian and Czech. In the next section (5.2) an appropriate unit for the analysis of speech rate in connection with speech rhythm will first be developed. After that (5.3) it will be analyzed what effect the intended tempo changes had on laboratory measurable rate. In 5.4 it will be tested whether the effect of rate change was equal in all languages.

5.1. An appropriate unit for rate measurements

Laboratory measurable speech rate (lsr) can be defined as the rate of a unit of speech per second. Experimental research varies widely in respect to what speech unit is chosen (see Dankovičová, 2001). Research also varies in the terminology of how to refer to rate measurements (see Laver, 1994). While the term 'speech rate' is often used to characterize measures including pauses, 'articulation rate' is used for measures monitoring rates during which the articulators are moving, i.e. excluding pauses (see Trouvain & Grice, 1999, for a discussion). In the current work this distinction is not made since the overall aim of the study is to investigate relationships between speech rate and rhythm measures based on durational variability of C- and V-intervals. Although it could be argued that pauses may play a role in the overall rhythm of speech, rhythm models based on C- or V-interval measures do not incorporate pauses yet. The relationship between rate and rhythm in speech is therefore manifested during filled speech only and not during silent passages (the only exception is voice onset time in plosive consonants). For this reason lsr in the present context refers to speech rate measures which exclude pauses.

In studies on speech rate typically syllables per second are used, although sometimes

phones or words are also common units. In language contrastive studies it is of particular importance to choose units that reveal language characteristic differences. Such differences will be hard to find on the level of a phone but they will manifest on a syllabic level since the complexity of syllables varies between languages and thus their durations. Measuring a phone rate would possibly lead to only marginal differences between languages.

As a contrast to global measures of speech rate Pfitzinger (1999) introduces the measure of 'local speech rate' combining syllable rate, phone rate and f0-movement. Pfitzinger is able to demonstrate a high correlation between laboratory measurable local speech rate and perceptual rate. A disadvantage of this rate measure (which is also the reason why it is not applied in the current study) is that the weighting of individual 'local speech rate' parameters varies between languages and has so far only been evaluated for German. So in order to perform a cross language comparison a considerable amount of adaptation work has to be done beforehand which would go beyond the scope of this study.

The reasons for which unit to choose for speech rate measurements are manifold and vary from case to case. In the case of rhythm studies based on C- or V-intervals it is questionable and, as will be shown below, misleading to use the syllable as the unit for measuring speech rate. For this reason a rate measure will be introduced here that is based on the sum of the total number of C- and V-intervals per second and thus monitors the rate of units that are the actual base constituents of rhythmic measures (see 3.3.2 for formulas).

But why is it important to shift the unit from syllables to C- and V-intervals per second? The main reason is to ensure that the units for rhythmic and speech rate analysis are the same. This is especially of importance when it comes to speech rate normalization. Ramus et al. (1999), for example, normalized their data by selecting sentences of roughly equal duration (about 3 seconds) and roughly equal number of syllables (15 to 19). With this technique they intended to select sentences that were matched in speech rate across different languages to eliminate possible speech rate artifacts in their results. Such a technique, however, assumes that the ratio 'number of syllables: number of C- and V-intervals' is equal across languages. If this was not the case then the rate of C- and V-intervals may vary between languages even when

syllable rate is identical. However, since C- and V-intervals are the actual units of rhythmic investigations, rate normalization in respect to rhythm measurements is only given when the sum of the number of C- and V-intervals per second is roughly the same. In the next section the syllable: C- and V-interval ratio will be analyzed across languages.

5.1.1. A comparison between CV- and syllable-rates

An important issue for the present work is the relationship between the average numbers of C- and V-intervals per syllable that a language allows. In chapter 3 it was demonstrated that syllable timed languages tend to have more simple syllables with the segmental characteristics v and cv while stress-timed languages also possess complex types like cvc, cvcc, etc. This means that in stress-timed languages a total number of two C-intervals per syllable may be more common than in syllable-timed languages. Consider the following two hypothetical examples, both consisting of a phrase with 8 syllables which are made up out of (a) non-complex syllable types and (b) complex syllable types. The following gives two examples, one for (a), one for (b). The syllables are separated by a blank and the segmental structure within the syllable is characterized by the letters ‘c’ (consonantal segment) and ‘v’ (vocalic segment):

(a) cv cv cv cv cv cv cv cv

(b) cvcc cvc cv cvc ccvc vc cvc cv

It is not possible to increase the number of C- and V-intervals per syllable higher than 1:2 (syllable: C- and V-interval ratio), however, the numbers of C- and V-intervals can be drastically reduced when syllables consisting of a single vowel (v) come into play. The following example from French shows the syllable types occurring in the French phrase 'je me suis rendue à Albi' (taken from the French reading material, see Appendix II):

(c) cv cv cv cvc cv v vc cv

The above phrase counts 8 syllables but only 6 C- and 6 V-intervals, i.e. a syllable: C- and V-interval ratio of 8:12 or 1:1.5. The example demonstrates that the syllable: C- and V-interval ratios can vary substantially as an effect of single V syllable type usage in the environment of a preceding syllable of the structure ‘cv’ and a following

one of the structure 'vc'. Assuming that all syllables in the above examples have the same duration then a speech rate normalization based on syllables/second would lead to the result that all three phrases (a, b, c) are of equal 'speech rate'. Nevertheless, in terms of rhythmic units we have a considerably lower rate of C- and V-intervals in example (c) compared to (a) and (b).

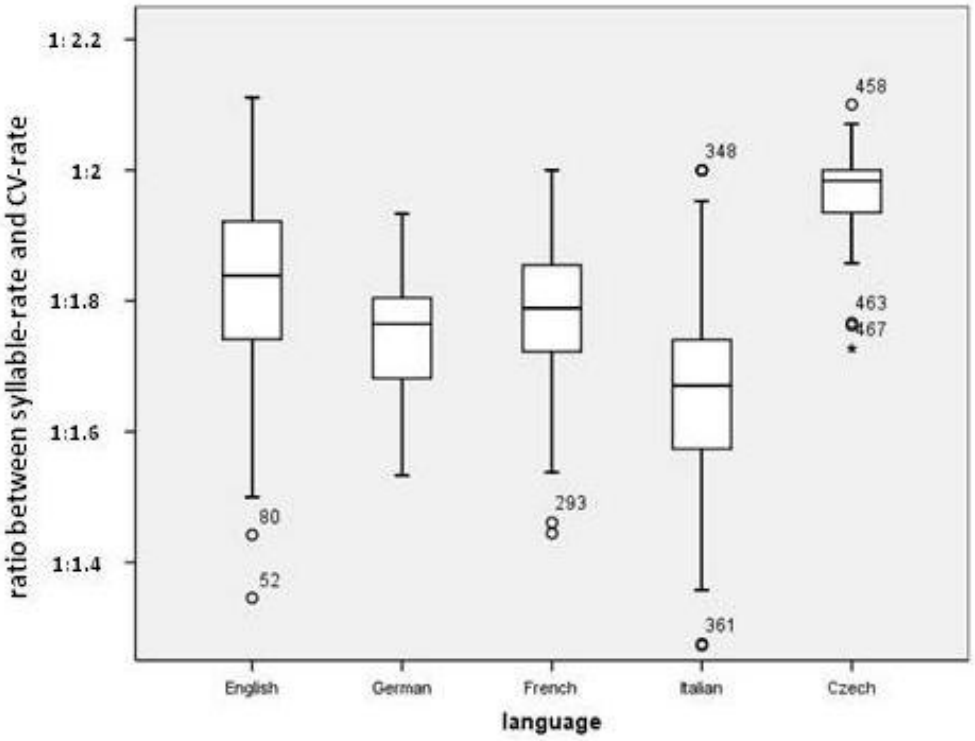

Figure 5-1: Ratio between syllable- and CV-rate for the languages English (1), German (2), French (3), Italian (4), and Czech (5) at normal intended speech tempo.

Since single vowels are more common in syllable- than in stress-timed languages (see chapter 3) we may assume that examples of type (c) are more common in syllable-timed languages. Figure 5-1 shows that this assumption is not correct. The box-plot shows the ratio between syllables and C- and V-intervals as a function of the five languages under investigation. It is apparent from the distribution that languages have different ratios between syllables and C- or V-intervals. Czech seems to be the most consistent in terms of altering a C- with a V-interval in each syllable.

All ratios in this language are distributed very closely around 2. All other languages vary considerably more. The full range of ratios that English allows for example is from about 1.5 to 2.1 C- and V-intervals per syllable.

An ANOVA with language as a fixed factor and syllable: C- and V-intervals ratio as the dependent variable reveals that there is significant variability between the distributions ($F[4,1399]=27.97$, $p<0.001$). A further post-hoc test reveals that the variation between the pairs English and French as well as German and French are not significant ($p=.168$ and .336 respectively). All remaining group variability is highly significant ($p<.001$).

The important message we receive from these examples is that languages possibly vary significantly in the numbers of C- and V-intervals per syllable and that this variability needs to be taken into account when rhythm measures are to be normalized for speech rate. The most straightforward way to take this into account is probably by using C- and V-intervals as opposed to syllables for rate measurements. This step seems plausible also in respect to the claim that C- and V-intervals are perceptually the more salient units in terms of speech interval variability than syllables (Ramus et al., 1999). If C- and V-intervals are salient units for rhythm then it seems conceivable that they can also be salient units for rate. Although this hypothesis will not be tested in the present work it will be used as an additional justification to shift the rate measurement unit from syllables to C- and V-intervals. In cases where speech is delexicalized to preserve durational information about C- and V-intervals only (chapter 8 and Ramus et al., 1999) the only possible rate indicator is the rate of C- and V-intervals since acoustic cues to syllable boundaries have been removed (unless they co-occur with C- or V-interval boundaries). In such cases the measurement of C- and V-interval rates instead of syllable rates is essential.

5.1.2. Conclusions about CV and syllable rates

In the present section it was demonstrated that the ratio between C- and V-interval and syllable rates varies significantly as a function of language. This is important for certain types of speech rate normalizations in which sentence material of equal number of syllables at equal durations are compared between languages (e.g. Ramus et al, 1999). This means that similar numbers of syllables but different numbers of C-

and V-intervals may be compared. The effects of such a comparison will be discussed further below in a closer evaluation of such rate normalization methods.
Since the present work aims to study relationships between speech rate and rhythmic measures based on C- and V-interval variability the rate of C- and V-intervals per second will be used as a speech rate measure. With this method the actual rate of C- and V-intervals can be better controlled for.

5.2. Intended and measurable speech rate

Above we demonstrated that there is a difference between C- and V-rate and syllable-rate and that it can be crucial when it comes to speech rate normalization to differentiate between these two types of rate. In this section it will be demonstrated what effect the intended decrease and increase of speech tempo had on CV-rate.
It may appear straightforward to assume that a deliberate increase and decrease in speech tempo goes along with a measurable in- and decrease in the rate of units. However, speakers themselves often state in informal questionings after the recordings that they merely increased the number of pauses when they slowed down speech rate. The results of this section will show whether this self-assessment of speakers is correct. Moreover, statements like that give rise to the assumption that there is a non-linear in/decrease in speech that is triggered by intentionally speaking slower and faster. Last but not least languages may vary in the way they allow speech rate to be increased. This is an important point to consider in respect to one of the overall aims of this research which is to investigate the influence of speech rate on rhythmic measures. If rhythm measures should turn out to be affected by rate and languages differed in the way they allow rate changes then rhythmic measurements in languages that allow fewer rate changes may be less affected by rate.

5.2.1. Changes in CV-rate as a function of intended speech tempo

5.2.1.1 Comparison of CV-rate between intended tempo categories

Figure 5-2 shows the distribution of CV-rate (the sum of the number of C- and V-intervals per second) for all five intended speech tempo categories. It is obvious from the boxes that there is a considerable increase in CV-rate as a function of intended

speech tempo. An ANOVA with intended speech tempo as a five class factor reveals highly significant differences between all distributions ($F[4,1399]=555.81$; $p<.001$). All post-hoc comparisons (Tukey) for the 'language' factor show highly significant differences between all groups ($p<.001$). These results clearly show that there is a significant overall increase in CV-rate from the slowest to the fastest intended tempo.

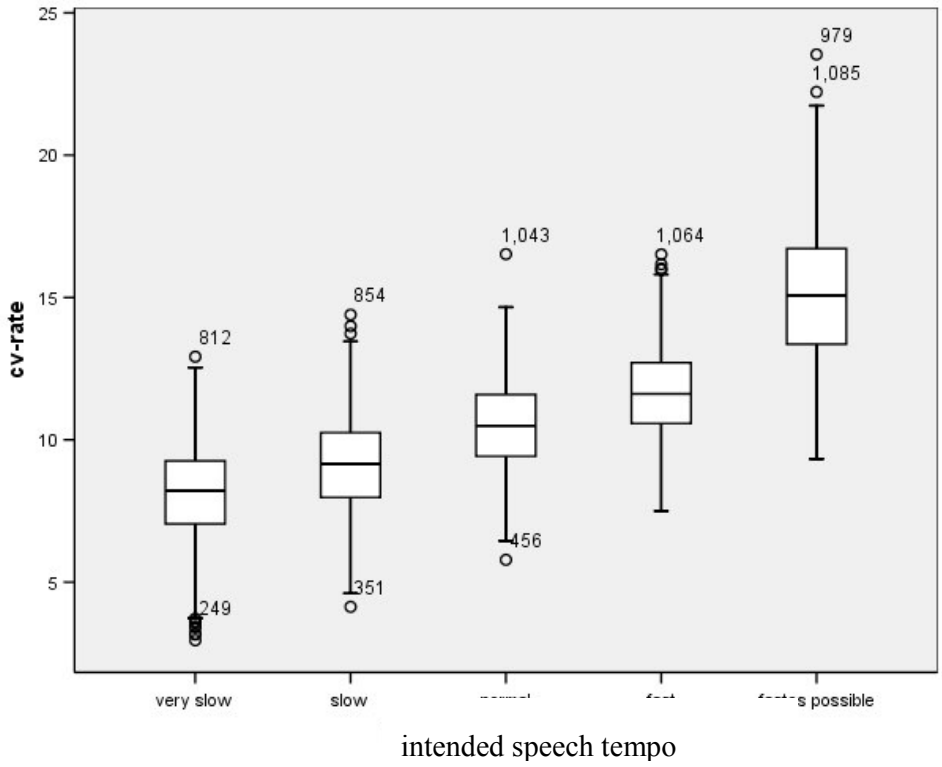

Figure 5-2: Box-plot showing the distribution of CV-rate at each of the five intended speech tempo categories for all languages under investigation (Czech, English, French, German, and Italian).

Comparing the boxes in the box plot with each other, however, it is also obvious that there is considerable overlap between the distributions. Apart from the fastest intended tempo category the second quartile of each distribution overlaps considerably with the third quartile of the next slower intended tempo class. These results suggest that a great amount of measurable speech rates cannot be uniquely attributed to a particular intended tempo category.

5.2.1.2 Variability between intended tempo categories

How safely intended tempo can be predicted from CV-rate was tested with a Discriminant analysis. The results of the analysis with the five intended tempo categories as a grouping variable and CV-rate as an independent variable confirm the view that CV-rates overlap highly between intended tempo categories (see results in Table 5-1). The highest classification scores are reached for the fringe classes (very slow [1] and fastest possible [5]). However, only 27, 30, and 45 % or correct classifications are reached for distributions in the classes 2, 3, 4 respectively. Moreover, a considerably larger amount of data (38.6 as opposed to 27.5 %) from the slow intended tempo category (2) is attributed to category 'very slow' (1).

		Predicted Group Membership (%)				
	isr	1	2	3	4	5
Presented Group Membership	1	56.8	29.6	7.9	5.7	.0
	2	38.6	27.5	21.8	11.1	1.1
	3	14.3	20.0	30.0	31.1	4.6
	4	2.1	12.1	22.9	45.4	17.5
	5	.0	.4	3.2	22.5	73.9

Table 5-1: Summary table of a Discriminant Analysis with intended speech tempo as a grouping variable and CV-rate as an independent variable. Columns show the predicted group membership of each data point for each intended tempo categories (rows).

All above analysis (ANOVA for intended speech tempo categories and Discriminant Analysis) was performed for individual languages and results resemble the general output. All languages showed highly significant increases in CV-rate as a function of intended speech tempi and classification results for the Discriminant Analysis are equally poor (a display of all data for five languages was found to be too space consuming and would in itself be redundant).

5.2.1.3 Conclusions about CV-rate variation according to intended speech tempo

The results presented above show that there is considerable evidence for the view that CV-rates decrease when speakers intend to speak slower and increase when they intend to speak faster. This is also true between the slower intended tempo versions where speakers' typical introspective observation was that the rate decrease from 'slow' to 'very slow' was produced by filling in pauses rather than changing word or segment durations. We can therefore safely conclude that the methodological design of the present study to ask subjects to vary their rate lead to a significant in/decrease in rate in all languages. With this methodology we received a wide variety of speech rates in each language which we will use in return to test for a possible interaction between rate and rhythmic measures (chapter 6).

5.2.2. Differences within and between languages

In the previous section it was found that CV-rates increase as a function of intended tempo. It will now be investigated whether languages differ in CV-rates obtainable at each intended tempo. Figure 5-3 plots laboratory measurable speech rate as the sum of the number of C- and V-intervals per second (CV-rate) averaged across inter-pause-intervals for all five intended speech tempo categories (very slow, slow, normal, fast, and fastest possible) for the languages English and German (representing stress-timing), French and Italian (representing syllable-timing) and Czech (auditory classification ambiguous). Regarding our hypothesis the descriptive data suggests that there is a general increase in CV-rate from the slowest to the fastest version. This increase is roughly proportional for all languages at all intended speech tempo categories apart from the fastest category for English where the language seems to put a constraint on tempo increase.

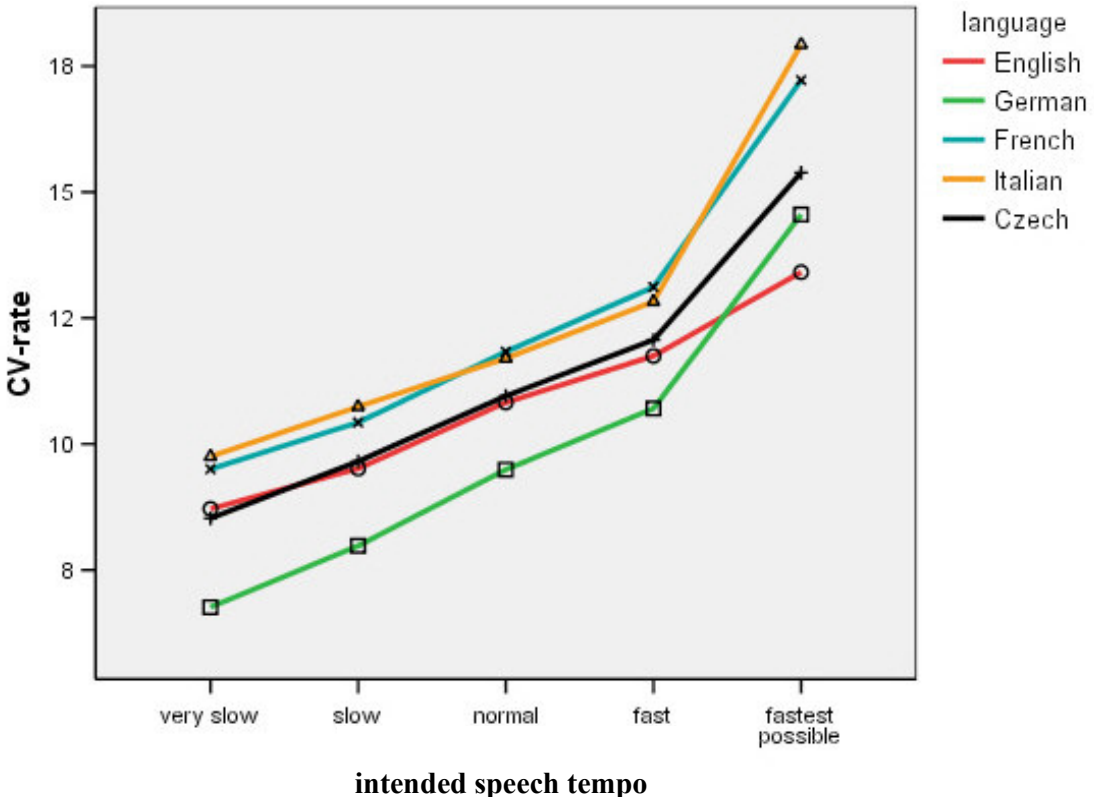

Figure 5-3: CV-rate (sum of the number of C- and V-intervals per second) averaged across inter-pause-intervals at the five different intended speech tempi for the languages Czech (Cc), English (Ee), French (Ff), German (Dd), and Italian (Ii).

The graph supports the assumption that some languages are produced faster in terms of CV-rate than others. Both French and Italian show the highest (and equal) rates at all intended speech tempi. German is clearly at the lowest extreme of the scale with the slowest CV-rates at all intended speech tempo categories apart from the fastest possible where English is slightly slower. English and Czech show rather equal CV-rates and lie somewhere in the middle between the French/Italian upper and the German lower extremes. In terms of rhythmic class distinction the two syllable-timed examples French and Italian seem to cluster well, however a large difference is detectable between English and German.

5.2.2.1 Method

By means of a repeated measure analysis with 'intended speech tempo' as a five level within-subject variable and 'language' as a five level between-subject factor it was tested whether languages differ in measurable speech rates (CV-rate) at different intended speech tempi (between-subject factors or interaction effect). The within-subject variability results may additionally confirm the analysis from above stating that there is significant variability in CV-rate between intended tempi.

The analysis was based on CV-rates averaged over the number of phrases of each subject. The design for repeated measures requires the cases of each factor level (here: intended speech tempo class) to be a separate variable (tempo1, tempo2, etc.). Cases will only be considered for repeated measures if they exist at equal numbers in all variables. For this reason average values for each phrase have been chosen since it could be assured that each phrase would occur as a case in all five intended tempo classes. The number of cases entered for each language was thus the number of speakers times 7 phrases (reading material for all languages was subdivided into 7 phrases in each language). So the total number of cases entered into repeated measure were 49 for English (7 speakers * 7 phrases), 105 for German (15*7), 42 for French (6*7), 21 for Italian (3*7) and 56 for Czech (8*7). Repeated measures analysis is robust against little variations from normal distribution, however, distributions for CV-rate was tested for each language at all intended speech tempo levels and was found to be close to the Gaussian distribution. Further, repeated measures does not require the between-subject factors to be of equal numbers, so the varying number of cases for each language have no influence on the results.

5.2.2.2 Results for within-subject variability

The Mauchly's test revealed a highly significant effect of sphericity for the within-subject analysis which is why a Huynh-Feldt correction was applied to all following data analyses. Within-subject variability is highly significant ($F=229.65$, $p<.001$) which indicates that the increase in CV-rate as a function of the 5 intended speech tempi observable in Figure 5-3 is a statistically recognized effect. Interaction between intended speech tempo and language are significant at a $p<.05$ level ($F=$

2.383). This is evidence for the view that languages do vary in the way they increase or decrease speech rate across the 5 intended speech tempi.

A post-hoc test reveals that French and Italian do not interact with each other (i.e. are not significantly different from each other; p=1.0), however, they do interact with all other languages (p<.001). English interacts with all languages (p<.001) but Czech (p=.993). German interacts with all other languages (p<.001).

5.2.2.3 Results for between-subject variability

Since English apparently lacks behind the other languages at higher speech rates some of the between subject interactions may be an effect of the cross-over between English and German between the fast and the fastest possible intended speech tempo classes. For this reason two further repeated measures have been conducted, one which takes the first 4 intended speech tempo classes as within-subject variables (very slow to fast) and one taking the last two intended tempo classes (fast and fastest possible). Within subject variability stays equally significant as above for both analysis (p<.001), however, the post-hoc test for the between-subject factor, language, changes slightly. There is stronger support for no interaction between Czech and English (p=.993) in the repeated measure analysis using the first two intended speech tempo categories. In the repeated measure analysis using the final two intended speech tempo categories an interaction is again present (p<.001). However, in the latter analysis English and German show no longer an interaction effect (p=1.0). All other interactions remain the same as above.

5.2.2.4 Discussion and conclusions of repeated measures

The results for within-subject variability of the repeated measure analysis confirm the results from the previous section that CV-rate varies significantly between the intended tempo categories.

The between-subject results reveal that languages interact differently with each other. French and Italian have been found to be equal in speech rate at all five intended tempo categories. English and Czech are equal in rates between the slowest and fast intended tempo categories but not between fast and fastest. Between the latter two

categories there is a cross-over between German and English.

The pattern that the traditional stress-timed languages (English, German) and the syllable-timed languages (French, Italian) reveal is very similar to what was found in chapter 4 concerning some rhythm measures. French and Italian share clear similarities on a higher CV-rate level while English and German seem considerably different from each other but are still both considerably slower than French and Italian. It may therefore be that CV-rate is an equally reliable rhythm class predictor as some of the C- or V-interval variability measures discussed in chapter 4. This hypothesis will be tested below at all different intended tempo categories with an ANOVA using rhythm class as a factor on CV-rate as the dependent variable (in analogy to ANOVA comparisons for C- and V-variability measures performed in chapter 4).

Another issue that will be tested is the question about proportional rate differences within a language. Although it was demonstrated that English for example does not reach the same CV-rate at the fastest intended tempo category this may well be an effect of the averaging that has been applied.

5.3. Differences in speech rate ranges

In order to test whether languages allow different rate ranges the average CV-rate of each percentile of the CV-rate was plotted for each language. The result of this analysis can be viewed in Figure 5-4. The analysis makes the proportional increase in CV-rates between languages directly comparable. The graph shows again that French and Italian are the two fastest languages at each percentile, German is the slowest at most percentiles and Czech is somewhere in the middle. English is the most interesting language of this constellation. While it could be argued that the total range differences in French, Italian, Czech, and German are more or less proportional the range differences in English are visibly smaller. English starts off with rates that are slightly higher than Czech and close to French and Italian at the lower percentiles, it is equal with Czech up to about the 50^{th} percentile, and it drops below all languages after the 90^{th} percentile. This analysis shows us that there are fewer proportional rate changes happening in English compared to all other languages.

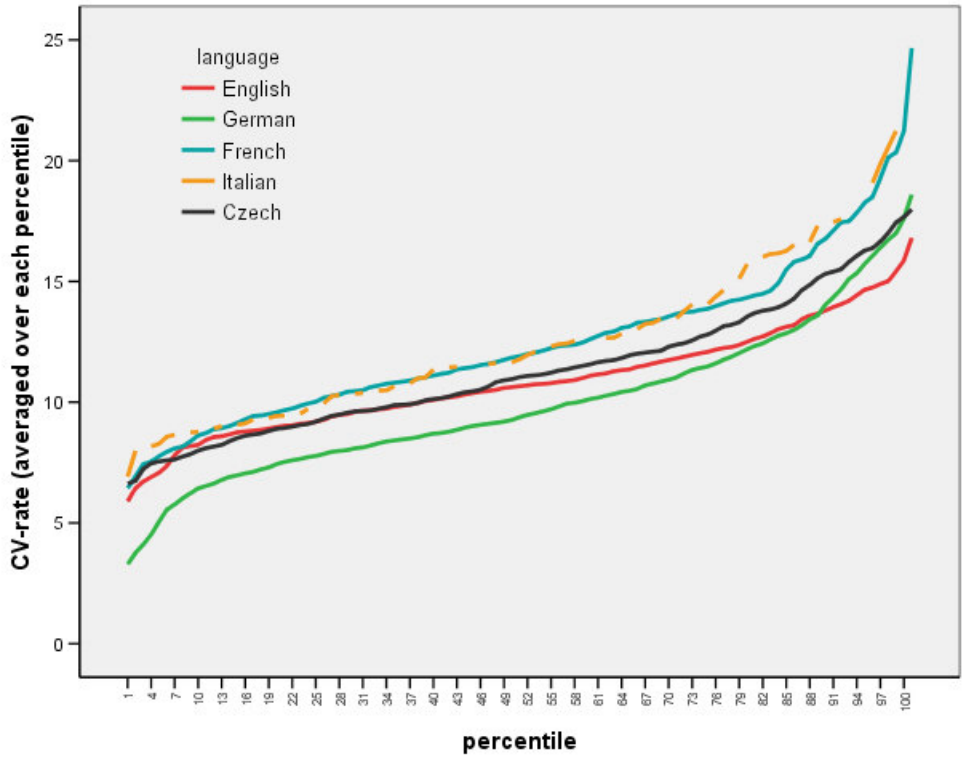

Figure 5-4: Mean CV-rate at each percentile (form 1st to 100th percentile) of the CV-rate distribution for the languages German, English, Czech, French, and Italian.

5.4. CV-rate and rhythm class

Above it has been demonstrated that languages traditionally classified as stress-timed (English, German) and syllable-timed (French, Italian) share similar speech rate features. This has been found especially true for the syllable-timed languages that are not separable from each other at all intended tempo categories.

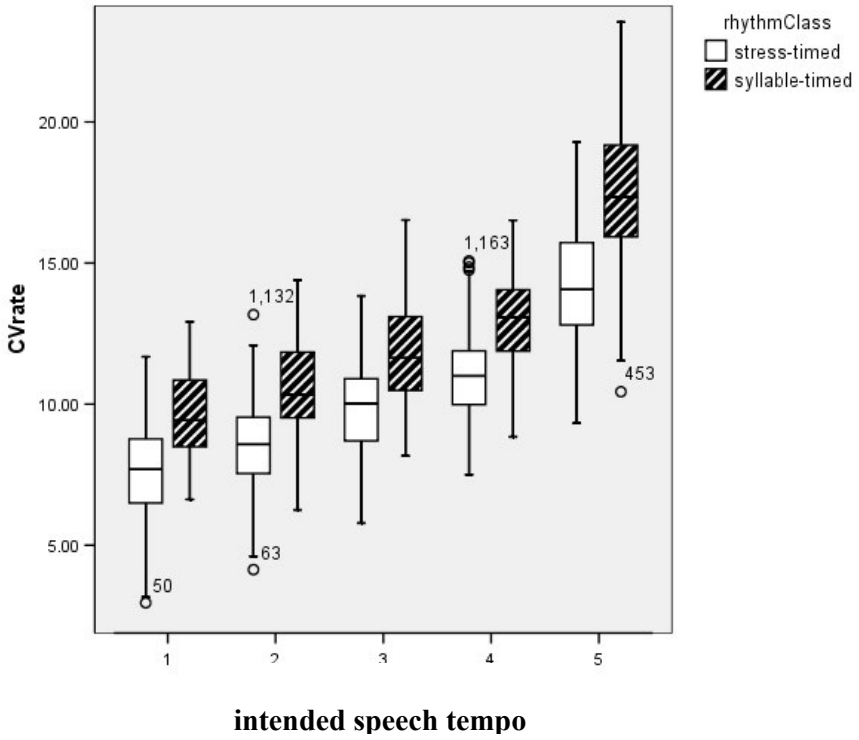

Figure 5-5: Box-plot showing the distribution of CV-rate (sum of the number of C- and V-intervals per second) for stress- (English, German) and syllable-timed languages (French, Italian) at all 5 intended speech tempo classes (1= very slow, 2= slow, 3 = normal, 4 = fast, 5 = fastest possible).

Figure 5-5 shows the distributions of CV-rates when stress- and syllable-timed languages are grouped together. The box-plot reveals that stress-timed languages are considerably slower in CV-rate at each intended tempo category compared to syllable-timed languages. Five independent-samples t-tests have been performed to check for statistically significant variability of CV-rate (dependent variable) between the stress- and syllable-timed classes (grouping variable) at each intended tempo category. Highly significant variability can be confirmed; all p values are smaller .001 (degrees of freedom = 215). Judged by the results it can be safely stated that CV-rate is a reliable predictor for rhythm class.

5.5. Discussion and conclusions on speech rate and rhythm class

Figure 5-5 and the results from the accompanying t-tests can probably be regarded as one of the most important findings at this stage since they show that speech rate is as reliable a predictor for rhythm class as all measures based on durational C- and V-interval variability presented in chapter 4.

In respect to differences between individual languages it was found that French and Italian show equal rates which are higher than the rates produced by English and Czech speakers which in return are higher than German. The reasons for why CV-rates in French and Italian are higher than in English, Czech and German are probably directly related to the syllable complexity characteristics which form the main rationale for the rhythmic measures: stress-timed languages are characterized by more complex syllables than syllable-timed languages (see chapter 1 for a discussion). In chapter 3 an analysis of the syllable complexity for the languages under investigation has shown that this concept is generally correct for the languages under investigation (see chapter 3 for results and discussion on syllable and C-interval complexity). It was pointed out that for the experimental reading text in BonnTempo the German syllable and C-interval structure shows more complex features than the English one. French and Italian, however, are rather similar. Czech is somewhere between English and the French/Italian complexity. But these results do not only directly reflect the pattern for speech rate differences between languages; they also reflect the pattern for the durational C-interval variability in chapter 4. There the C-interval variability could be demonstrated highest for German and lowest for French and Italian which do not differ in C-interval variability. Czech and English were shown to be somewhere between these two extremes. So it seems that consonantal complexity is the driving factor for both speech rate and durational C-interval variability.

Another assumption that arises from this data situation is that speech rate and the consonantal correlates of linguistic rhythm are simply related. Dellwo (2006) has raised the assumption that since ΔC measures the absolute standard deviation it may be strongly dependent on the mean value for C-intervals. Positive correlations between standard deviations and mean values are typical in empirical data and can be

assumed for C-intervals too. Relatively short C-intervals in fast speech are unlikely to have the same absolute standard deviation as comparably long C-intervals in slow speech. For this reason it may be assumed that the results for the C-interval variability measures in chapter 4 are strongly driven by speech rate in the way that French and Italian has less C-interval variability because the underlying intervals are shorter than in English and Czech which in return have shorter C-intervals than German. In the next chapter it will therefore be tested whether the C-interval measures ΔC and rPVI do still reveal between-language variability when speech rate is normalized for.

The V-interval variability measures %V and nPVI do not show the same pattern for between-language variability as speech rate and the consonantal rhythm correlates. %V has been shown to be equal for English and German as well as French and English. Czech was shown to be somewhere between these groups. It will be tested, however, whether an interaction of %V with speech rate is possible.

Using the high within-language speech rate variability it will be assessed whether consonantal rhythm correlates and speech rate interact. It will then be tested whether these influences vary between languages.

5.6. Conclusion about speech rate factors in BonnTempo

In this chapter a variety of issues about speech rate in the BonnTempo corpus were addressed. First of all it was shown that in respect to an investigation of measures based on durational C- and V-interval variability it is important for rate standardizations to switch from the syllabic unit to C- and V-intervals per second as rate indicators. For CV-rates (the sum of the number of C- and V-intervals per second) it was then shown that the intended speech tempo classes produced the wanted effect of highly variable speech rates. This high variability of rates will be important in the following chapter when rhythm measures will be studied under rate influences.

It was also shown in this chapter that languages vary in speech rate at each intended tempo category and that languages traditionally classified as syllable timed languages turned out to be faster than traditional stress-timed languages. This was

attributed to the fact of different complexity of the different languages. Czech, a yet unclassified language, overlaps mostly with English. The way languages allow increase in speech rate was found to be equal for all languages apart from English that proved to be at a similar rate of Czech for slower speech but slower than all languages at higher speech rates.

Grouped into rhythmic classes it was found that CV-rate is an equally good indicator of rhythmic class like all other rhythm measures based on durational C- and V-interval variability. Thus it can be concluded that rhythm classes, if they exist, can be predicted by speech rate only and it is assumed that there may be an interaction between speech rate and rhythmic variability measures. This will be tested further in the next chapter.

6. Speech rate influences on rhythm measures

In chapter 5 it was demonstrated that a high variation of measurable speech rates could be produced by prompting speakers to vary their speech rate from very slow in several steps (slow, normal, fast) to as fast as they can. It was further demonstrated that at any of the intended speech tempo versions CV-rate works equally well as a predictor for rhythmic class as any of the rhythm measures introduced in chapter 4. The general idea for the present chapter is to see whether there is an interaction between speech rate and rhythmic measures. In a first step it will be analyzed whether durational C- and V-interval variability measures are dependent on variations of speech rate.

6.1. Dependency of rhythmic correlates on speech rate

6.1.1. Method

In order to investigate whether rhythmic measures interact with speech rate each rhythmic measure will be correlated with CV-rate. The full bandwidth of realized rates triggered by the intended speech tempi will be used for this analysis. In a first step each rhythm measure will be cross-plotted against speech rate for a descriptive analysis. With a curve analysis procedure it will be tested which mathematical function best describes the relationship. In the following, first the original measures from Ramus et al. (1999), %V and ΔC, will be analyzed, after that the Grabe & Low (2002) measures nPVI and rPVI.

6.1.2. Ramus et al. (1999) measures %V and ΔC

Figure 6-1 shows the %V (top) and ΔC (bottom) cross-plotted against CV-rate (the sum of the number of C- and V-intervals per second). The graphs show clearly that there is a relationship between CV-rate only between ΔC and CV-rate but not between %V and CV-rate. The relationship between ΔC and rate can be described as a negative correlation, i.e. an increase in CV-rate leads to a decrease in ΔC.

A linear and a logarithmic curve have been fitted to both data plots. As expected the R square for the %V/CV-rate relationship turns out very poor. In both cases it returns a value of 0.035. It can thus safely be concluded that there is no relationship between %V and CV-rate.

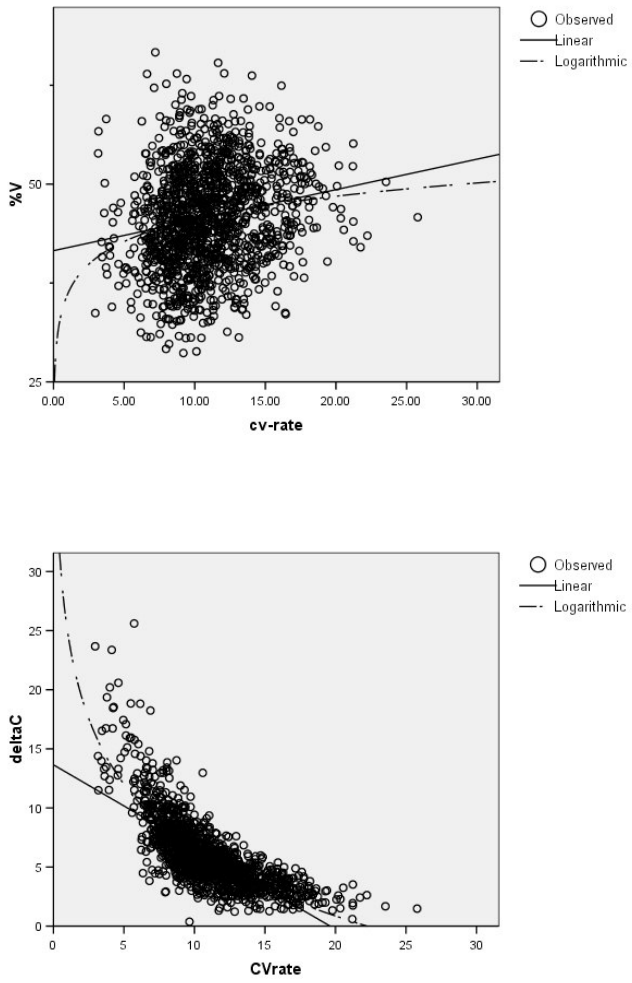

Figure 6-1: Scatter plot of %V (top) and ΔC (bottom) as a function of CV-rate (plot points represent the data for one sentence) with a linear and logarithmic curve fitted.

For ΔC the returned R square value for the linear fit results in 0.535 and for the logarithmic fit in 0.63. It can be concluded that a logarithmic model is best for

describing the relationship between the two parameters.

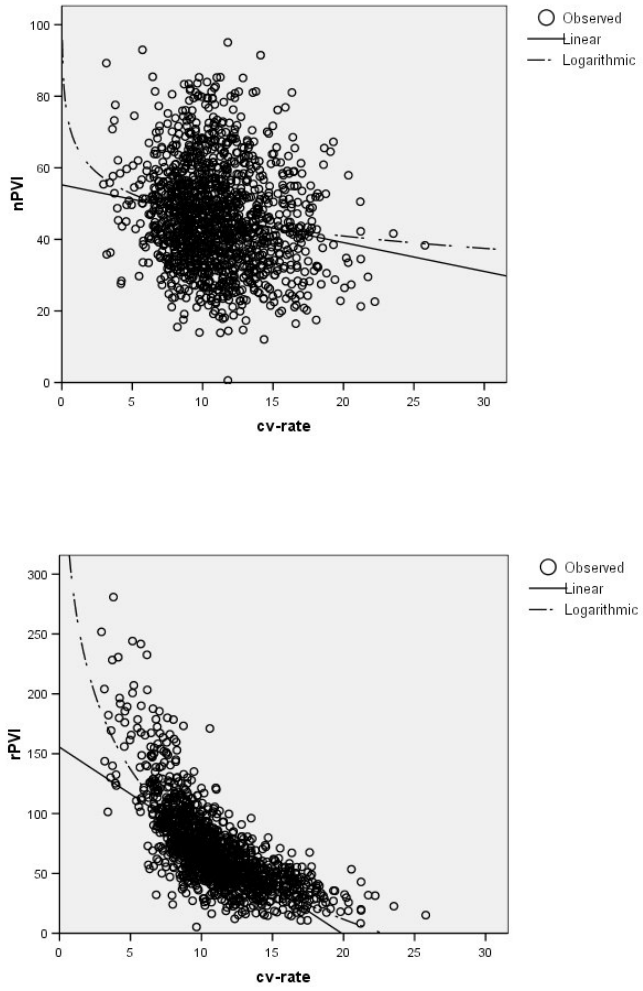

Figure 6-2: Scatter plot of vocalic nPVI (top) and consonantal rPVI (bottom) as a function of speech rate (plot points represent values for one sentence) with a linear and logarithmic curve fitted.

6.1.3. Grabe & Low (2002) measures nPVI and rPVI

A very similar situation as for %V and ΔC occurs with nPVI and rPVI where the vocalic measure does not correlate with CV-rate, however, the consonantal rPVI correlates negatively with CV-rate. Figure 6-2 displays these results (top: nPVI, bottom: rPVI). The curve estimation procedure shows again very poor model

prediction values for nPVI (R square linear: 0.03, logarithmic: 0.031). R square values for nPVI suggest a logarithmic function as the best predictor model (linear: .492, logarithmic: .577).

6.1.4. Discussion of CV-rate/rhythm measure relationships

Ramus and Grabe & Low measures show very strong relationships. Both the V-interval variability measures %V and nPVI do not correlate with CV-rate but the C-interval measures rPVI and ΔC proved to be strongly affected by rate. And both the C-interval measures reveal a logarithmic relationship with speech rate.

The assumption formulated in the discussion of the previous chapter can be considered as correct: the consonantal measures ΔC and rPVI are dependent on speech rate. When speech rate is slower, C-intervals are longer and thus C-interval variability is higher. This affects the standard deviation of C-interval durations (ΔC) and the absolute durational variability monitored by the rPVI.

Grabe & Low (2002) argue that a dependency of the consonantal rhythm measure on speech rate may not matter and that normalization is not required. However, it was demonstrated in chapter 5 that speech rate can equally distinguish between languages and/or rhythmic classes. The question that remains is: can consonantal rhythm measures distinguish between rhythm classes after they have been normalized for speech rate? If they do not distinguish between rhythmic classes after normalization this may suggest that speech rate is the only reason for the rhythmic class distinction. In the next section both C-variability measures, %V and rPVI, will be normalized for rate and in the section after that it will be investigated, how well the rate normalized measures distinguish between rhythm classes.

6.2. Normalizing consonantal measures for rate

Ramus et al. (1999) normalized for rate by using sentences of 15 to 19 syllables which are of roughly 3 seconds in duration. A closer analysis of this normalization shows that it does still allow strong rate variability. If it was assumed that each sentence had a duration of exactly 3 seconds, then a rate variability between 5 and 6.3 syllables per second would be tolerated. Since the total number of C- and V-intervals per syllable varies from language to language (see chapter 5) this would

result for example in a CV-rate of 8.95 to 11.3 C- or V-intervals per second for French (1.79 C- and V-intervals/second, see chapter 5) and 9.25 to 11.7 for German (1.85 syllables/second). For German and French this variability represents roughly the variability of the inter-quartile ranges of these languages at normal intended speech tempo, however the absolute figures for German are slightly lower and for French are slightly higher. So if German and French was standardized by this method a slightly higher than normal CV-rate for German would be compared with a slightly lower than normal CV-rate for French. Furthermore, since it was demonstrated in chapter 5 that speech rate varies significantly between languages of different rhythmic classes at normal intended reading tempo, it is conceivable that the data used by Ramus et al. (1999) contained considerable speech rate differences to allow variability in ΔC between rhythm classes to be significant. In addition to that, if it was assumed that 'roughly 3 seconds' includes a variability of +- 10% then sentence durations of 2.7 to 3.3 seconds were allowed in the data of Ramus et al. (1999). This would lead to a maximum possible variability of 4.54 syllables (15 syllables/3.3 seconds) to 7.04 syllables (19 syllables/2.7 seconds) for the sentences in the data set which, in return, would be resulting in CV-rate range differences that are equally high as the data at normal intended speech tempo in BonnTempo. From this analysis it can be hypothesized that the rate normalization procedure in Ramus et al. did most likely not control for rate effectively and allowed a rate variability that is likely to be equally high as in the uncontrolled data at normally produced rate in BonnTempo. Or in other words: it is conceivable that the rate controlled data in Ramus et al. (1999) reveals rate variability that is similar to the rate variability in normally produced speech.

In Grabe & Low (2002) speech rate was only normalized for the V-interval variability measure and the effect of this normalization procedure was demonstrated to be extremely effective since nPVI does not correlate with rate even when there is a maximum CV-rate variability (see above). The C-interval measure rPVI is not normalized by Grabe & Low and a normalization procedure will be suggested in the next section.

6.3. Normalizing ΔC for rate

6.3.1. Data considerations

Most of the data analysis is based on the assumption that the underlying dataset is normally distributed, i.e. the sample set resembles a Gaussian distribution. A Gaussian distribution is only important for ΔC; the calculation of a standard deviation assumes the underlying data to be normally distributed. %V as well as nPVI and rPVI do not require the data to be normally distributed. Basically any of the manifold works on speech rhythm based on standard deviations of any interval in speech and any work using numerous inferential statistical models need to assume normally distributed data.

Durations of speech intervals are not necessarily normally distributed. After all, speech segments must be assumed to reveal some floor value of maximum shortness (a segment can hardly be shorter than 5 ms) but not of length (especially vowels can be of very long duration, for example as an effect of phrase final lengthening). For this reason it could be assumed that speech units of any type (consonants, C-intervals, vowels, V-intervals, or syllables, etc.) may well be positively skewed, i.e. the right part of the data distribution graph possesses a long tail which is the result of a comparatively small number of data points of high values.

Figure 6-3 shows that this assumption is correct. The graph displays the distributions of C- and V-intervals as well as syllables (s). A black line superimposes a Gaussian normal distribution. It is clearly visible for each case that the bulk of the data is shifted to the left of the normal distribution peak and higher values occur at low frequency. Furthermore the peak values of the central data are much higher than for a normal distribution. This phenomenon is referred to as positive kurtosis.

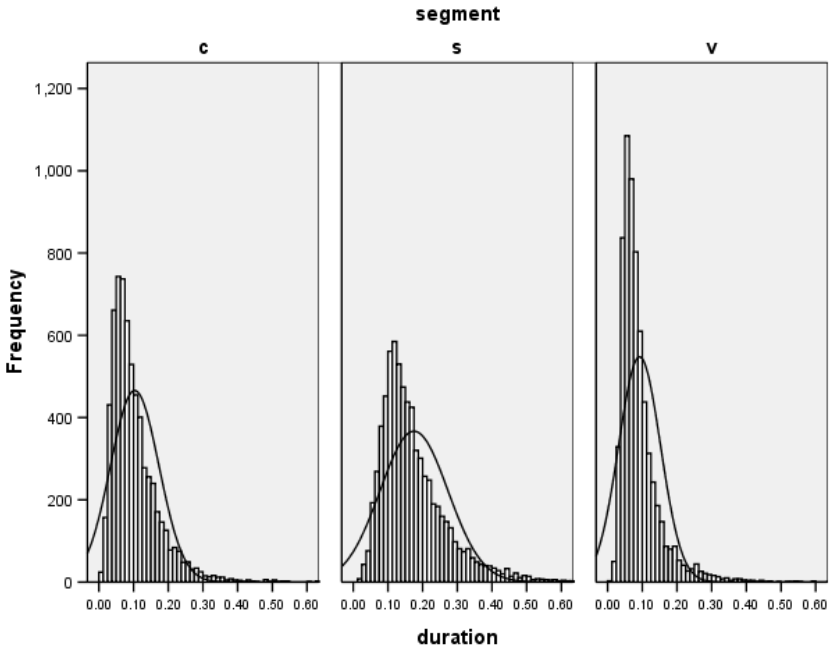

Figure 6-3: Three histograms showing the distribution of C-, V- and syllabic interval durations (in seconds; C, V, S respectively) with superimposed Gaussian normal distribution curve (black line).

A way of reducing positive skew and kurtosis is by applying a logarithmic transformation to the data in which the logarithm of each value typically to the base e (Euler's number) is calculated. The results of such a transformation (henceforth: ln-transformation) are visible in Figure 6-4 which shows the distribution frequencies of C-, V- and syllabic intervals for the ln-transformed data. From a visible impression the data can now be regarded as approximating a normal distribution.

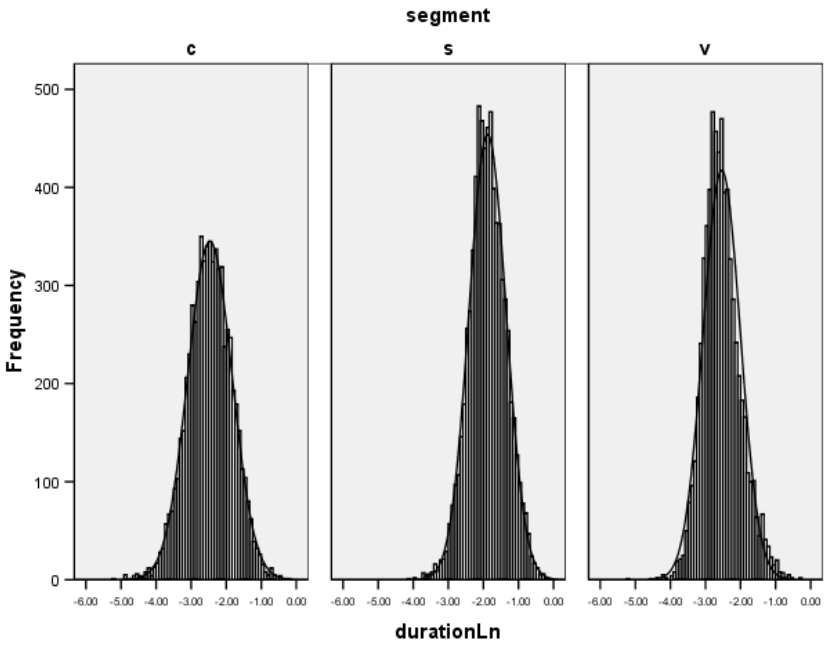

Figure 6-4: Three histograms showing the ln normalized C-, V- and syllabic interval durations (C, V, S respectively) with superimposed Gaussian normal distribution curve (black line).

The visible impression can be measured objectively by calculating skewness and kurtosis coefficients for the syllables, C- and V-intervals before and after the ln-transformation. Table 6-1 displays the results of such an analysis skewness and kurtosis values for the raw (raw) and logarithmic transformed data (ln) are available in the columns for syllable (s), C- and V-interval duration. It is apparent from the table that skewness got significantly reduced from positive values between roughly 2 and 3 to values around 0. Only in the case of V-intervals a skewness coefficient of .45 still remains. On the one hand this may be regarded as acceptable, on the other hand it needs to be pointed out that the only measure in the current context that assumes non skewed data is ΔC, thus for V-interval measures the normalized data will not be considered anyway. The logarithmic transformation also had a tremendous effect on kurtosis values which were reduced from values between about 5 and 14 to values not higher than 0.5. In the case of C-intervals which, again, are the most relevant intervals to be normalized, kurtosis came down to a negligible 0.084.

	Skewness			Kurtosis		
unit	raw	ln	standard error	raw	ln	standard error
S	1.72	-.06	.016	4.8	.24	.032
V	2.87	.45	.017	14.38	.47	.033
C	2.11	-.01	.017	7.98	.084	.033

Table 6-1: Values for skewness and kurtosis before (raw) and after (ln) ln-transformation for syllabic (s), consonantal (C) and vocalic (V) intervals.

It has been demonstrated in this section that syllabic as well as C- and V- intervals are strongly positively skewed and reveal a considerable amount of kurtosis which would have a strong influence on all analysis procedures assuming normally distributed data. The most important rhythm measure for which this is the case is ΔC. It could be demonstrated that a logarithmic transformation of C-interval durations leads to a data distribution that comes very close to a Gaussian normal distribution.

6.3.2. Method

Dellwo (2006) made an attempt to control for speech rate by calculating the variation coefficient for ΔC (varcoC) which is defined as the percentage difference of ΔC to a respective mean C-interval value (see equation 3-2 in section 3.3.2). Dellwo found that there was a reduction of speech rate influence on mean ΔC values for intended speech tempo classes. As all of the published analysis in the field which were viewed for the present study also Dellwo (2006) implicitly assumed that the durations of C-intervals are normally distributed. However, it has been demonstrated above that all intervals under investigations (C, V and syllabic) are extremely positively skewed and may contain a considerable amount of kurtosis. It is therefore argued here that mean values, standard deviations and variation coefficient of standard deviations are extremely influenced by these non-normal distribution characteristics. ΔC will therefore be processed for the ln-normalized data first. If speech rate influences remain for this data the coefficient of variation as proposed by Dellwo (2006) will be further applied.

6.3.3. Results & Discussion

ΔC plotted across CV-rate for ln normalized data can be viewed in Figure 6-5. The plot includes the values for each phrase in BonnTempo for the languages under investigation (English, German, French, Italian and Czech). It is evident from the plot that no relationship can be detected between rate and ΔC based on ln normalized data (ΔCln). A linear regression analysis confirms this visual impression with a low R-square value of 0.045.

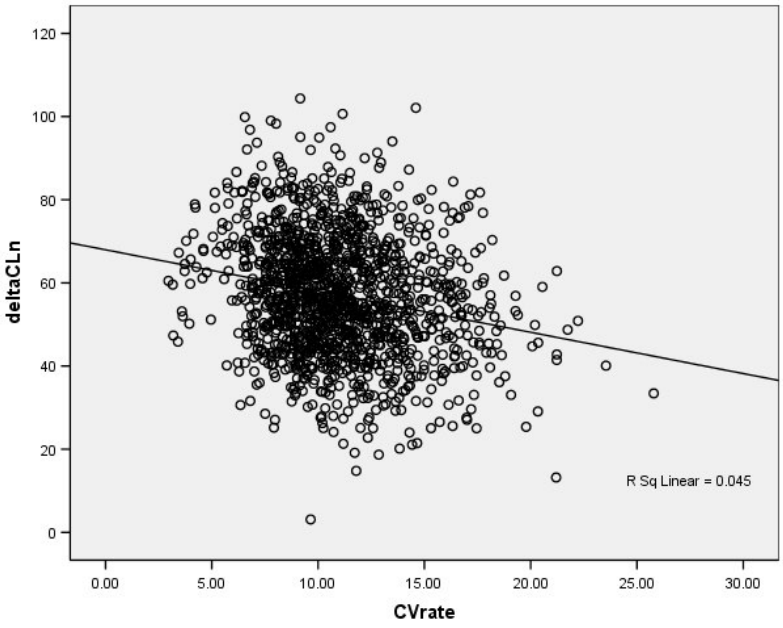

Figure 6-5: Scatter-plot showing ΔC plotted over CV-rate for ln-transformed data (dots represent each phrase in BonnTempo for the languages under investigation).

This result shows that an ln-transformation of the data is already sufficient for normalizing for speech rate effects. An explanation for the effect is straightforward. A logarithmic transformation leaves intervals of short duration at more or less equal duration whereas long durations are shortened drastically. So the mean differences between fast and slow rates are reduced drastically and what remains is the individual variability of the tokens. A further normalization using the coefficient of variation as proposed by Dellwo et al. (2006) cannot possibly introduce a higher

normalization effect. Possible variability between languages will be discussed and analyzed below.

6.4. Controlling rPVI for rate

In an analysis of the PVI Wagner and Dellwo (2004) normalize for speech rate using the z-scores of each C- and V-interval. This normalization procedure underlies the same criticism as above: calculating z-scores assumes a normal distribution of the underlying data which is not given for C- and V-intervals. With this analysis a processing with ln-transformed data is not easily possible. Since the PVI calculates the differences between consecutive C- and V-intervals these differences would decrease logarithmically as well, thus results for nPVI or rPVI may represent an entirely different picture after a transformation is applied. However, PVI calculations do not rely on normally distributed data. For this reason a different procedure of data transformation is suggested here that does not correct for the non-normal distribution but does correct for speech rate influences. This procedure takes the percentage for each interval of the total duration of the phrase it has been taken from. With this method, all phrases of whatever duration are set to 100% and the variability of the respective intervals is expressed in relation to this.

6.4.1. Method

As explained above normalization for speech rate by using z-scores is not suitable for nPVI since z-scores require the underlying data to be normally distributed. A normal distribution using ln-transformation, however, warps the differences between consecutive intervals in unwanted ways. For this reason normalization is required that makes the actual phrase durations relative without changing the proportions of durations. Such normalization can for example be reached as performed in Grabe & Low (2002) for vocalic intervals by each individual durational difference between two consecutive V-intervals by the average duration of the V-intervals. Although Barry et al. (2003) criticize that this normalization is only a local one and does not account for speech rate changes over a longer piece of discourse this has demonstrated as incorrect (see discussion above). The normalized PVI for V-interval calculations normalizes effectively for maximum speech rate changes as found in the

BonnTempo corpus. In order to test whether this normalization procedure is applicable for the rPVI in identical normalization procedure as for the nPVI was applied according to the following formula. The resulting measures could be called nPVI (from: normalized PVI) as well but in order to avoid confusion with this measure it is referred to as nPVI-C in the following.

Equation 6-1:

$$\text{nPVI-C} = 100 \cdot \frac{\sum_{i=1}^{N_C-1} \left| \frac{C_i - C_{i+1}}{(C_i + C_{i+1})/2} \right|}{N_C - 1}$$

N_C = number of C-intervals sampled
C_i = duration of the i^{th} C-interval

6.4.2. Results and discussion

The results of the rate normalized consonantal nPVI-C can be viewed in Figure 6-6. The graph shows the nPVI-C plotted across speech rate and it is obvious that the normalization procedure serves as an effective control for speech rate in case of the consonantal variability measure too. With an R square of 0.032 as a result of a linear regression analysis it can be concluded that there is no correlation between the CV-rate and the nPVI-C.

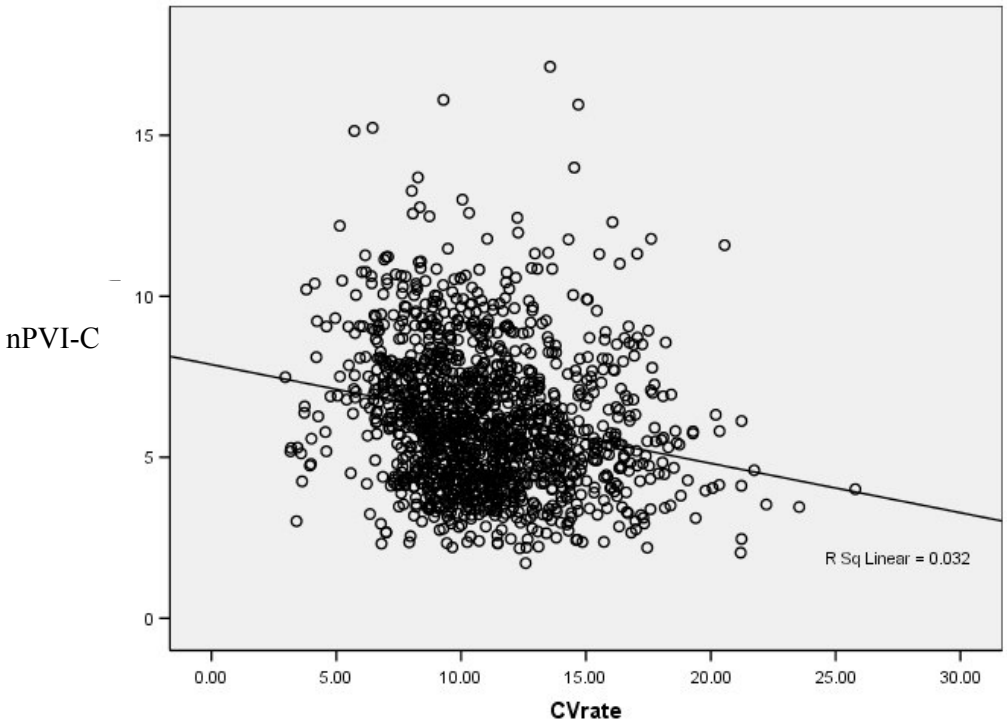

Figure 6-6: nPVI-C plotted across CV-rate.

Grabe & Low do not normalize for rate since they argue that rate can have an effect on C-interval durations in two ways: a) consonantal segments can be affected by rate in that they change their duration, b) consonantal clusters can be affected by rate in the way that certain consonants underlie elision effects, i.e. they are eliminated with increasing speech rate. Languages with a low amount of complex clusters (here: Italian and French) are in most cases only liable to reduction effect a) since single consonant clusters should be more robust against elisions.

If Grabe & Low's hypothesis was correct then complex C-cluster languages (traditionally stress-timed languages) should show stronger variability as an effect of speech rate then simple C-cluster languages since the latter languages allow C-cluster durational variability on two dimensions (consonantal duration variability and number of consonantal variability). Languages with a higher amount of simple clusters allow only consonantal duration variability within each cluster. In the next section evidence for this hypothesis will be gained.

6.5. Robustness of rhythm measures across speech rates

In this section it will be tested how reliable rhythmic measures are across different speech rates. This will include a test whether a) the previously developed normalization procedure had an effect on the rhythm measures and b) languages behave differently according to rhythmic variability as a function of rate.

There is the possibility that different languages react differently to changes in speech rate. Two previous observations lead to this assumption:

a) Different languages were shown in chapter 3 to possess different amounts of consonantal complexity. It was demonstrated for the reading material for BonnTempo that French and Italian possess a lower number of complex intervals than English and German (with Czech somewhere in the middle between these two groups). Languages containing a lower complexity have been shown to be lower in ΔC and ΔCln (because they possess less variability) and higher in CV-rate (because they possess shorter C-intervals). This observation gives reason for an assumption that languages with shorter less complex intervals also show less variability of C-intervals over different speech rates. The major amount of C-intervals in Italian and French consist only of a single consonant, so only the duration of this consonant can be changed over different speech rates. Languages like German contain a high amount of much more complex C-intervals (see chapter 3) and at higher speech rates numerous consonants may be lost as an effect of unit elisions, or degraded as an effect of coarticulation. These effects may well introduce a higher variability in C-interval durations with a change in speech rate.

b) It was demonstrated in chapter 5 that languages vary in the range of speech rate differences they allow. In this context it was found that English allows less speech rate differences than the other languages under investigation (German, French, Italian, and Czech). If languages allow less overall changes in speech rate, the variability of C-intervals can overall be expected to be less affected.

6.5.1. Method

In order to compare speech rate influences on rhythm measures between languages rates have to be made comparable across the different languages under investigation. This was already the case when speech rate ranges were made comparable between

languages. A method showing the rate for each percentile was used in this context (see chapter 5). In the present case, a method is needed again that contrasts comparable rates between each language with each other. Calculating averages of the C- and V-variability measures for each percentile showed, however, too much noise in the data. For this reason rhythmic measurements have been summarized over a larger percentage of the data, the quintile, i.e. the data ranges in each 20% step of the distribution. In the following the lowest quintile refers to the lowest 20% of the data (i.e. quintile 1 is the lowest 20% of the data, from 0 to 20%, and quintile 5 the highest 20 %, from 80 to 100%).

6.5.2. Results and discussion for speech rate dependent measures

In the present section the speech rate dependent measures rPVI and ΔC will be analyzed. The results for rPVI before and after the normalization can be viewed in Figure 6-7. The graph plots the average rPVI (y-axis) for each language (lines) at each of the five quintiles (x-axis). In the top graph the clear influence of speech rate on the rPVI discussed above is apparent again. All languages show a clear decrease in rPVI with increasing CV-rate. However, the data gives the impression that this decrease in rPVI is much stronger in the case of German than in the case of English and Czech which is again stronger than French and Italian. Five ANOVAs with rPVI as the dependent variable and the CV-rate quintiles as a five class factor were calculated, one for each language. All ANOVAs show that there are highly significant differences between quintiles for each language. Post-hoc the analysis reveals that there are highly significant differences ($p<.001$) between the rPVI distributions at each quintile in German while in Czech and English such differences can only be found between the extreme quintiles (1 and 5). In French and Italian these differences can only be found marginally for the extreme quintiles.

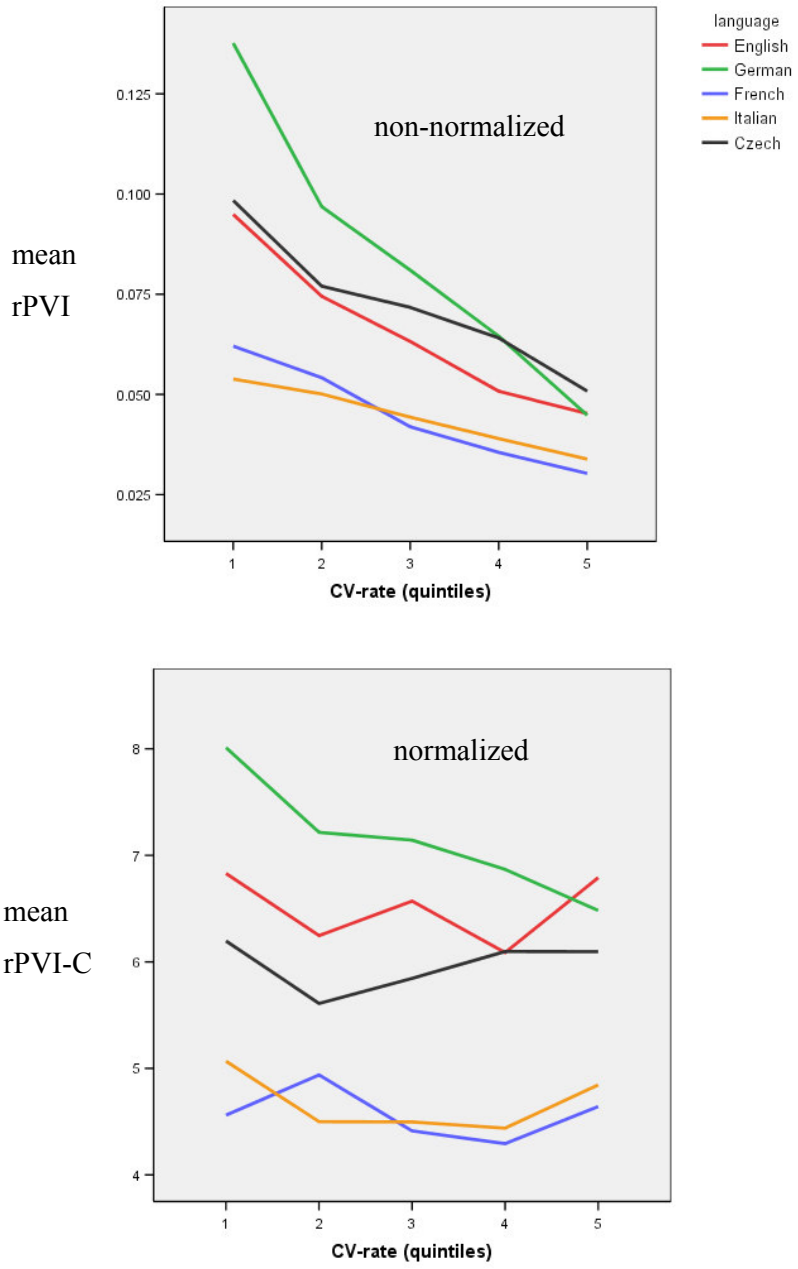

Figure 6-7: rPVI averaged over quintiles of CV-rate, top: non-normalized, bottom: normalized.

After normalization (bottom graph) speech rate effects on rPVI have been reduced in any of the five languages under investigation. An ANOVA produced with the same parameters as above shows that there is only significant variability between the

quintiles in the case of German($F[4,524]=4.06$, $p= 0.003$) but not for any other language (English: $F[4, 524]=0.54$, $p=0.71$; French: $F[4, 209]=1.14$, $p=0.351$; Italian: $F[4, 104]=0.52$, $p=0.72$; Czech: $F[4, 314]=1.02$, $p=0.34$). A post-hoc analysis for German shows that significant variability of rPVI across the CV-rate quintiles is only existent between quintile 1 and 4 ($p=0.03$) and 1 and 5 ($p=0.001$). This analysis shows that speech rate normalization in the case of rPVI was very effective. Speech rate differences only remain in German between extreme rates. These differences, however, can be neglected since the type of rate variability in BonnTempo is unlikely to occur in real speech situations. It can therefore be concluded that rPVI is very robust variability measure across different speech rates in all languages.

An additional ANOVA tested rhythm class differences at each quintile with rPVI as dependent variable and rhythm class as 2 class factor. Czech was excluded from this analysis, English and German were attributed to the stress-timed, French and Italian to the syllable timed group. The analysis shows that at each quintile highly significant differences are obtainable between rhythm classes ($p<0.001$). This analysis shows that the normalized rPVI is also a very robust measure to distinguish rhythmic class across variable speech rates in all languages.

In analogy to the analysis in chapter 4, languages were compared with each other within and between rhythm classes to investigate how reliable rhythm classes are represented. For this again 5 ANOVAs have been produced with language as a 5 class factor and the speech rate normalized rPVI as a dependent variable. For each ANOVA highly significant differences ($p<0.001$) can be detected between languages. A post-hoc analysis reveals perfect support for rhythm classes can be found at the 3rd, 4th and 5th quintile where comparisons between English/German and French/Italian show no significant differences. Further all between-class comparisons of languages (English-French, English-Italian, German-French, and German-Italian) show significant differences with p at least smaller than .05. For the slower CV-rates at the 1st and 2nd quintile, differences within rhythmic class can be detected between English and German. Between rhythm class comparisons of language pairs are all significant. This analysis shows that the normalized rPVI is a very robust measure to distinguish between rhythmic class across different speech rates.

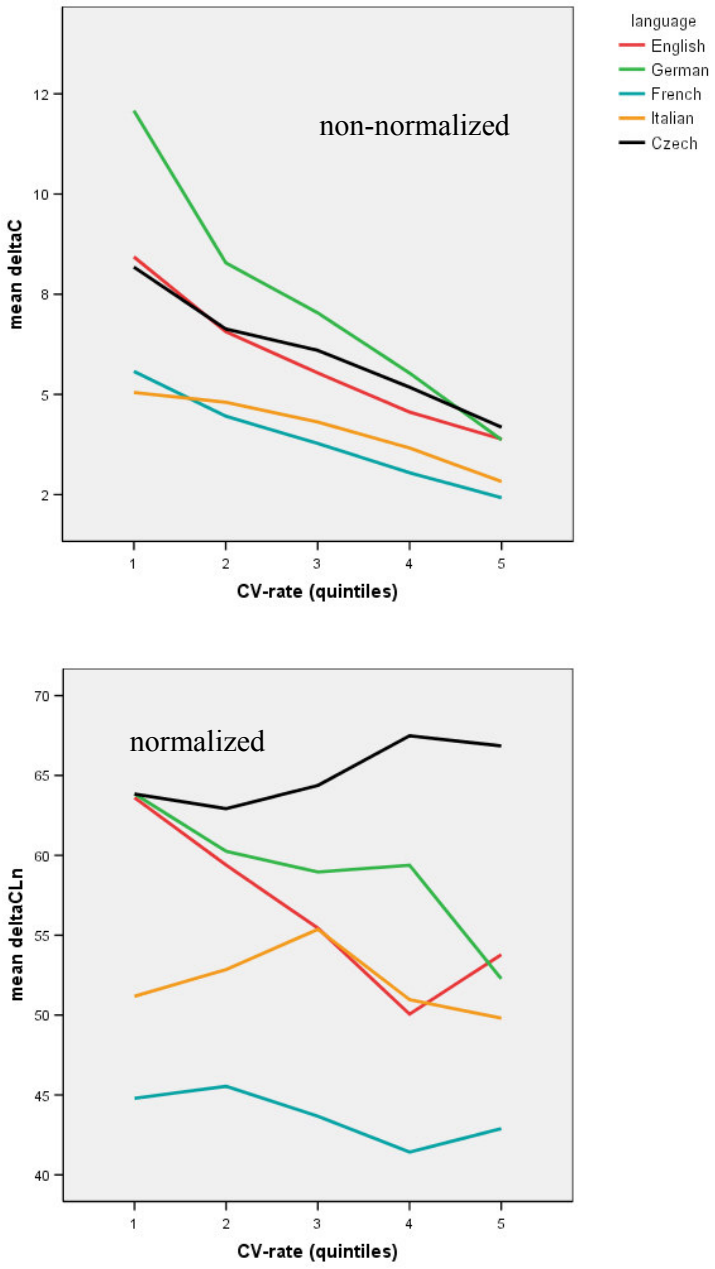

Figure 6-8: ΔC averages over quintiles of CV-rate. Top: underlying data are raw durations, bottom: data is ln normalized.

For the normalized ΔC the rate normalization by processing the measure for the log normalized data did not have comparable effects. Figure 6-8 plots the average ΔC (y-axis) for each language (lines) at each quintile (x-axis) before (top graph) and after

(bottom graph) the rate normalization. The top graph (before normalization) reveals that ΔC produces very similar overall results concerning the distribution of the different languages compared to the rPVI (see top graph in Figure 6-6). Languages (rhythmic classes) are equally well separated and rate influences seem to affect the measures similarly. ANOVAs produced in the same manner as for the rPVI with ΔC as the dependent variable and language or rate-quintiles as a five class factor, revels nearly matching results in both cases. It can therefore be concluded that ΔC and rPVI seem to measure very equal aspects in the signal when they are not normalized for rate. However, rate normalization in the case of ΔC creates a very different picture. Languages seem to be very unequally affected by the normalization and a rhythm class distinction is not very well visible any more. While at quintile 1 and 2 English, German and Czech are grouped together, nearly equal values for Italian and English can be obtained at quintiles 3 and 4. At the 5^{th} quintile (fastest CV-rates) the languages German, English and Italian group together, Czech is significantly higher and French significantly lower. ANOVAs for testing within-language variation (quintile as a five class factor and language as the dependent variable) help interpreting the situation: only for the languages English and German a significant variability between the quintiles can be detected (English: $F[4,244]=6.5$, $p<.001$; German: $F[4, 524]=12.1$, $p<.001$) but not for any of the other languages under investigation (French: $F[4,209]=.99$, $p=.42$, Italian: $F[4, 104]=.82$, $p=.51$, Czech: $F[4, 314]=1.7$, $p=.143$). Post-hoc the analysis reveals for English that there is significant difference between the quintile pairs 1/3, 1/4, and 1/5 as well as pair 2/4. In German only the 5^{th} quintile is significantly different from all other groups.

So rate normalization did seem to be effective apart from little influences that still remain in English and German. In conclusion it can be said that the rate normalization procedures applied to rPVI and ΔC were effective, however, in the case of ΔC languages reacted differently with English and German still showing traces of interactions between ΔC and rate. However, because of these traces rhythmic class separation of ΔC is strongly distorted after normalization. For this reason, rPVI is considered the more robust measure for rhythmic class distinction under rate influences.

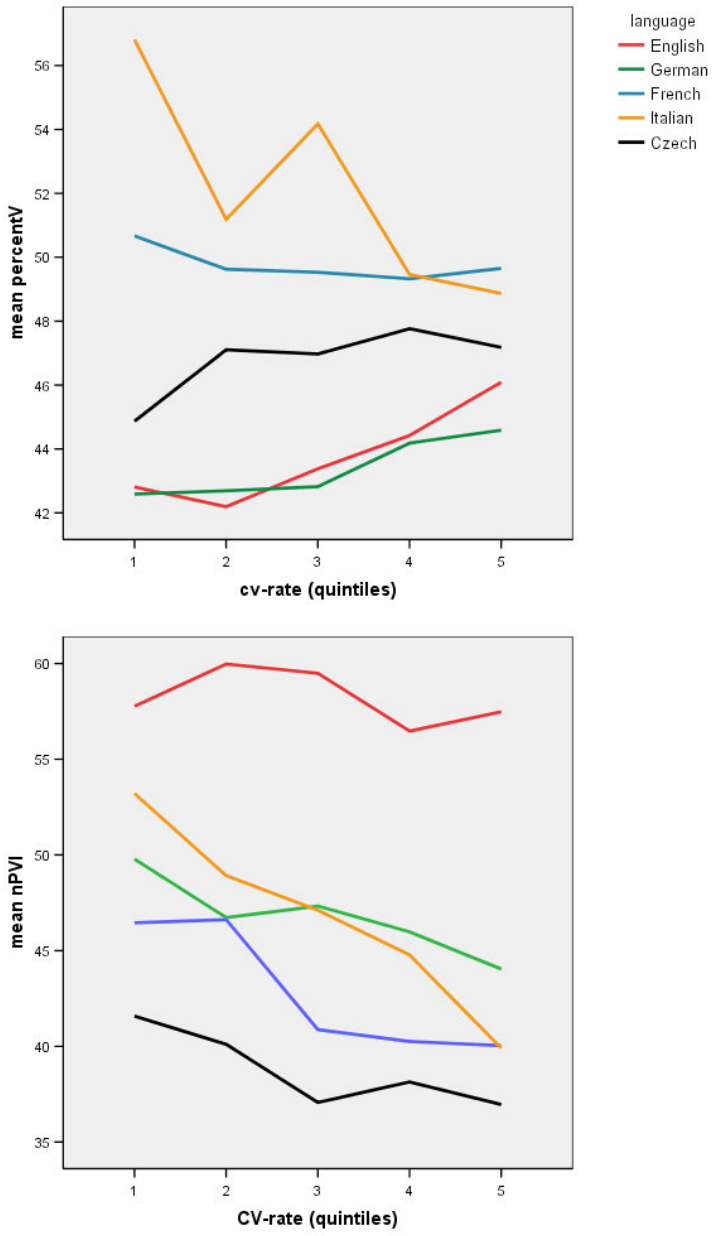

Figure 6-9: %V (top) and nPVI (bottom) for quintiles of CV-rate for the languages English (1), German (2), French (3), Italian (4), and Czech (5).

6.5.3. Results and discussion for speech rate independent measures

In the same manner as above the vocalic measures %V and nPVI have been tested for within- and between-language variability across different rates. For %V (see Figure 6-9, top graph) descriptive results for within-language variability show that,

apart from Italian, is rather constant across all quintiles with a little tendency of an increasing %V in the languages English and German. An ANOVA testing within-language variability (ANOVA for each language with quintile as a 5 class factor and %V as the dependent variable) shows that this variability is non-significant in the cases of French ($F[4, 209]=11.5$, $p=.62$) and Czech ($F[4, 314]=2.14$, $p=.06$). All other languages show significance levels at least at the $p<.05$ level. Post-hoc the analysis reveals that in English there is only a significant difference between quintile 5 and 2, all other comparisons are non-significant. This means that the tendency of %V to increase with rate in English is not a very prominent feature overall. For German the situation is similar; the effect can be considered as not very strong ($F[4, 524]$, $p=.028$). An analysis of eta-squared however reveals that only 2% of the variance are accounted for and post-hoc no significance between any of the groups can be detected. For this reason the variability of %V as a function of CV-rate will be neglected in English and German.

For Italian there is the strongest variability of %V across the different CV-rate quintiles ($F[4, 104]=7.8$, $p<.001$). However, this variability seems rather randomly distributed across the five quintiles. Post-hoc the analysis shows that there is significant variability for the pairs 1/2, 1/4 and 1/5 as well as 3/2, 3/4 and 3/5. It is questionable whether this strong variability of Italian is really a language dependent effect. It may be that the rather small dataset for Italian contributed to this variability (3 speakers as opposed to 5 speakers and more in other languages). This point will be further discussed in the next chapter.

ANOVAs to test for between-language variability (dependent variable: %V, language as a 5 class factor for each quintile) reveals that there are highly significant differences at each quintile ($p<.001$). A post-hoc test shows that at quintile 2, 4, and 5 the rhythm class distinction is perfect. There are no significant differences between the language pairs French/Italian and English/German. At quintiles 1 and 3 there are significant differences between French and Italian. For all between rhythm class comparisons of languages (English/French, English/Italian, German/French, German/Italian) the pairs reveal highly significant differences ($p<.001$). Additional ANOVAS with rhythm class as a two class factor (Czech not included) confirms this result and shows that rhythm classes are distinguished at a high significance level

($p<.001$) at each quintile.

Czech is descriptively somewhere in the middle between both language pairs (see figure) and in the inferential domain significance effects are typically weak and vary in random fashion between Czech and other languages across the quintiles. In conclusion it can be said that Czech is difficult to attribute to a rhythm class according to %V.

For the remaining part of this section the nPVI will be analyzed. Descriptively it appears (see graph) that the nPVI is decreasing with an increase in CV-rate for the languages French, Italian and Czech. Inferentially an ANVOA for each language (dependent variable: nPVI, quintiles as a 5 class factor) reveals that this effect is given but that it is very weak in each case (significance levels are all closely below the $p<.05$ mark and eta-square values reveal that only up to 4% of the variance is accounted for). Post-hoc the analysis shows that only very few of the quintile comparisons show significant effects and if they appear they are very weak (i.e. being close to $p=.05$). It can therefore be concluded that within languages the nPVI does not show any effects of CV-rate. Between languages, however, the situation is different, since at all quintiles highly significant differences can be found for ANOVAS with language as a five class factor and nPVI as the dependent variable. Post-hoc the analysis reveals a picture that is not in support of rhythmic class distinction. English is throughout the quintiles highly significantly different from all other languages ($p<.001$; apart from quintile one where it is not different from French). German shows no significant differences to either French or Italian at any of the quintiles. Czech is different from all other languages at quintiles 1 and 2 but shows no differences with French at 3 and French and Italian at 4 and 5. In summary this means for the rhythm class distinction that the nPVI groups German together with syllable-timed French and Italian and separates English from all the rest. In conclusion: of all the rhythmic correlates nPVI is the measure that supports the rhythm class hypothesis least.

6.6. Conclusions about speech rate influences on rhythm measures

In this chapter it was first demonstrated that the C-interval variability measures ΔC

and rPVI vary strongly as a function of speech rate (CV-rate). It was shown, however, that speech rate influences could be separated successfully from the measures with two suggested normalization procedures. After the normalization an analysis for the variability of individual languages across the quintiles of speech rate has been performed. This analysis revealed that for the languages under investigation, %V and the normalized rPVI are the measures that are a) least affected by speech rate variability in all languages and b) the strongest measures to distinguish rhythm class.

7. Additional influences on rhythm measures

The main topic of this work is the relationship between measures of speech rhythm based on the durational variability of C- or V-intervals. However, during the work with BonnTempo it became apparent that next to speech rate there are other strong influential factors on durational C- or V-variability measures. Such influences are for example speaker specific variations in the realization of C- or V-interval variability and also individual differences between sentences. These influences turned out to be significant across all intended speech tempo categories. Although these influences do not fit fully into the line of argumentation of the current work it was found that they are strong influential factors on rhythm measures that give an outlook for future work in the field. In addition to that the results presented here are of relevance for the methodology of the perceptual testing in chapter 8 where phrases with high variability will be needed in order to study whether within-language variability of %V is perceptually salient.

Since the intention of the present chapter is not to give an exhaustive description and analysis of all rhythm measures only %V was selected as an exemplary

7.1.1. Between speaker differences for %V

Variability for %V has been studied as a function of speaker. Exemplary results can be viewed in Figure 7-1. The box-plot shows the distribution of %V for six randomly selected German speakers (top) and the six French speakers (bottom) in BonnTempo. The data plots suggest that there is a higher variability between speakers in German than there is in French. An ANOVA with %V as the dependent variable and the six speakers as factors reveals significant differences between speakers in each language (German: $F[5,209]=18.58$, $p<.001$; French: $F[5,209]=3.62$, $p<.05$). However, a post-hoc test showed that there is significant difference between a wide variety of speakers in the case of German but in French only between speakers 6 and 2, 3, 5.

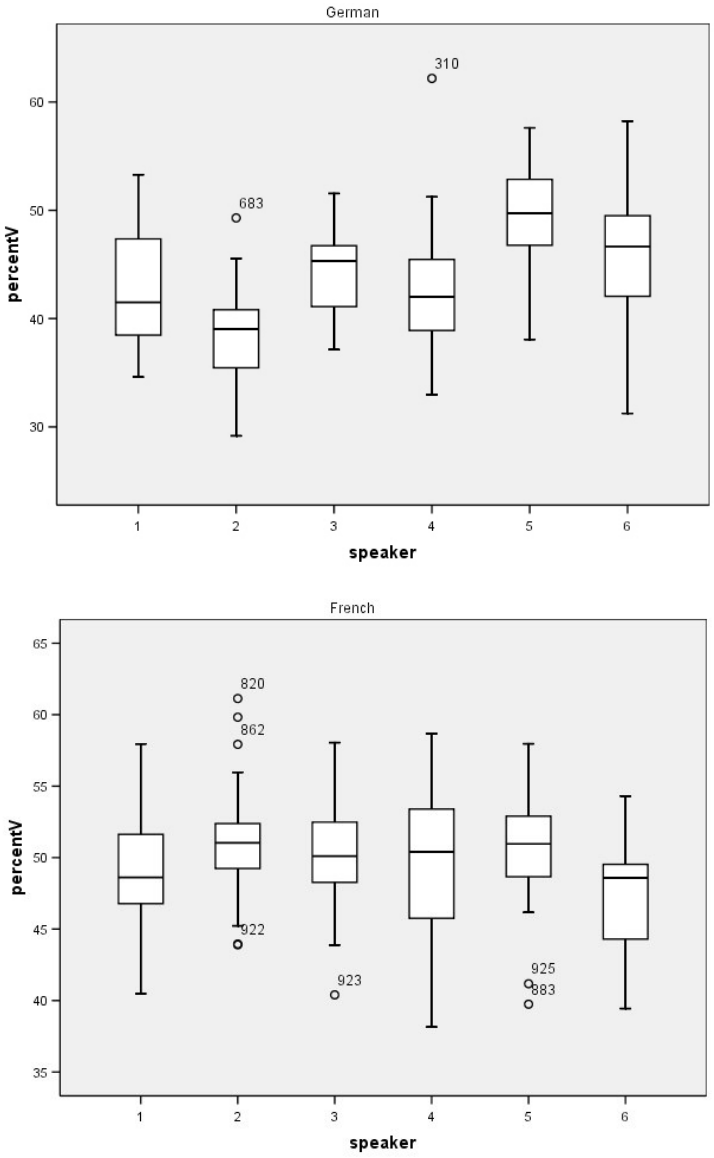

Figure 7-1: Box-plot showing the distribution of %V as a function of 6 speakers in German (top) and six speakers in French (bottom).

These results suggest that speaker dependent variability of %V may be different in languages. French seems more stable in %V across the six test speakers, German seems to allow more speaker specific variability.

7.1.2. Between phrase differences for %V

In the methodology all reading material was subdivided into 5 phrases that roughly

represent intonation units (since intonation units can vary highly between speakers' realizations this terminology was avoided). In the present section the between phrase variability for %V as an exemplary rhythmic measure will be studied.

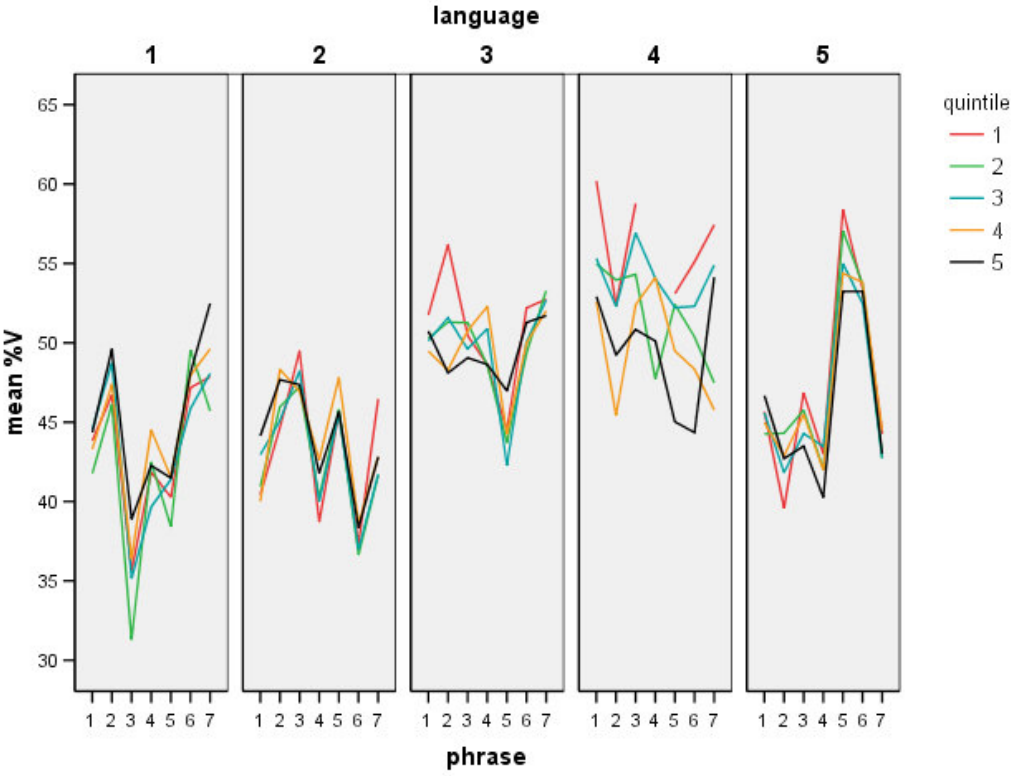

Figure 7-2: Line graph showing mean %V at each phrase for the languages English (1), German (2), French (3), Italian (4) and Czech (5).

Figure 7-2 is a line graph showing the results for mean %V for each phrase for the languages English (1), German (2), French (3), Italian (4) and Czech (5). The data has been separated (lines) by the quintiles of CV-rate thus quintile 1 represents the lowest 20% of CV-rate data, quintile 2 the CV-rate data between 20 and 40%, etc.
It can be seen that in all languages but Italian (4) there are very strong differences for %V between the phrases with very large agreement between the 5 different speech rate categories. In English (1) the range of %V values reached for each phrase lie between about 35 and 50%, a slightly smaller but similar range is true for German. In Czech, most phrases have %V values between 40 and 45%, however, phrase 5 and 6

lie about 10 to 15% points higher than this (at around 55%). This variation finds high agreement across all speech rates.

Italian appears to be a language with very low agreement of phrase differences across different speech rates since the lines show high variability and no consistent pattern is detectible. However, this effect has with a high probability been caused by the low number of subjects (3) in this language. An identical line graph for English and German with only three subjects showed equally random patterns across speech rates. It may therefore be likely that Italian showed equal agreement across speech rates for a larger number of speakers. This result shows nicely that for empirical analyses in the field of durational C- or V-interval variability and CV-rates high numbers of speakers are necessary to obtain reliable results. It is likely that a lot of the disagreement met in previous studies is due to low speaker numbers. Ramus et al. (1999) used four speakers per language; Grabe & Low (2002) used only one speaker representing the entire language.

7.2. Discussion and conclusion about other influences on the durational variability of C- or V-intervals

The data analysis in the previous section showed on the example of %V that rhythm measures are influenced by a number of different aspects in the speech signal. It was demonstrated that %V is extremely dependent on the content of the phrase. Preliminary investigations on the characteristics of these phrases did not lead to a straightforward answer for this result. It was further tested whether the duration of the phrase or the number of vowels in a phrase could have an effect but all these tests remained negative. It is assumed that segmental characteristics on a consonantal and/or vocalic level must be the driving force.

An example for two widely diverging phrases in German is (a) 'Es ist eine Fahrt ans Ende der Welt' and (b) ' Hinter Giessen werden die Berge und Wälder eintönig'. Phrase (a) has an average %V of 58% and phrase (b) 35% across all speech rate intervals. This difference is important in terms of the results obtained in chapter 4 where we found that German has an average %V of about 44% and French of about 49%. So of the two phrases picked, phrase (a) is drastically higher than that of a typical average syllable timed language, phrase (b) is lower than that of a stress-

timed language. Additionally it must be mentioned that both phrases do not differ in CV-rate in any of the intended tempo categories.

In order to find what the driving factors in the strong %V variability between phrases are further knowledge about the statistical occurrence of segment types are necessary that are not available in BonnTempo at the current stage. It may also be that certain intonational or stress patterns are the driving factor for %V to vary. However, these prosodic parameters have been demonstrated to be widely affected by changes in speech rate. So if %V was dependent on prosodic parameters and prosodic parameters change with rate it should also be assumed that %V changes in one or the other respect with rate. For this reason it should be assumed that %V is dependent on a parameter that does not change with speech rate. Such a parameter, however, is hard to find regarding the fact that also segmental duration parameters are strongly affected at higher speech rates. As discussed in chapter 5, higher speech rates introduce coarticulation, assimilation and even segment elision. However, %V was shown to stay robust against all these changes.

The results about the robustness of %V obtained here are well in accordance with other findings in the literature. Smiljanic & Bradlow (forthcoming) report about the robustness of %V in clear and conversational speech. The authors come to the conclusion that %V stays the same even when syllable durations vary widely between clear and conversational speech. The types of speech used may well be directly comparable to the speech in BonnTempo in that clear speech resembles BonnTempo speech at slow intended tempo and conversational speech resembles fast intended speech tempo in BonnTempo. Although BonnTempo speech is read, the fast versions well resemble the types of assimilation, elision and coarticulation found in conversational speech.

It seems impossible to give an answer for why %V remains robust against a wide variety of segmental and suprasegmental changes in the speech signal. However, current hypothesising on where to look for stable regions in the speech signal is motivated by findings from Dellwo et al. (2007) and Adrian Fourcin (personal communication). Dellwo et al. used the measures %V and ΔC and applied them to voiced and voiceless intervals in speech respectively (instead of C and V intervals as it has been done traditionally). This analysis leads to a drastically different

segmentation since now all voiced consonants are part of the 'voiced' intervals and only voiceless consonants remain in the 'voiceless' class. When calculating %V and ΔC for voiced and voiceless stretches it was found that an equally reliable classification of languages into rhythmic classes is possible. Dellwo et al. argue that such an analysis has multiple advantages. On a methodological level a labelling of intervals into voiced and voiceless stretches is much easier to obtain since it can be derived e.g. from a laryngographic recording that reliably monitors voiced and voiceless signals parts. In respect to the argument in Ramus et al. (1999) that infants can process %V and ΔC information without any knowledge of the language it was argued by Dellwo et al. that a processing of voiced/voiceless intervals requires even less amount of knowledge about the phonological system of speech. Also, it may well be that infants have the chance to be trained by the voice signal prior to birth because it may be perceivable in the mother's womb after most of the other information has been eliminated through low pass filtering. Coming back to the robustness %V it may be that %V is highly dependent on the voice impulse. Fourcin (personal communication) assumes that there is a rhythmic mechanism steering the voiced and voiceless intervals that may be independent of other prosodic parameters in speech. And it may be that such a voice timing mechanism is well robust against changes in speech rate. So if %V was dependent on such a rhythmic voicing mechanism it may be that this was the driving force for its stability. Further research on this topic is currently being carried out and answers are expected soon.

In conclusion it can be said that %V is robust against changes in speech rate and multiple segmental and prosodic variability and it can currently not be explained what are the constant factors throughout widely variable speech rates that make %V stay stable. One aspect in this discussion, however, remains. It was shown that within a language certain phrases can take values that are typical for syllable-timed languages and other phrases take values typical for stress-timed languages. Does this mean that listeners would also perceive these phrases as either stress or syllable timed? One aspect of a perception experiment in the next chapter is to give answers to this question.

8. Influences of rate on the perception of rhythm

8.1. Introduction

In chapter 4 it was demonstrated that languages are differentiable on an acoustic level in terms of durational C- and V-intervals variability. However, it was demonstrated that, apart from %V, there is considerable variability between languages within a rhythm class and non-variability of individual languages between rhythm classes. It was therefore concluded that the rhythm measures are powerful tools to distinguish individual languages but that the concept of continuous rhythmic variability rather than a dichotomous rhythm class is supported.

In chapter 5 it was then demonstrated that not only durational C- or V-interval variability can distinguish between rhythm classes but also speech rate. For this reason it was further investigated in chapter 6 to which extent speech rate and rhythmic measures are dependent on each other. It was demonstrated that rate did only affect the C-interval variability measures ΔC and rPVI. However, it could be demonstrated that after normalization for speech rate, ΔC and rPVI still significantly distinguished rhythmic classes, however, rPVI proved to be the more robust measures across all speech rates.

In summary it can be said that from all rhythmic measures %V and the normalized rPVI are the most reliable indicators for rhythmic class in the languages under investigation. They are not speech rate dependent; however, chapter 7 showed that such measures (exemplary for %V) can vary considerably as an effect of phrase and in some languages also as an effect of speakers. Since %V and rPVI have been shown to be the most robust rhythm class indicators on an acoustic level it should be assumed that these measures are the ones listeners use most to distinguish between rhythm classes or at least between languages that vary significantly for this value. In the present chapter this hypothesis is tested on two languages: French, representing syllable-timing, and German, representing stress-timing.

Stimuli from both languages will be delexicalized in a way that only durational cues

of C- and V-intervals remain. It will then be tested a) whether listeners can distinguish between the languages and b) on the basis of which acoustic cues a language distinction (if present) is performed.

In similar experiments by Ramus et al. (1999; see also chapter 1) speech rate between languages was highly controlled. As demonstrated in chapter 6 (under normalization of consonantal rhythm measures) Ramus and colleagues normalized their speech rate by choosing sentences of a roughly 3 second duration and 15 to 19 syllables per second. Compared to the BonnTempo data this leads to speech rates that can either be higher or lower than what speakers produce normally in a language. The fact that rhythm classes show significant differences in speech rate leads to the question whether C- and V-interval variability is perceptually salient in the same way when speech rate is not normalized for. This will be tested in the present chapter.

8.2. Perception experiments in speech rhythm

Perhaps the most important question about acoustic features of human speech is whether these features are perceptually salient, i.e., used to encode some type of meaning in the speech communication process. Differences between acoustic utterances of speech that are not salient can be regarded as rather unimportant by-products of the speech production process. Perceptually salient acoustic features do make a difference when modified or missing. For example, they can change the identity between two sounds and thus the meaning of two words if the sounds are phonemic. They can also carry information about paralinguistic features such as emotions.

Although speech rhythm has been studied exhaustively on an acoustic level, reports on perception experiments concerning speech rhythm are not very numerous in the literature. One of the reasons for this may well be that rhythm is a very difficult concept to incorporate in perception experiments. Subjects in perception experiments do not usually have problems with tasks asking which of two sounds appears higher or lower, longer or shorter, or which speech sample is faster or slower, etc. However, when asked to distinguish between more or less rhythmical, or in the context of stress

and syllable-timed rhythm, between more or less stress- or syllable-timed the necessary concepts are often not straightforward to phonetically naive participants. Miller (1984) tried to find a way around this problem in an experiment in which expert and non-expert listeners were asked to classify a set of 8 different languages according to their rhythm class. The expert listeners were regarded as possessing the necessary concepts while the non-expert listeners were trained with a tapping instruction. Miller found that both phoneticians and non-phoneticians showed consistencies in grouping some languages as stress- and others as syllable-timed but that no clear stress- and syllable-timing categories could be established. Another method is presented by Ramus & Mehler (1999) and Ramus et al. (1999) who use delexicalized stimuli for their experiments in which speech is made unintelligible while certain acoustic cues are still preserved. This allows the researchers to test how listeners respond to a particular cue in an experimental situation. One of the delexicalization methods of Ramus & Mehler (1999) used stimuli that preserved only rhythmical cues to speech by transforming all consonants into /s/ and all vowels into /a/ sounds while the fundamental frequency of the /a/ vowels is kept constant. This means that all semantic cues to word or sentence meaning as well as prosodic cues like intonation were not present in the signal any more thus listeners were only presented with durational cues of consonantal and vocalic intervals. Ramus & Mehler (1999) showed that subjects were able after a certain amount of training to distinguish stimuli of mora-timed Japanese from stress-timed English.

8.3. Method

The present experiment aims to find evidence for the perceptual salience of C- or V-interval variability and CV-rate on the basis of delexicalized stimuli from French and German. In the following the delexicalization method, listening subjects, choice of stimuli will be discussed in detail.

8.3.1. Delexicalization

> We leave it to the reader's imagination to invent other interesting manners to manipulate speech resynthesis. (Ramus & Mehler, 1999)

Methods for delexicalization are manifold (e.g. Heuft, 1999; Sonntag, 1999; Wagner, 2002). All delexicalization methods are characterized by a common problem that in some way can be called a paradoxon: the acoustic information that is to be judged by the listener is degraded to a more or less high degree. Early methods are based on extraction of fundamental frequency or a band-pass filtering between 70 and 200 Hz (Lehiste, 1979). More recent methods are able to maintain information about prominence by preserving parameters like amplitude and fundamental frequency together. The PURR-Method (Prosody Unveiling Restricted Representation; see Sonntag, 1999) transforms each fundamental period of a speech signal into a sinusoidal period of the same duration and amplitude and adds a 2nd and 3rd harmonic (with a ¼ and 1/16 of the original amplitude respectively). Voiceless parts of the signal are represented as pauses. According to Sonntag the signal thus contains all relevant perceptual prosodic characteristics like pitch, duration and loudness.

Linear prediction coding (LPC) based methods e.g. by de Pijper & Sanderman (1994) change the vowels of a speech signal into neutral vowels (Schwas) and makes consonants unidentifiable. The authors claim to preserve a high degree of acoustic-prosodic information with this method.

Both methods described above are valuable but do not fulfil the criteria required for the current experiment, which are a representation of the duration of vocalic and consonantal intervals without presenting other prosodic parameters like amplitude and fundamental frequency. For this reason a new method has been developed which uses elements from both methods described above. Derived from Sonntag (1999) the vocalic intervals were turned into sinusoidal signals with an added 2nd and 3rd harmonic, preserving the duration but not the fundamental frequency. The frequency of the fundamental was kept constant at 150 Hz. Derived from de Pijper & Sanderman (1994) consonantal intervals were made unidentifiable, here by turning them into white noise. The general idea of this delexicalization method is that listeners do rely on durational information only and do not perceive variation of the fundamental frequency and or intensity, nor do they perceive different vowel or consonantal qualities.

The delexicalization process was carried out automatically in Praat by the use of a Praat script. This script loops through all inter-pause stretches of each speech

utterance. It writes each utterance into a new audio file while it turns all consonantal intervals into white noise and all vocalic intervals into the PURR based complex tones. The new file is automatically named by the speaker and the inter-pause-interval (first occurring interval = 1, etc.).

8.3.2. Subjects

Subjects for the experiment were 18 listeners of English and German descent. Listener's age range was from 18 to 47 years; mean age was 29 years with a standard deviation of 8. There were 12 female and 6 male listeners. None of the listeners reported any hearing problems. Listeners were more or less acquainted with phonetic experiments. However, listeners were able to comment on the experiment while no comment expressing discomfort with the task has been received.

8.3.3. Stimuli

The stimuli for the experiment were chosen for speakers of French and German from the BonnTempo database. These languages were chosen because they can be reliably distinguished in the acoustic domain by all rhythm measures (see chapter 4). In order to receive comparable stretches of speech only phrases were chosen that are also intonation units, i.e. are preceded and followed by a pause. This ensured that naturally occurring phrase final lengthening features (which may to a great extent influence perception of regularity) was included equally in all phrases. For this reason a Praat script was written that automatically created a third tier for all relevant phrases that included the inter-pause intervals and calculated the values of the rhythm measures (ΔC, %V, nPVI and rPVI and varcoC) as well as the CV-rate (sum of the number of C- and V-intervals per second) for each inter-pause interval.

In order to receive a selection with a wide variety of extreme values of rhythm measures, the two highest as well as the two lowest (outliers not considered) ΔC and %V phrases (2*2*2 = 8 phrases) for each of the two languages (16 phrases) were chosen to serve as stimuli for the experiment. Additionally a selection of four random phrases for each language (8 phrases) was added to the original phrase collection (24 phrases in total; 12 per language). The variability of nPVI and rPVI has not been manipulated in the experiment since it was found that selection of

stimuli included a naturally wide variety of nPVI and rPVI values. A distribution of all stimuli in the ΔC/%V and nPVI/rPVI plain can be found in Figure 8-1. In both cases languages could be distinguished to a significance level of $p<.05$ in an independent samples t-test by the each of the rhythm measures.

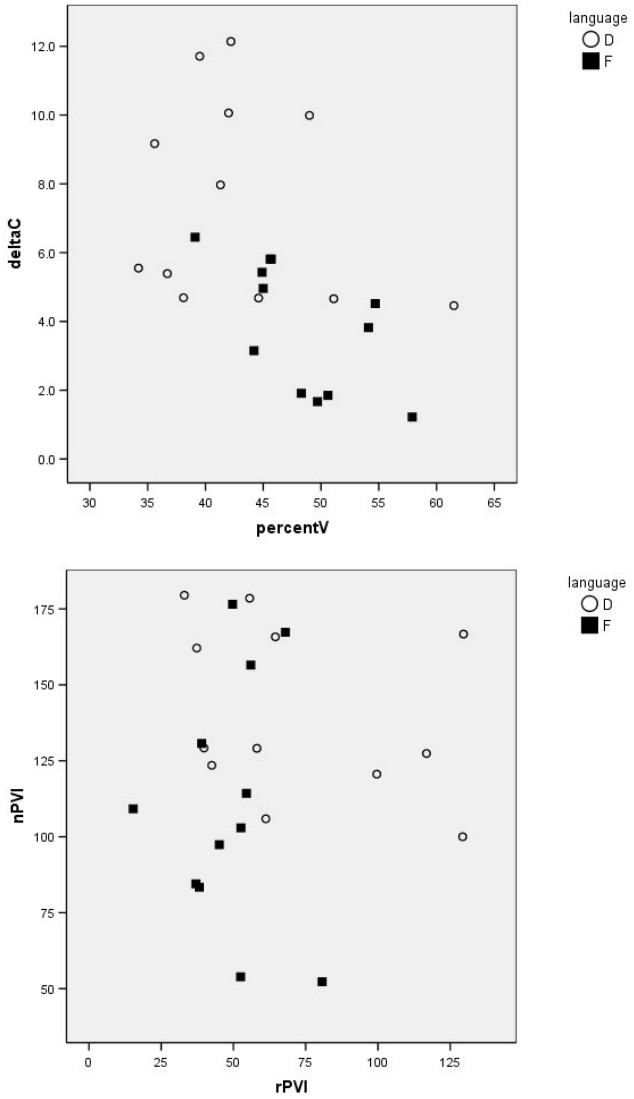

Figure 8-1: Distribution of stimuli along the ΔC/%V (top) and rPVI/nPVI (bottom) plains.

The number of V-intervals per phrase varied randomly across all selected stimuli between 7 and 12. This was due to a natural variability within the domain of the

inter-pause interval and also to choosing different intervals for French and German based partly on a random selection and partly on highest and lowest rhythmic values for the particular phrases,. It will be investigated below whether this variability had an influence on listeners' perception of the stimuli.

Speakers of stimuli varied between 6 (French) and 8 (German). This variability got introduced by some speakers in one language containing stimuli that both showed the highest and the lowest value of a particular rhythm measure (%V, ΔC).

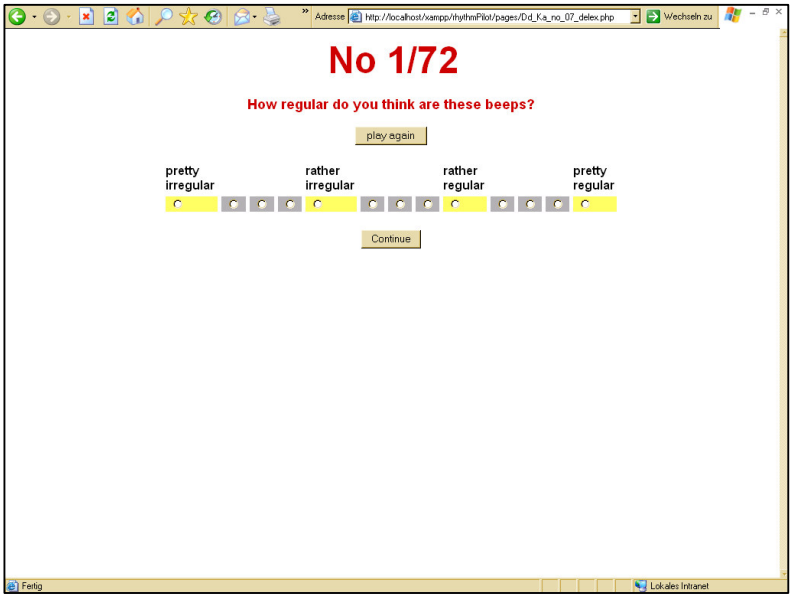

Figure 8-2: Web-interface for the perception experiment.

8.3.4. Procedure

The 12 French and 12 German stimuli (24 stimuli) were put in random order and played 3 times (72 stimuli in total). A php web-interface was written for stimuli presentation (see Figure 8-2). On this interface subjects were presented a stimulus automatically upon presentation of the page. Subjects were asked to rate on a horizontal 13 point scale (see next section for an in depth discussion of the scale) how regular they think the series of tones were. The scale was subdivided with landmarks from pretty irregular (leftmost scale point) to 'rather irregular' (scale point 5 from the left), to 'rather regular' (scale point 9 from the left), to 'pretty regular'

(rightmost scale point). It was made clear to listeners that the most extreme ends of the scale are ‘pretty irregular’ and ‘pretty regular’ and that the other points should be roughly in equal steps between these extremes. All German subjects in the experiment were fully literate in English and thus received the same interface; however, oral explanations were given in German language.

By asking for 'regularity' of the sequence subjects were given maximum degrees of freedom to decide which variability in the signal they choose as a predictor to rhythm class. Ramus et al. (1999) have shown that after considerable training listeners were able to distinguish between similarly delexicalized stimuli mainly on the basis of %V. In the present design subject were deliberately not trained to a particular parameter. The aim behind this design is to simulate a real variability discrimination situation more closely in which listeners are free to choose on which parameters they base their decision.

Unlimited replays of a stimulus were possible but in instructions prior to the experiment subject were encouraged to make their decisions quickly and intuitively after the first presentation of each stimulus. Before the actual start of the experiment subjects were presented 20 trial stimuli (including at least 1 example of highest and lowest representative of each parameter; ΔC and %V). During this trial phase it was expected that listeners cognitively adjust to the range of variability present in the stimuli.

To explain regularity to subjects the experimenter (author of the present work) produced quasi isochronous series of tones with his voice to simulate ideal regularity for subjects. Subjects were told that the tone sequences in the experiment never reached such an ideal regularity and that the aim of the experiment was to place the stimuli that come closest to ideal regularity on the left most part of the scale and the ones that are least close to ideal regularity to the right most part of the scale. Subjects were explicitly encouraged to make use of the entire scale.

8.3.5. Data assumptions and analysis procedures

The response scale used in this experiment is a so called ordered-response scale which has high similarities to a Likert scale. Such scales have originally been developed for psychological measurements of attitude (Likert, 1932) but are now

widely used for a wide variety of behavioral experiments such as speech perception experiments. Likert-type scales are bi-polar, thus they run from one extreme (here: pretty irregular) to the opposite extreme (pretty regular). Such scales typically have a neutral point (here: neither irregular nor regular) although forced-choice Likert-type scales exist which consist of an even number of items and thus force subjects to be on one side of the scale. Prototypical Likert scales have five response categories (e.g.: strongly agree, agree, neither agree, disagree, strongly disagree) although variants with 7 categories are common too.

In speech perception and hearing research Likert-type ordered-response data is commonly assumed to be interval scaled (Paul Iverson and Stuart Rosen, personal communication). It may be argued, however, that in particular for low-item Likert scales the number the cognitive distances between individual categories on the scale are not the same. As such it could be questioned whether the resulting data can be treated as interval scaled and whether standard statistical calculations of, for example, mean values and standard deviations, or inferential statistical comparison of groups using t-tests or ANOVAs, are justified. To ensure that the data deriving from the ordered-response scale used in the present experiment can be treated as interval scaled a number of measures have been undertaken and are explained in the following (8.3.5.1 – 8.3.5.4). In addition it was checked whether a normal distribution of the response data can be assumed (8.3.5.5).

8.3.5.1 Increased number of items on the scale:

On canonical Likert-scales typically five items are used as subjects' response categories. It can be argued that small numbers of items on an ordered-response scale may under-sample the response data. Thus a small number of items may lead to situations in which subjects have not enough options for differentiation. Subjects may then be forced to use the difference between two items to make a small distinction on one part of the scale while on another part of the scale they use the same distance for a large distinction. In addition, not the same distinction may be made for every response. Such a situation would inevitably lead to the data not being interval scaled. To avoid this, an ordered response scale of 13 points was used in the present experiment. On this scale subjects have room for fine grained scaling

between the four major anchors (pretty irregular, rather irregular, rather regular, and pretty regular).

8.3.5.2 Neutral point not marked:

In Likert-type ordered-response scales a neutral point is typically marked. Here such a neutral point would be 'neither irregular nor regular'. For the present experiment it was believed that such a point may be of disadvantage since there is no 'neutral' rhythmical pattern present. Subjects may inevitably have mistaken such a point and interpreted it with 'insecurity' rather than 'neutrality'. As such, subjects might have used this item for responses to stimuli they would find difficult to place on the scale although they were assuming them either regular or irregular. It was believed that not marking a neutral point contributes to the assumption that the data can be treated as interval-scaled.

8.3.5.3 Instruct subjects to use full scale:

A common problem with ordered-response scales is that subjects may be biased to use stay in a certain part of the scale (e.g. subjects may tend to remain neutral or tend to generally agree or disagree; see also below for another discussion). Subjects were explicitly instructed to try and use the full scale (see previous section).

8.3.5.4 Average data across all subjects:

As argued above, individual subjects may have a response bias. Such response biases may also lead to some parts of the scale being more carefully scrutinized than others. For example, if a subject responds preferably on the 'irregular' half of the scale, this person is likely to differentiate more between the irregular points of the scale than between the regular ones which inevitably speaks against the assumption that the resulting data can be treated as interval scaled since differentiations on the 'irregular' part of the scale are unlikely to be of the same magnitude as on the 'regular' part of the scale. To avoid such individual influences, responses of all subjects were averaged for each individual point of the scale.

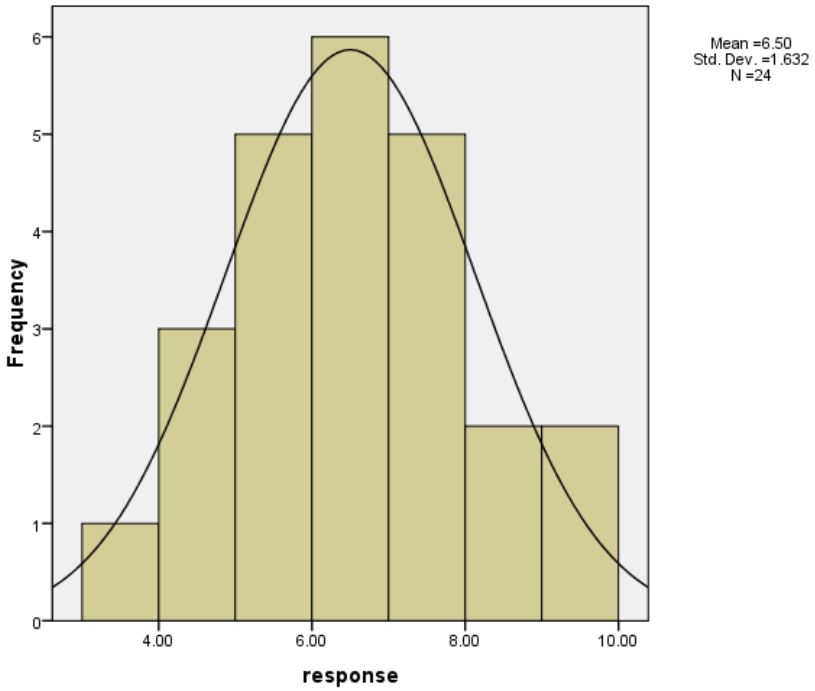

Figure 8-3: Histogram presenting showing the distribution of all responses collected via the ordered-responses scale from the present experiment. The superimposed black line shows an ideal normal distribution for this data.

8.3.5.5 Testing the distribution of the response data for normality

A common method for testing to what degree subjects behavior resulting from a ordered-response scale, like the one in the present experiment, was treated as interval scaled is by testing whether the distribution of the resulting response data is close to normal (or two underlying normal distributions in case of strongly bi-modally distributed data). For a distribution that is skewed and/or contains kurtosis standard statistical procedures should not be applied. For the present response data the distribution is plotted in Figure 8-3. The figure shows the frequency of a response (y-axis) as a function of the response itself (x-axis). A curve showing ideal Gaussian normal distribution for the underlying data has been superimposed as a black line. It is obvious upon visual inspection of this graph that the real distribution of the data overlaps highly with the predicted normal distribution (apart from a little deviation in the highest two bins). This visual inspection will be taken as sufficient to argue that

the data fulfills the criteria of normal distribution and are thus eligible for being analyzed using statistical procedures requiring normal data distribution.

Taken into consideration the above measures and the evidence from a preliminary data analysis it will be assumed that the response data in the present experiment can be treated as interval-level scaled data. For this reason statistical procedures like linear correlations will be applied which require interval-level scaling and a close to normal distribution of the data. Correlations were performed between subjects' responses and a number of rhythm and rate measurements. The significance of the correlations were tested using ANOVAs. It will be assumed that if there is a significant correlation between listeners response and an acoustic parameter (rhythm or rate) that subjects will have used this acoustic parameter for the overall judgment of 'regularity' of the stimuli.

8.4. Results

In a first step it will be analyzed whether the languages French and German were distinguished by the listeners on the whole on the basis of C and V durational information only. This will be done by testing whether there is a significant difference in the distribution of listener's response as a function of language. If the interval variability is lower in German this may be reflected in the fact that German receives significantly higher number of low-regularity responses than French.

After that it will be analyzed which of the interval variability or rate measures is the best predictor for listeners' response. The general idea for this analysis step is to perform a linear regression with each measure as a predictor for listeners' response. The strength of the correlation (reflected in R square) will be taken as an indicator to how well listeners used the respective measures in their decision making process.

8.4.1. Language separation

Figure 8-4 shows a box plot of the group results for the responses as a function of the language condition. It can be seen that the German stimuli were rated as more irregular than the French stimuli. An independent samples t-test with 'response' as the test variable and 'language' (German and French) as the grouping variable revealed a significance level of $p<.001$ for the two means to be not equal. It can therefore be

concluded that overall listeners could separate the languages German and French well on the basis of durational cues of C- or V-interval variability only. According to this French was perceived by the listeners as more regular. In the next sections it will be investigated which variables were mostly responsible for the separation.

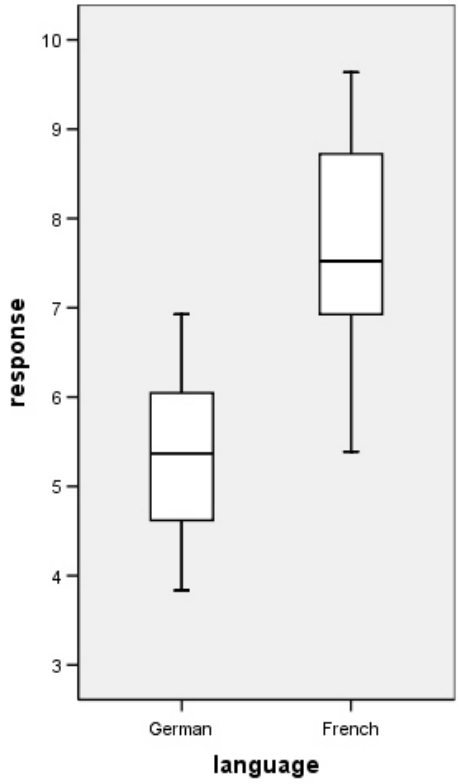

Figure 8-4: Box plot containing the group results of responses (average response/stimulus) for German (D) and French (F). Possible responses reach from 1 (pretty irregular) to 13 (pretty regular), however, the range of average responses was only between about 4 and 10 on the scale.

8.4.2. Influence of the number of intervals in a stimulus

As mentioned in the methodology the number of V-intervals was not controlled for in the stimulus selection and thus varied randomly across stimuli. This may have a direct influence on the results. It may be that listeners were cognitively higher challenged when stimuli contained more tones. For this reason it is possible that they were less able to distinguish between the durational variability of individual intervals

in longer stimuli and as an effect rated these stimuli less variable. Such an effect would be observable by a positive correlation between for example the number of V-intervals (as an interval number measure) and listeners' response.

With respect to Figure 8-5 such a hypothesis must however be falsified. The graph plots listeners' response as a function of the number of vocalic intervals (N_V). Judged by the distribution of the dots across the graph there is no correlation between the two parameters which is confirmed by a low R square value of 0.011.

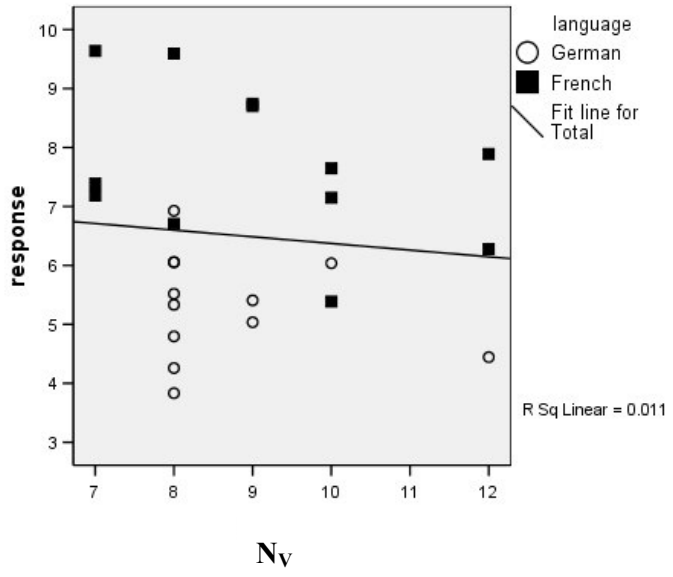

Figure 8-5: Listener's response as a function of the number of V-intervals (N_V) contained in a stimulus.

In addition, a possible correlation between response and the number of intervals in a phrase could not have influenced the above results that listeners distinguished between French and German. Although most of the German stimuli contain between 8 and 9 V-intervals per phrase and French stimuli between 7 and 12 (see Figure 8-5) no significant difference could be obtained for the number of intervals in a phrase between the two languages French and German ($p=0.51$ for an independent samples t-test with number of V-intervals as a dependent variable and language as a grouping variable).

In conclusion it can be stated that the number of V-intervals in a phrase has no influence on listeners' perception of regularity in the delexicalized stimuli. Further there is no basis for the assumption that the number of intervals in a stimulus could have contribute to listeners' ability to distinguish between the languages German and French on the basis of these stimuli.

8.4.3. Listeners' response as a function of rhythm measures

Results for the Ramus et al. (1999) interval variability measures %V and ΔC and for the Grabe & Low measures, nPVI and rPVI, can be found in Figure 8-6. The figures plot the values for each respective variability measure as a function of response and superimpose a line of best fit for the total score in each plot. The strength of correlation (expressed by R square) between each of the variability measures and listeners' response suggests to what extent listeners' made use of the respective variability to judge 'regularity' of the tone sequences in each stimulus.

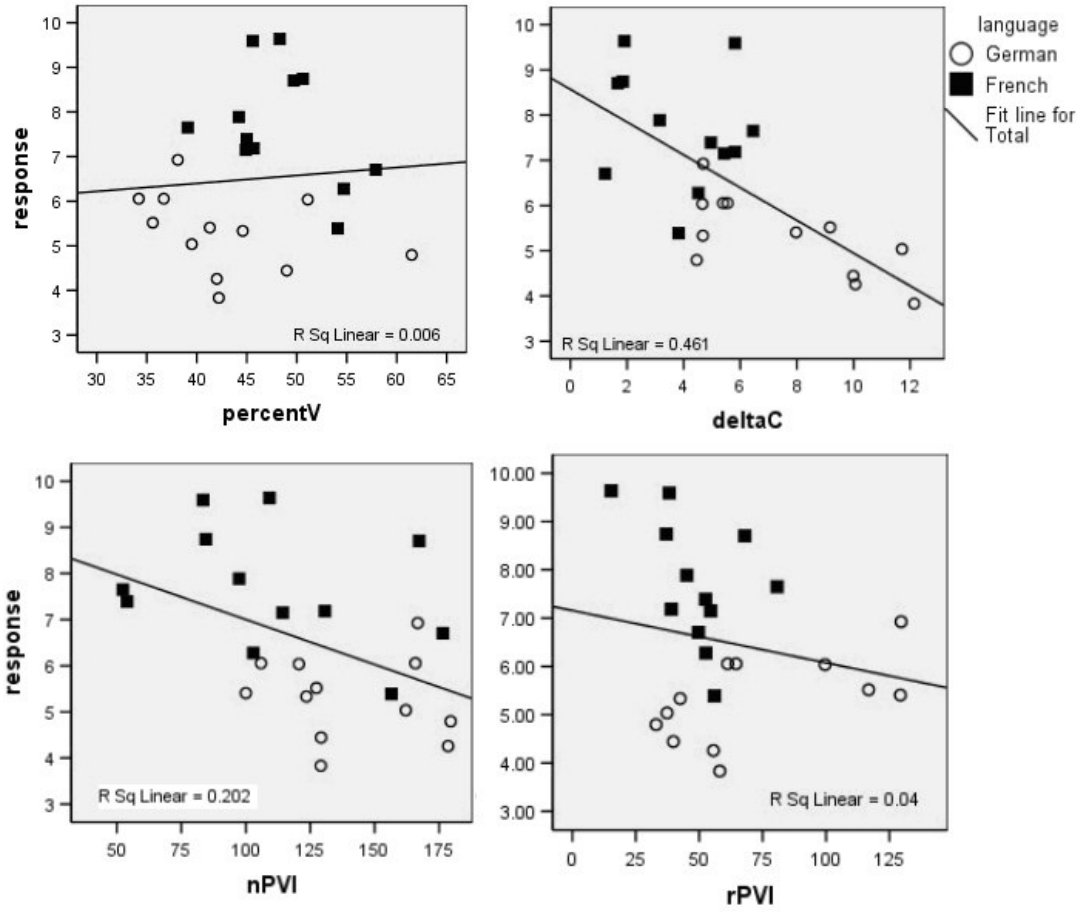

Figure 8-6: Cross plot of the interval variability measures %V (top left), ΔC (top right), nPVI (bottom left) and rPVI (bottom right) as a function of listeners' response (dots represent averages for each of the 12 stimuli per language). Superimposed is a line of best fit for a linear correlation between the two parameters.

For the Ramus measures it can be seen that there is no correlation between response and %V (R square = 0.006) but there is some correlation between response and ΔC (R square = 0.461). This means that in the case of ΔC nearly half its variance can be predicted by response. This relationship is considerably weaker for when ΔC is calculated on the basis of the ln-transformed data (ΔCln; not plotted). The linear R square for response and ΔCln is 0.295 which means that only about a third of the variability of ΔCln can be predicted by listeners' response.

For the Grabe & Low measures the situation is reversed. In case of the V-interval

variability a low correlation could be detected between response and nPVI (R square linear = 0.202). In the consonantal domain no correlation could be found between response and rPVI (R square = 0.04). After a rate normalization of nPVI (not plotted) the correlation does not increase (R square = 0.044).

Overall the correlations show that C- and V-interval variance can only be poorly predicted by listeners' response. A medium strong predictability in the case of ΔC loses strength after the data has been normalized for speech rate. This already is a sign for the possible fact that speech rate has an influence on perception. The results for the correlation between CV-rate and listeners' response will be described in the next section.

8.4.4. Listeners' response as a function of CV-rate

Figure 8-7 contains the results for CV-rate. The plot shows CV-rates as a function of response; each dot represents the average for one of the 24 stimuli (12 stimuli per language). The strongest correlation between listeners' responses and an acoustic correlate could be found for CV-rate where about two thirds of the CV-rate variance can be explained by response (R square = 0.655). This means that listeners' interpretation of regularity was to the most part influenced by rate differences between the stimuli.

However, the graph shows that languages differ considerably in linearity. While there is a strong linear increase in response with an increase in speech rate, this pattern stops at a CV-rate of about 11 C- and V-intervals per second. At about this point languages change from German to French. This gives rise to the assumption that there is less of a correlation between response and CV-rate in the case of French than there is in the case of German.

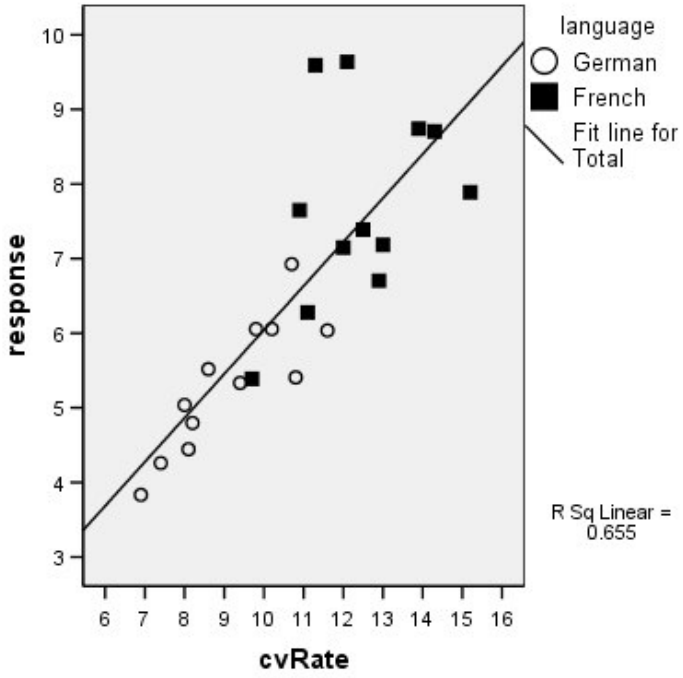

Figure 8-7: Listener's response as a function of laboratory measurable speech rate (CV-rate: the sum of the number of C- and V-intervals per second). Dots represent average responses for each stimulus.

Figure 8-8 finds support for this assumption. The box plots display the data of Figure 8-7 split up by the languages German (left) and French (right) with superimposed lines of best fit. For the German stimuli a strong correlation between response and CV-rate could be found (R square = 0.733) whereas the correlation in the case of French is rather poor (R square = 0.158).

From the results presented in this section it can be concluded that CV-rate is a good predictor for listeners' response. At rates higher than 11 intervals per second however, predictability becomes poorer. It happens that the class difference between German and French stimuli is also at about this rate point.

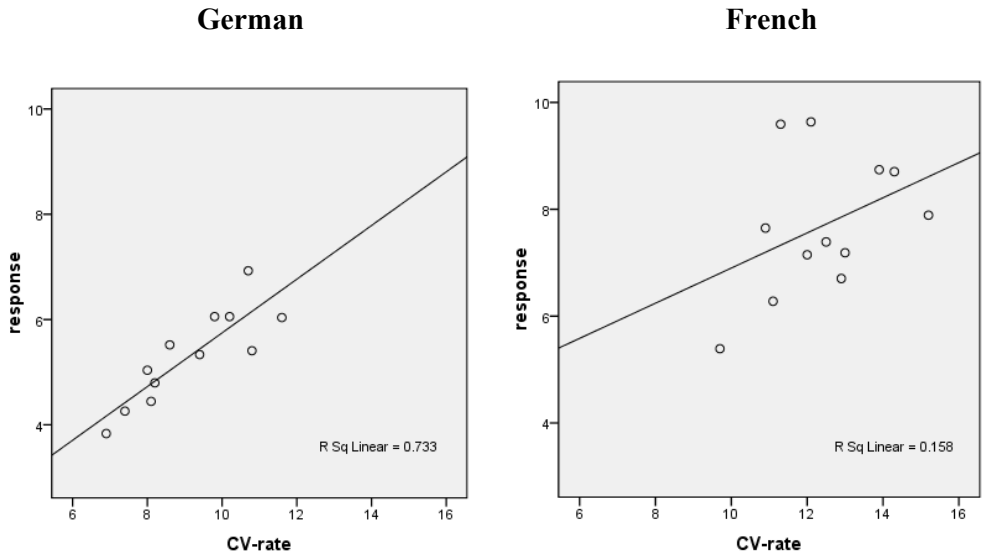

Figure 8-8: Scatter plots of CV-rate as a function of response for the languages German (left) and French (right) with superimposed lines of best fit.

8.4.5. Discussion

The language discrimination results presented in this section are in some respect well in accordance with results from perception experiments in Ramus et al. (1999) who found that languages of different rhythm classes can be distinguished on the basis of C and V durational cues only after listeners had a considerable training in discriminating delexicalized stimuli and speech rate variability between the stimuli was minimal. The present experiment replicated the fact that languages from different rhythm classes can be distinguished by listeners based cues on C- and V-interval variability.

The present experiment, however, varied considerably from Ramus et al. in that it did not train listeners on the type of stimuli used but asked them to rate stimuli according to the regularity of tone sequences obtained from the delexicalization of real speech stimuli. The interpretation of 'regularity' was left to the listener. In addition to that the stimuli in the experiment were not normalized for speech rate but revealed natural variability which is typical for the languages German and French when reading a normal text. For the stated differences the results from Ramus et al.

and the present experiment are difficult to compare directly. However, the data presented here reveals important details about how listeners make distinctions about rhythmic variability between languages when durational variability parameters and speech rate vary together. The data shows clearly that when natural speech rate fluctuations are allowed, listeners make their distinction on CV-rate rather than the durational variability of C- or V-intervals.

Durational C- or V-interval variability is, however, not a neglected cue by listeners as nPVI and ΔC reveal low correlations with listeners' responses. It was shown though that for speech rate dependent measures like ΔC a considerable amount of its prediction power is lost when the underlying data is normalized for speech rate. As opposed to that the rPVI measure remains an equally poor predictor with and without speech rate normalization.

A surprising observation is that %V is amongst the poorest predictors for listeners' response. This finding is unexpected since %V has been shown to be a measure that distinguishes best between rhythmic classes for the languages under observation (see chapter 4) and it has been shown to be robust against changes in speech rate (see chapter 6).

Although this experiment provides insights into the perception of durational variability of C- and V-intervals between languages, the data is far from allowing a comprehensive statement. Interval duration perception is influenced by far more factors than absolute interval duration only. O'Connor (1965) showed for example that the sequence of presentation has a high influence on the perception of regularity. In an experiment with irregular click intervals subjects judged exactly these intervals as regular in which the second interval was longer. For our stimuli this would mean that if subjects are to detect a variability difference between consecutive intervals, a significant increase between the first and the second interval needs to be present. Without such an increase the perception of a durational difference is probably not given. The rhythmic variability measures presented in the current work do not take account for sequencing differences. Whether a long interval is followed by a short interval or the other way round is not captured by the measures.

9. The production and perception of rhythm in second language speech

9.1. Introduction

The idea for the following experiment arose during work with the second language data in BonnTempo (L2 data, see chapter 3 for an overview of what data is available). Rhythmic differences between the L1 and the L2 domain are not topic of this work, so they will not be elaborated any further. However, a major finding that has been made in a comparison study between L1 and L2 rhythm created a unique environment to test the perception of rhythmic variability further without using delexicalized speech as in the previous chapter. This finding was that from all rhythmic correlates only %V is significantly different between native and non-native speakers of a language. This finding will be illustrated in the following for the German L1 speakers in BonnTempo reading the English and French text versions. In the present chapter it will be tested whether these differences are perceptually salient in the way that they contribute to the quality of pronunciation of the L2 speakers. If, as it was often argued in the early rhythm work (see chapter 2 and discussion below), rhythm is the main prosodic feature that is violated in L2 speech (Lloyd James, 1929), it should be assumed that when listeners are asked to rate the quality of pronunciation their judgement is heavily influenced by speech rhythm. In the present experiment, speech rhythm is measured by the use of rhythmic correlates as suggested by Ramus et al. (1999) and Grabe & Low (2002). If the change of %V between L1 and L2 speech reflects a change in speech rhythm between the two pronunciation varieties, native listeners should classify the pronunciation of L2 speakers as poorer if their %V deviates more from that of the average native speaker. This hypothesis was tested in the present chapter in a perception experiment in which native listeners of French and English had to rate the pronunciation of Germans speaking French and English respectively. It is hypothesised that a positive correlation between %V and listeners' ratings is obtainable for the case that a) changes in speech rhythm between L1 and L2 speech are perceptually salient and b)

these changes of speech rhythm are accurately reflected by %V.

In the following the acoustic differences between L1 and L2 speech will first be illustrated in more detail. On the basis of the results from the perceptual experiment will be analysed and discussed.

9.2. Acoustic correlates of speech rhythm in L2

The data used for the analysis of L2 speech rhythm is taken from the BonnTempo-Corpus. For the present study only the recordings from the reading condition were used in which readers were asked to read a text normally. Since speech rate was demonstrated to have an influence on the perception of rhythmic variability (see chapter 8) it was assumed that choosing speech rates across the whole intended speech tempo spectrum in BonnTempo would cause an additional amount of confusion. Stimuli in the normal reading version, however, also show variability in CV-rate. With the same argumentation as in chapter 8, this natural CV-rate variability was not normalized for. If the variability of %V should turn out to be perceptually salient then this saliency is only a valuable indicator for rhythmical differences between native and non-native pronunciation if it can withstand naturally occurring speech rate fluctuations.

Table 9-1 gives an overview of the numbers of speakers and the total number of syllables, C- and V-intervals analyzed in this chapter.

language	speakers	syllables	C-intervals	V-intervals
English (E)	7	2684	2475	2444
French (F)	6	2734	2420	2455
German (G)	15	5698	5028	4832
German speaking English (Ge)	8	3087	2873	2833
German speaking French (Gf)	8	3504	3119	3151

Table 9-1: Overview over the number of speakers, syllables, C-, and V-intervals for each of the language conditions used.

9.2.1. Comparability of the L2 proficiency of the different groups

The criterion for selecting a native speaker for BonnTempo was that this speaker needed to be a native-speaker of a standard variety of this language. Dialect speakers have not been considered but speakers do sometimes vary accentually according to regional pronunciation varieties of the standard. For the non-native speakers no record has been taken of the proficiency level of the speakers in the second language since it was felt that formal proficiency levels may be easy to define but usually do not tell us a lot about the homogeneity of the group.

The problem that arises is whether the groups are at all comparable. Especially with the number of eight L2 speakers per group there is some chance that we picked rather bad speakers for one language and rather good ones in the other. This would certainly have a major influence on our results since we would expect better L2 speakers to perform generally better on L2-rhythm. The solution that we found here is to have L1 speakers of the particular languages rate the quality of the L2 speech that our speakers produced. The methodology for gathering this data is identical with the methodology for the perception experiment below, where it is described in detail. Native speakers of English and French were asked to rate the quality of the speech of Germans speaking English and Germans speaking French respectively on a scale from 1 (poor) to 10 (fluent) with an additional value (11) in case they thought the speaker was native of the language.

The results for this analysis are shown in Figure 9-1. The left graph of the figure is a box-plot showing the distribution of the quality rating of the two native listener groups English (left box) and French (right box). The right graph is a plot of the mean values (circles) for the two groups with their standard error (t-bars). Both graphs give evidence that the quality of speakers is rather comparable between both groups. The French native listeners have a slightly lower median and mean (7) than the English listeners (8), which shows that the French listeners rated the proficiency of French speaking Germans as slightly poorer than the English listeners the English speaking ones. This is also supported by the fact that the range of the data for English listeners does not include value 1 (i.e. none of the English L2 speakers was at the absolute low end of the quality scale) while on the French scale this label was used. The small standard errors for the data observable from the left graph indicate that the

mean values obtained here are highly representative. An independent samples t-test reveals that the difference between the two mean values is significant ($p<.05$). However, the absolute difference is considered rather small thus the proficiency levels of both groups are regarded to be sufficiently comparable for the present study. Nevertheless, for the interpretation of the results (see discussion) the fact that French are slightly poorer rated will be considered.

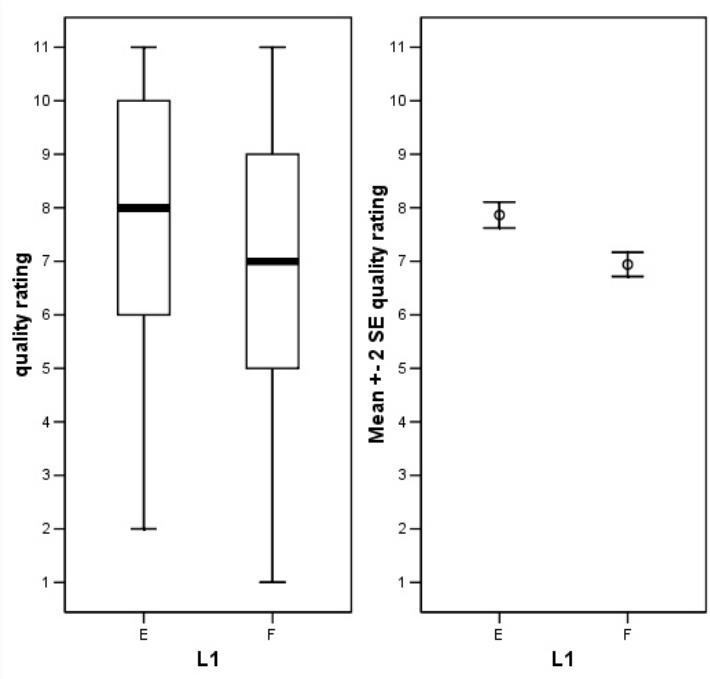

Figure 9-1: Box plot (left) and mean with error bar graph (right) to compare L2 proficiency of the non-native speakers. Germans speaking English were rated by English natives (E) and Germans speaking French were rated by French natives (F); (1 = poor quality, 11 = native speaker).

9.2.2. Results for the acoustic measurements of speech rhythm and rate

The results of the rhythm measurements for the native groups English (E), French (F), and German (G) and the non-native groups, Germans speaking English (Ge) and Germans speaking French (Gf) are presented in Figure 9-2. The left graph shows the Ramus et al. (1999) measures, ΔC as a function of %V, the right graph the Grabe & Low (2002) measures, rPVI as a function of nPVI. The centre of the crosses shows

the mean value, the arrows the standard deviations of the respective groups and measures.

For the native groups E, F, and G the results are the same as presented in chapter 4 for the rhythm measures where the level of significance according to which languages vary are analysed. In case of the non-natives Ge and Gf, both the Ramus and Grabe & Low measures reveal that the non-native groups are comparatively higher on the V-interval variability axis (%V and nPVI). On the C-interval variability axis (ΔC and rPVI) non-native speakers perform as well as native speakers. T-tests comparing the variability of Ge with E and Gf with F for all rhythm correlates (%V, ΔC, nPVI, rPVI) reveal that the deviations in vocalic variability between native and non-native groups are highly significant ($p<.001$) but that there is no significant difference between these groups for the consonantal variability measures (ΔC, rPVI).

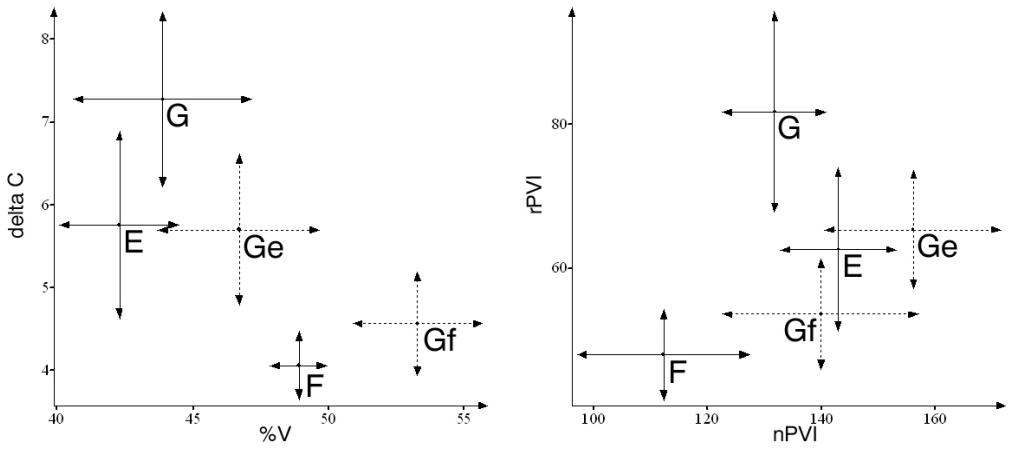

Figure 9-2: Acoustic correlates of speech rhythm for the native groups E, F, and G and the non-native groups Ge and Gf. The left graph shows the Ramus et al. (1999) measures %V and ΔC; the right graph the Grabe & Low (2002) measures rPVI/nPVI.

Figure 9-3 shows the results for speech rate (sum of the number of C- and V-intervals per second) in a box plot for the native and non-native groups under investigation (E, F, G, Ge and Gf). This data has been exhaustively discussed for the native groups in chapter 5. In summary, for these groups it shows that G produces the lowest CV-rate, F the most and E somewhere in the middle. A post-hoc analysis to an ANOVA reveals that there is a significant difference between G and F ($p=0.001$)

but not between G and E (p=0.267). Concerning the L2 groups it appears from the graph that G speaking either English or French do not vary very much compared to the rate of their native tongue. A t-test shows no significant difference between any of comparison pairs (G-Ge: p=1; G-Gf: p=0.776). Comparing the L2 speakers to the respective native speakers though it seems that Ge manage to imitate the speech rate of their native peer group better than Gf do. This is also supported by a t-test showing significant difference between Gf and F (p=0.013) but no significant difference between Ge and E (p=0.281). In other words this means that Germans are better at imitating the speech rate for English than they are for French. This may have to do with the observation that the native speech rate of German is more similar to English. This, however, is a question that will not be analyzed further in the current context.

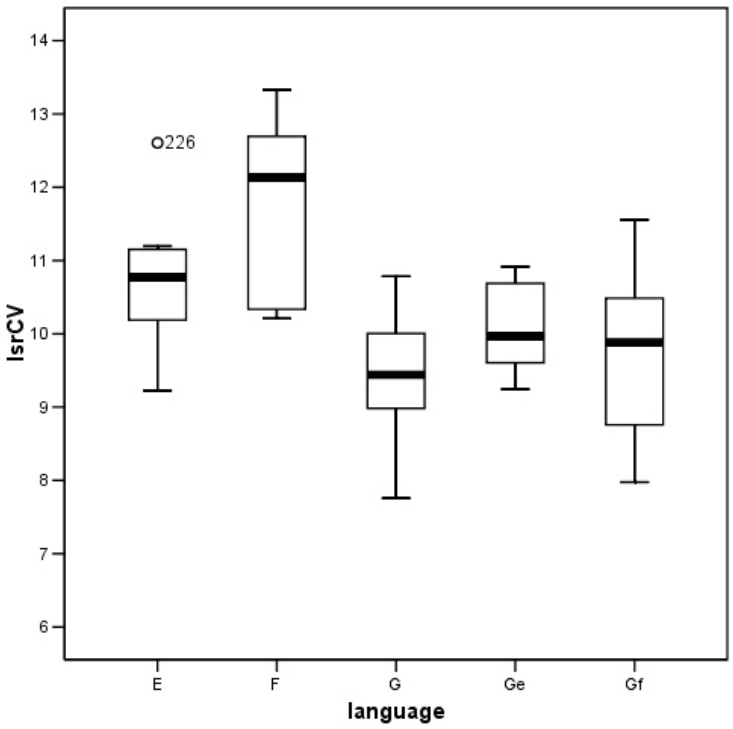

Figure 9-3: Box-plot of the distribution of CV-rate (sum of the number of C- and V-intervals per second) for the different native groups E, F, G and the non-natives Ge and Gf

9.2.3. Discussion of acoustic measurements in the L2 domain

In conclusion this section showed that vocalic variability measures %V and nPVI are significantly higher for non-native speakers than they are for native speakers. This is true for Germans speaking French and English.

For the C-interval measures, rPVI and ΔC, no variability could be detected between native and non-native speakers. This result seems surprising but may be interpreted according to the following: In case of the C-interval complexity German and English are probably rather similar while French is less complex than both these languages. For this reason Germans should be well able to perform the consonantal complexity of English but should also have no problems doing this for French. The C-interval measurements should therefore not vary too much between the native and non-native groups. This explanation implies that speakers of languages like French that does not possess consonant clusters with a comparable complexity like English or German should not perform well on the C-interval variability which should be reflected in both a higher ΔC and nPVI. First results drawn from the BonnTempo-corpus on two native speakers of French speaking English show much higher ΔC values for French than for English. The amount of data though only shows tendencies and no statistical significance. With a constantly growing database we expect to be able to report more detailed answers in the future.

For the V-interval variability the rhythm class hypothesis does not support the current findings. When Germans speak English they do perform a language that is supposedly in the same rhythmic class as their mother tongue, when they speak French they speak a language from a supposedly different rhythm class. If rhythm classes reveal rhythmically similar structures in their languages it should be assumed that Germans have less problems to produce the rhythm of English than they have with the rhythm of French. The present results, however, indicate that they have similar problems since %V and nPVI vary in similar amounts in both L2 conditions. This assumption though is only true under a further assumption, namely that the rhythmic correlates %V and nPVI reliably indicate changes in speech rhythm. It further assumes that equal absolute changes in %V and nPVI as obtained here, are also produced by the equal absolute amount of variability. This, however, does not need to be the case. Since the durational variability of C- and V-intervals in French is

per se smaller than in English (see chapters 4 and 6) it could be argued that the same absolute changes in variability in both languages lead to a higher amount of change in variability for French than for English. In order to test this, a more detailed study of variability would however be necessary. Durational variability of C- or V-intervals can be reflected on multiple levels and it is not necessarily the case that equal changes have been performed in French and English by the German natives in the present case.

If the difference in V-interval variability between stress- and syllable-timed languages is mainly reflected in the use of vowel reductions, German speakers, who do use vocalic reduction in their L1, should perform well producing vowel reduction in English. In the case of French, which does not use vowel reduction, one may assume that the vowel reducing Germans force this concept in one or the other way on French. If that was the case the measures for V-variability should indicate this by showing lower values for the non-native speakers, which would make them move closer to the stress-timed. Clearly this assumption is not supported by our results. In case of the V-interval variability German natives seem not to apply German type vowel reduction on French and they seem not to be able to perform well on English vowel reduction systems either. This finding lets us conclude that there must be other factors than vowel reduction that are violated by Germans speaking English and French. In order to interpret the finding further we would first need to know whether the degree of vocalic variability in both English and French L2 is specifically German or whether speakers of other native languages performed similar. Again, the current data situation in BonnTempo does not allow us to make clear statements at this point but we do find tendencies from other languages that show us that this is not the case. We would also need to see whether Germans perform similar in other language conditions.

Although all these questions raised here in the discussion are interesting in nature for future research they lead too far away from the path of the main question of this chapter which is to see how much the changes in vocalic variability produced by the Germans are perceived by native listeners of the respective languages English and French.

The findings for the rate differences can be given a straightforward interpretation. If

the CV-rate was mainly dependent on the structural complexity of C- and V-intervals, then non-native speakers of a language in general should be able to produce similar rates to native speakers since they use the same C- and V-intervals. Of course non-native speakers may struggle more with a correct pronunciation of C- and V-intervals in foreign languages and may therefore tend to produce less intervals per second but then why is our Gf group so bad at imitating native-like speech rates when structural C- and V-interval complexity is rather low compared to their native language and why our Ge group better at getting the right speed in a language with more complex C- and V-interval clusters? The answer for this may be that speech rate in our current example is to a certain extent an inherent property of the native language and is transferred to L2 conditions. So in other words: It seems conceivable that Germans produce a comparably slow CV-rate in French because they are used to doing that in their native language.

9.2.4. Conclusions

In conclusion our main findings from the acoustic measurements are that there is a significant difference between L1 and L2 speakers of French and English in the case of the V-interval variability measures %V and nPVI. We also found that Germans produce French significantly slower (in terms of CV-rate) than French natives. This is not the case between Germans speaking English and English natives. In the next step it will be investigated in a perceptual experiment whether these changes are perceptually salient.

9.3. Perception of L2 speech rhythm

In chapter 8 two methods of perception experiments in the rhythm domain were discussed. It was a Miller's (1984) method using real speech samples assessed by trained naive and phonetic expert listeners and Ramus et al.' (1999) study using delexicalized stimuli which was also applied in the previous chapter. Both methods bear disadvantages. In Miller's method listeners' responses may well be influenced by their knowledge of which language is supposed to belong to which rhythm class. In Ramus' method speech is reduced to durational cues only while other possibly important rhythmic cues are not present any more.

In the present chapter a methodology is introduced that does not rely on the listeners' knowledge of rhythm classes and uses real speech. It is only possible in certain frameworks like the one presented here with L1 and L2 speech but it is considered a powerful method to investigate whether measurable speech rhythm variation is perceptually salient. Subjects were not directly asked in the experiment whether they can hear a difference in rhythm between L1 and L2 speech but whether they can perceive a difference in the pronunciation quality of the L2 speech in terms of which speech samples sound more or less native-like. It was assumed that speech samples of L2 speech that vary highly in measurable rhythm from speech samples of L1 speech should be considered of poorer quality by native speakers of the language under investigation (see introduction to this chapter).

With respect to the acoustic measurements of L2 speech rhythm it was found that there was statistical significant variability only in case for the V-interval variability measures %V and nPVI. For this reason it should not be assumed that the consonantal measures ΔC and rPVI are in any way salient. However, for the reason of completeness they were included in the analysis too.

9.3.1. Method

Subject for the experiment were 5 native English (academic staff and students of the University College London) and 9 native French listeners (academic staff and students of Lyon University). The stimuli for the experiment were selected from the exact same speech material for the non-native speakers that were used for the acoustic measurements above. Four equal sentences have been extracted for each of the 8 Ge (8, 10, 10 and 11 phonological syllables) and 8 Gf (8, 10, 12 and 13 phonological syllables) speakers. In chapter 7 it was shown that %V can vary significantly between phrases. For this reason it was ensured that samples of the non-native speakers were chosen which where representative for a higher %V in comparison to their native counterparts. The 64 stimuli (2 languages * 4 sentences * 8 speakers) were put into random order and presented on computers using a php scripted web interface. On some initial pages subjects received a brief introduction about the rating before they started with the experimental pages. One stimulus was presented on each page together with a rating scale. Stimuli were presented instantly

when subjects called the respective page but unlimited replays were allowed with a 'replay' button. Subjects were asked to rate on a scale from 1 to 10 how good they thought the pronunciation of the L2 of the respective speaker was with '1' signifying 'poor' pronunciation and '10' fluency. If they thought that a speaker had native like pronunciation they had an additional option to indicate this (value 11 in the experiment).

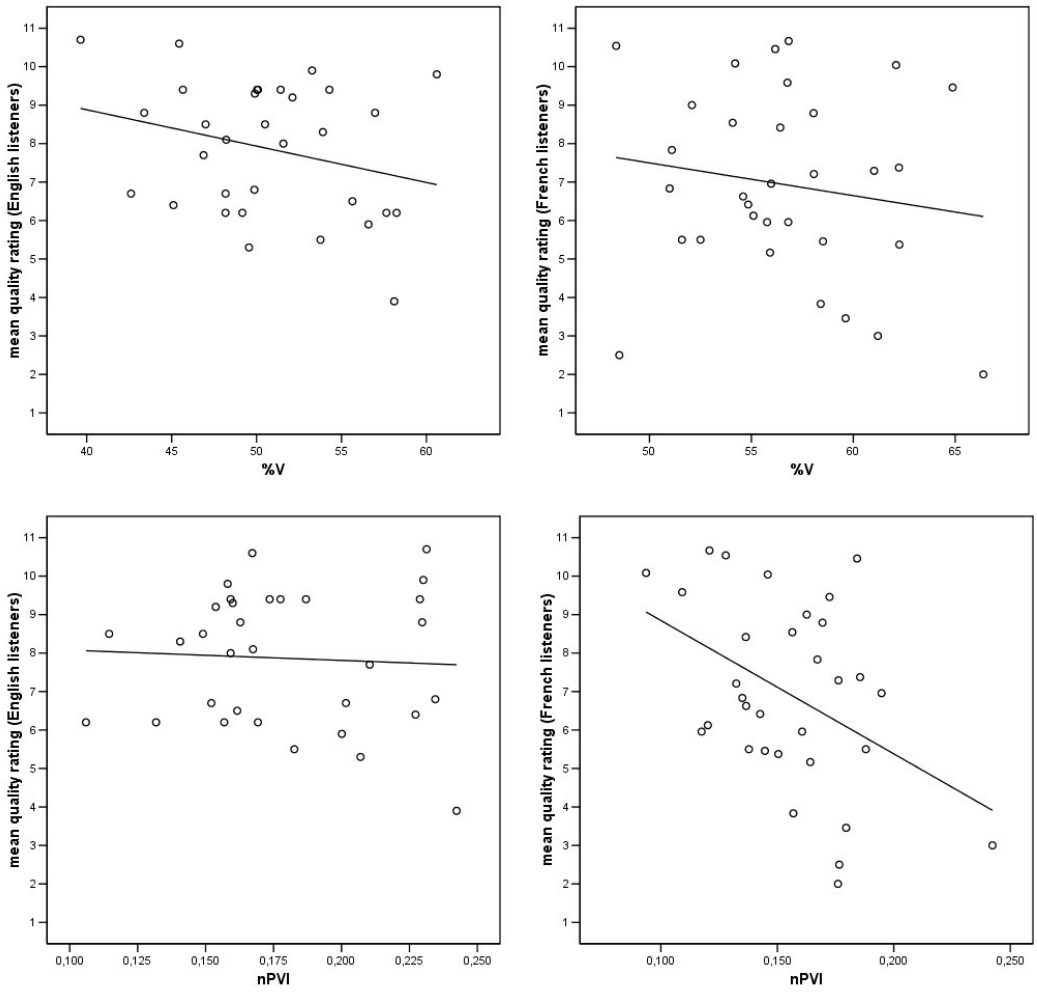

Figure 9-4: Quality rating plotted over the V-interval variability measures %V and nPVI for English (left) and French (right) native listeners.

9.3.2. Results

The results of the perception experiment are presented in Figure 9-4 for the V-interval variability measures %V (top graphs) and the nPVI (bottom graphs), in Figure 9-5 for the C-interval variability measures ΔC (top graphs) and rPVI (bottom graphs) and in Figure 9-6 for the speech rate measure CV-rate (sum of the numbers of C- and V-intervals per second). The graphs on the left hand side in each figure show the data for the English, on the right hand side for the French native listeners.

All graphs are scatter plots in which each dot shows the mean rating for each sentence of each speaker across a rhythm or speech rate measure. In each graph a line of best fit has been added. If this line shows a rather steep slope with the points scattered closely around it, a liner relationship between the two variables will be argued for. In case a linear relationship was assumed, r square was calculated (and the probability (p) for r square to be significant).

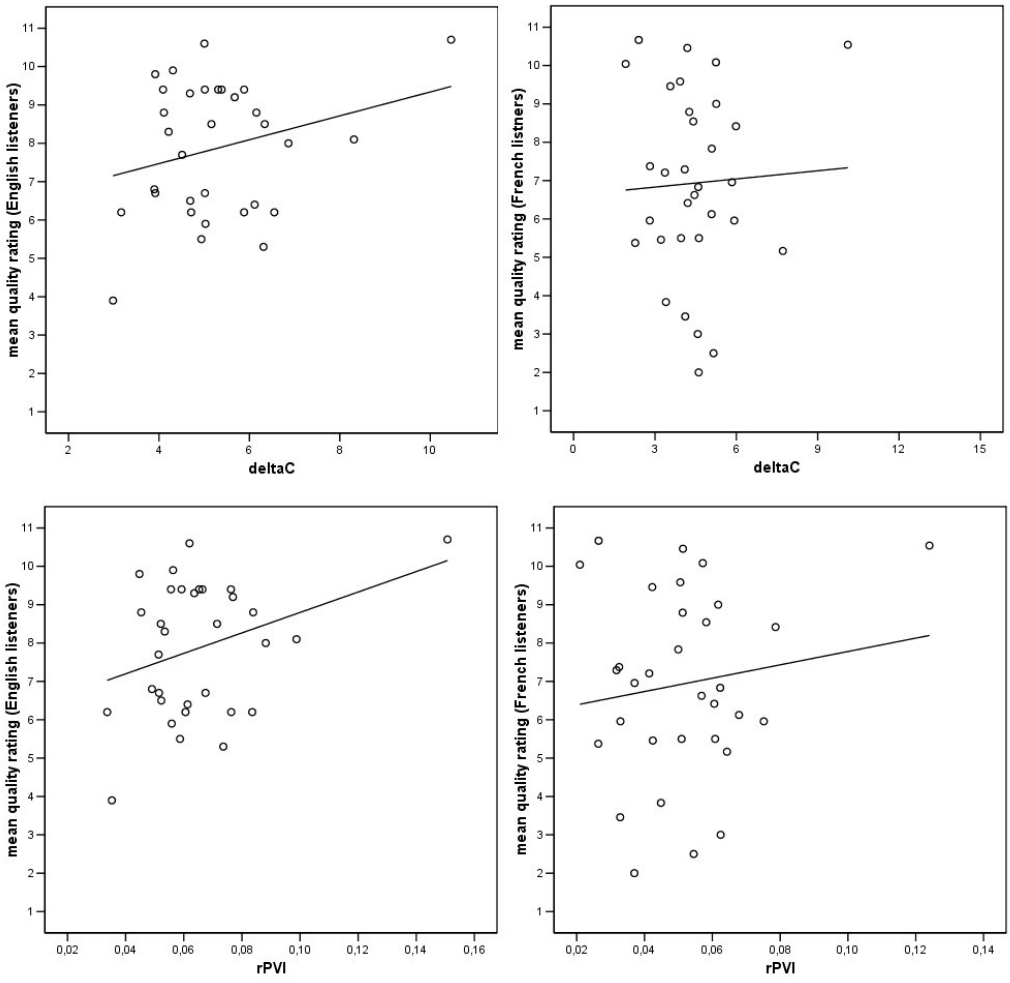

Figure 9-5: Quality rating plotted over the C-interval variability measures ΔC and rPVI for English (left) and French (right) native listeners.

For the C- and the V-interval variability measures it can be seen in the data plots that relationships between listeners' response and any of the rhythmic correlates is very poor. All lines of best fit have the tendency to horizontal slopes and in the case of the consonantal measures the data is rather concentrated around a certain area but not along a line. Slightly non-horizontal slopes proved to be the result of outliers in the data (e.g. figure 6, rPVI for English listeners, the value in the far top right corner). These results support the view that neither C- nor V-interval variability, as measured by the Ramus and Grabe & Low measures, are perceptually salient correlates of pronunciation quality.

Nevertheless a linear relationship could be detected in case of the way French listeners perceive speech rate (Figure 9-6, right). The graph shows that the faster the speakers uttered the stimuli the higher their quality was rated. The multiple correlation coefficient 'R' was calculated which indicates the strength of the relationship. R proved to be high 0.742 ($p<0.001$). In comparison to this the English native peer group (left graph) showed an R value of 0.068 ($p=0.710$). It can therefore be concluded that speech rate does have a strong influence on the perception of speech performance quality in French but not in English.

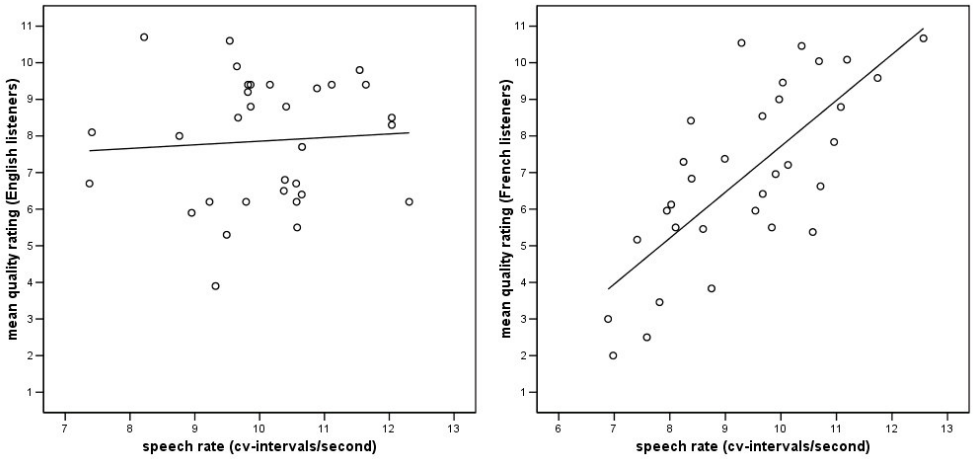

Figure 9-6: Quality rating plotted over speech rate measured in CV-rate (sum of the number of C- and V-intervals per second) for English (left) and French (right) native listeners.

9.3.3. Discussion of the rhythm perception experiment

The results of the perception experiment are clear: When asked to rate the quality of non-native speech on a scale from poor to native-like, native speakers of the languages English and French do not base their judgements on the here discussed acoustic correlates of rhythm. In the case of French though we do find that speech rate rather than rhythm may be an indicator of pronunciation quality. The reasons for why this is only the case in French and not in English may be related to the particular choice of groups. Although the two non native groups, Ge and Gf, proved to be rather comparable in expertise, group Ge did imitate the native English speakers much closer in speech rate than group Gf the native French. If the Germans would

have been equally different in speech rate from the English group than they had been for the French group, then maybe the English native listeners would have reacted to speech rate too. This question cannot be answered on the basis of the current results and will remain subject to future research.

9.3.3.1 Limitations of the experiment (or of rhythmic cues)

Do these results tell us that the acoustic correlates of speech rhythm are not perceptually salient at all? The answer to this question is 'no'. The present methodology is ambiguous in the way that no particularly attention was drawn to rhythmical cues in our stimuli in subjects' instructions. This is one of the compromises that had to be taken in order to use real speech as opposed to delexicalized speech (see chapter 8). These circumstances, however, result in the fact that subjects may have paid attention to cues other than speech rhythm. Such cues might have been segmental qualities of vowels and consonants or other prosodic cues like intonation. However, it may be questioned what value correlates of speech rhythm have in case they do not function as a cue to the perceived overall pronunciation quality of speech? Rhythm is one of the suprasegmental or prosodic cues that do not have a linguistic function (unlike for example intonation), i.e. speakers are not able to distinguish the meaning of two different words or two different phrases on the basis of rhythmical cues alone. But also on a paralinguistic level it is difficult to attribute clear functions to speech rhythm. Speakers can express e.g. certain emotions on the basis of intonation or loudness variations but it seems we are not able to express emotions based on varying rhythm alone. So what do rhythmical cues tell us at all? One of the answers to this question is already given by Lloyd James (1937) and is discussed in detail in chapter 2: rhythm is a requirement for speech intelligibility. The rhythm of a language is of no particular interest when it is performed correctly. Rhythm or the overall timing of speech units is particularly important for a general acoustic well-formedness of speech. Thus it is a cue that is based on 'form' rather than 'function'. A situation where rhythmical features are usually affected is a non-native speech situations in which L2 speakers often perform badly on timing. In this respect the experimental method seems to make a perfectly legitimate statement about the use of rhythmical cues. If measurable deviations in

speech rhythm are not detectable as an indicator for good or poor pronunciation then it may be asked what they could possibly be detected as?

9.3.3.2 The importance of rate differences between L1 and L2

The results of the present experiment suggest that rhythmic correlates monitoring deviations in the vocalic variability of L2 speakers do not play a role in assessing pronunciation quality. However, indications were found that speech rate can contribute to listeners' judgements of pronunciation quality, as in the case of French, where French native listeners rate the performances of Germans speaking French higher when they are faster. The question that arises from this finding is whether speech rate here is a primary perceptual correlate for pronunciation quality or whether it is a secondary effect arising as a result of the improvement of other cues in the signal. It may be that the speakers who spoke French at native or native like rates were simply the better speakers on a number of levels (segmental, intonation, etc.) than the ones who spoke slower. So the fast speakers were probably the ones who were additionally the better speakers in terms of native like vowel qualities, native like intonation, etc. The question whether speech rate alone can be an indicator for pronunciation quality in French remains unanswered from the current data situation and is an interesting topic for follow up experiments. Such an experiment could work with artificially de- and accelerated speech (e.g. with PSOLA). If a simple acceleration of the poorly rated French stimuli would lead to a higher rating regarding their pronunciation quality this may deliver support for speech rate being an independent cue. However, for such a study further methodological issues would need to be taken account for. A simple acceleration of speech stimuli with standard signal processing methods like PSOLA leads a) to a linear time compression of the signal and b) to the introduction of algorithm noise. Although linear compressions of speech signals to higher speeds are more intelligible to listeners than human productions at equivalent speeds they do sound more unnatural to listeners (see Janse, 2003). Further, the introduction of noise by the compression algorithm may further cause artefacts in listeners' pronunciation judgements. A method for such an experiment would therefore need to ensure that non compressed stimuli are comparably degraded in quality. These issues, however,

are issues for a follow up study that are currently not discussed in more detail. If it should turn out, however, that speech rate is a reliable indicator to pronunciation quality such a finding could find direct applications in a teaching environment. Students of languages such as French could probably be trained on native speech rates in order to enhance their pronunciation.

10. General summary and a revised view on rhythm in speech

The present work followed mainly two lines, (a) an acoustic analysis (chapters 4 to 7) of correlates of linguistic rhythm and speech rate and their interaction and (b) a perceptual analysis on the influence of speech rate on rhythmic correlates. In the following the main results from these two lines of research of the current work will be summarized and discussed further.

The methodological basis for the work is laid out by recent findings in the field of speech rhythm studies, showing that rhythm classes (e.g. stress- and syllable-timed languages, detailed definition further below) can be distinguished by the listener based on the durational characteristics of consonantal (C) and vocalic (V) intervals (e.g. Ramus et al., 1999, or Grabe & Low, 2002). The main working hypothesis thus derived from the fact that all suggested correlates of linguistic rhythm are based on measurements of C- and V-interval durations. It was therefore assumed that the durations of C- and V-intervals are highly affected by speech rate (faster rates causes intervals to shrink, slower rates to grow). It was further assumed that, because of variable segment structures (phonotactic structures), languages my show differences according to the way the durational variability of C- and V-intervals is affected by changes in their overall rate. Assumptions about interactions between speech rate and rhythmic correlates have often been addressed as a potential measurement artifact (Ramus et al., 1999, Ramus, 2003, Grabe & Low, 2002, Dellwo & Wagner, 2003, Barry et al., 2003, Dellwo, 2006), however, these influences have only been studied poorly. The present work is probably the first attempt to study speech rate influences on acoustic correlates of speech rhythm in a systematic way. In the following the findings from the acoustic- (10.1) and the perceptual measurements (10.2) will be summarized. In 10.3 a new model of speech rhythm will be suggested and in 10.4 evidence for this model from psychoacoustic research will be discussed. In 10.5 implications of the model on function of speech rhythm will be discussed and final remarks will be drawn.

10.1. Durational variability of C- and V-intervals

As shown in the previous section the data in BonnTempo replicated results from Ramus et al. (1999) and Grabe & Low (2002) which demonstrated that the duration of C- and V-intervals varies significantly between languages that have traditionally been classified as stress-timed (here: English and German) or syllable-timed (French and Italian). In the acoustic domain it was then shown that the consonantal rhythmic correlates ΔC and rPVI are positively correlating with speech rate. Such influences on speech rate could however be normalized for with various procedures (chapter 6). For all normalized and speech rate independent measures it was demonstrated that rPVI and %V are the two measures that distinguish between rhythm classes best at all speech rates. It could be shown that rPVI and %V are significantly higher in syllable timed languages than in stress-timed languages.

In respect to speech rate, however, it was found, that syllable-timed languages typically have higher speech rates than stress-timed languages. This finding was shown to be true even when speakers intend to speak slow and fast. The analysis revealed that also speech rate differed significantly between the stress- and syllable-timed languages under investigation.

Both results (CV-rate and durational C- and V-interval variability) find support in the structural C- and V-interval complexity analysed in chapter 3. For the syllable-timed languages under investigation it was found that the syllable as well as the C-interval complexity is lower than in stress-timed languages (see chapter 3.2.2). The results show that languages with a simpler syllable structure reveal proportionally less C- and V-interval variability and a higher CV-rate than languages allowing more complex syllable structures.

10.2. Perception of durational variability in speech

Ramus et al. (1999) and Nazzi et al. (1998) argued that rhythmical cues such as %V and ΔC may be processed by listeners at infant age already in order to distinguish between languages when in a bilingual environment. Ramus et al. found in a perception experiment using delexicalized speech that listeners were able to distinguish between the languages English and Japanese on the basis of durational

cues only when speech rate between these languages does not vary. The results from the present study, however, showed that speech rate can be an indicator for rhythmic class as well.

The question arising from this is how listeners perceive durational variability when speech rate varies naturally between languages. This was addressed in a perception experiment using delexicalized speech that preserved only durational cues with the languages German (maximum low CV-rate and highest proportional durational variability of C- and V-intervals) and French (maximum high CV-rate and lowest proportional C- and V-interval variability). Listeners' regularity rating of the delexicalized stimuli revealed that their choices were not made according to the variability of C- and V-interval durations but they were strongly influenced by CV-rate. It was concluded that C- and V-variability cues are not salient when rates vary. Since rates typically vary between stress- and syllable timed languages it can be assumed that rate is an important cue for listeners to distinguish between rhythm classes.

A second perception experiment in the L2 domain supports these results (chapter 9). It was argued that an observable deviation of %V and nPVI between native and non-native language productions should be salient in terms of rhythmic variability between these two productions. It was expected that listeners would rate pronunciations of subjects as better when nPVI and/or %V come closer to native like variability. However, no evidence could be found that listeners used nPVI or %V as cues to pronunciation quality. In the case of French, where speech rate between the native and non-native productions varied widely, this cue correlated with listeners' responses. Results suggest that speech rate is an important cue for pronunciation quality. However, further experiments are necessary to test this hypothesis since rate in the experiment presented here may simply be a secondary factor that in return may correlate with other (e.g. segmental) cues for pronunciation quality.

10.3. Hypothesizing about a new model of rhythm class

Two findings from the acoustic and perceptual study will be compared in more detail in the following. It is the finding about the speech rate differences between the

languages under investigation from chapter 5 and the finding of the perception of CV-rate from chapter 8. Both figures are reprinted in Figure 10-1.

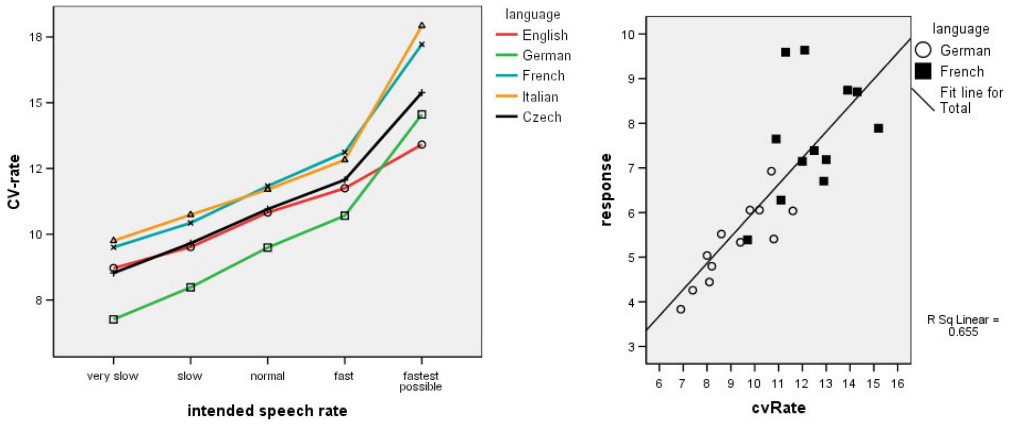

Figure 10-1: Findings about acoustic CV-rate differences between languages from chapter 5, Figure 5-4, (left) and the influence of CV-rate on listeners' perception of regularity of delexicalized speech stimuli from chapter 8, Figure 8-6, (right).

The findings from the acoustic rate analysis of English, French, German, Italian and Czech show that at a normal rate the distinction between the languages that have traditionally been classified as stress- and syllable timed is at about 11 C- or V-intervals per second. For the current interpretation it will be assumed that languages faster than 11 C- or V-intervals per second are syllable timed, lower than this threshold stress-timed.

The results from the perceptual study showed that there is a linear relationship between CV-rate and listeners' response. However, the graph suggests that the linearity is breaking down at about 11 C- or V-intervals per second. This was also supported in chapter 8 where the rating for the French stimuli showed a drastically lower correlation with CV- rate than for the German stimuli. The linearity is maintained up to about 11 C- or V-intervals per second but at rates higher than this the linearity begins to break down.

It is assumed that this rate threshold of 11 C- or V-intervals per second may be a crucial threshold for the auditory distinction between the two categories. Languages below 11 C- or V-intervals per second may sound 'irregular' in terms of interval

durations because listeners are able to distinguish between variable rates within these languages (here: German). In languages that are typically produced at rates higher than 11 C- or V-intervals per second the percept of 'irregularity' may not be perceived because the relatively high rates of C- and V-intervals evoke the perception of 'regularity'.

According to this model rhythmic classes would simply be a result of rate differences between languages. In its purest form this hypothesis does probably not find a lot of support since there are numerous other factors apart from absolute duration which contribute to the perception of durational regularity or irregularity. Such factors are for example intonation patterns, stress variability, segment qualities, etc. These influential factors have not been considered in the present work. In experimental set-ups which looked at the perception of durational variability only (Ramus & Mehler, 1999, Nazzi et al., 1998, Ramus, 2002, Ramus et al., 2000, Rincoff et al., 2005, Toro et al., 2003) rate influences may have been a potential artifact.

O'Connor (1965) for example shows that the sequence of intervals plays an important role in the perception of regularity. He demonstrated that in order for two intervals to be perceived as isochronous in duration the second interval has to be considerably longer (typically around 25%) than the first interval in the sequence. This type of perceptual warping as a result of sequencing is not accounted for in the current study. It is a type of variability that would be difficult to account for in durational variability typically occurring in the intervals discussed in the present work since they vary in multiple dimensions. However, a line of research on auditory perception of general rhythmic cues will be presented in the following which does supply some relevant findings for the interpretation of the present data.

10.4. The perception of regularity

The perception of durational variability has been argued to underlie Weber's law (see Abel, 1972, and a review in Friberg & Sundberg, 1995). Weber's law states that the size of the just noticeable difference (jnd) between two stimuli and the magnitude of the initial stimulus are lawfully related. Weber's law is expressed in the following formula:

$\frac{\Delta I}{I} = k$; where ΔI is the jnd threshold, I the initial stimulus intensity and k is a constant value arising from this ratio. The fraction is also generally referred to as the Weber fraction. Once k has been identified for one particular relationship between ΔI and I, ΔI can be predicted for each I according to:

$$\Delta I = k \cdot I$$

A typical example for Weber's law is the case of weight, where the jnd between two weights can be identified for example by increasing a 5 kg weight in increments of 0.1 kg. Empirical studies have shown that an increase of 0.5 kg is necessary to trigger the sensation of a just noticeable difference in humans. This increment is then ΔI and using the first formula above, k would be equal to 0.1 (0.5/5). Knowing k we can now predict the jnd (ΔI) for example for a 2 kg weight which would be 0.2 (0.1*2).

Weber's law has proved to hold in numerous dimensions of human sensation, as for example visual impressions, temperature sensation, and, most important for the current context, the sensation of time. However, for any dimension of human sensation, Weber's law is only linear within a certain range of values but it breaks down outside these ranges. In the case of the weight example it occurs that below a certain value of I, ΔI increases to values higher than I since the difference between any value smaller than I cannot be detected any more by humans (e.g. weight differences below a gram lie well outside human perception).

For the perception of variability between two auditory intervals, Abel (1972) finds that when interval durations are smaller than about 3 ms, the linearity predicted by Weber's law breaks down. This means that below 3 ms a proportionally higher amount of variability is necessary in order to trigger the perceptual sensation of a durational interval difference. Figure 10-2 shows the findings for empty intervals (distinguished by click sounds). The graph plots ΔT (the just noticeable durational difference between two empty intervals) as a function of T (the duration of the first interval in the sequence). Between 3 and 300 ms the relationship is linear where ΔT is about 30% of T, below 3 ms ΔT starts increasing proportionally. At about 1 ms the just noticeable difference ΔT is about 300% of T. The graph also shows the effect of

marker condition (e.g. long and short clicks) on the perception of durational variability which will not be discussed in the present context.

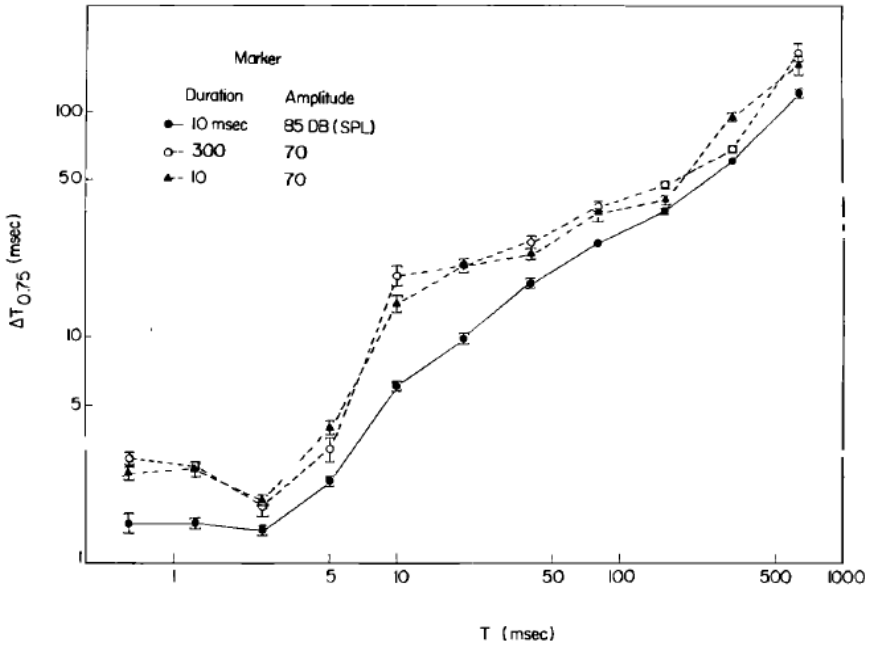

Figure 10-2: Graph taken from Abel (1972) showing ΔT (difference between two silent intervals) as a function of T (initial silence interval). The graph shows that Weber's law breaks down at about 3 ms for all conditions under investigation.

What Abel's findings show is that below a certain rate the proportional variability between stimuli needs to be increased a lot for it to be perceivable. With increasing rate an increase in variability is necessary to be perceived by the listener. On the other hand this means that with increasing rate an equal amount of variability is perceived as smaller. In respect to C- and V-interval variability in speech it has been repeatedly shown in previous studies (see discussions above) as well as in the present study that syllable-timed languages possess less variability of these segments than stress-timed languages. In addition the present study showed that syllable-timed languages are produced faster compared to their stress-timed counterparts. In the light of the psychoacoustic studies quoted above it therefore seems highly likely that the higher CV-rates in syllable-timed languages contribute to the percept of a

durational regularity of the underlying units.

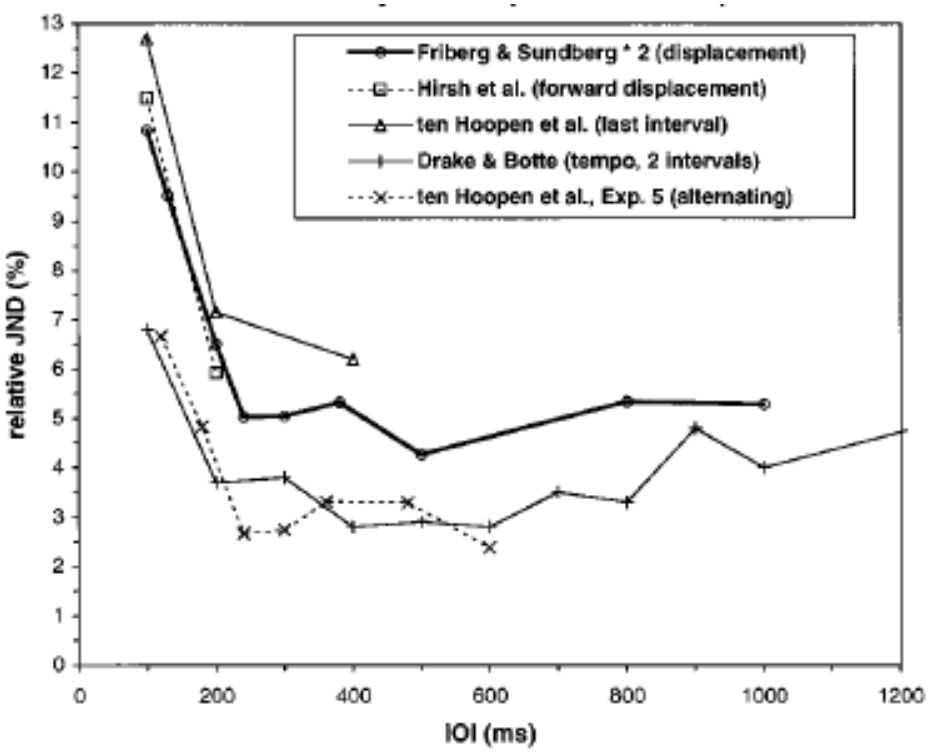

Figure 10-3: Graph reprinted from Friberg & Sundberg (1995) showing the results of the just noticeable difference in % between varying inter-onset-intervals (IOIs). Lines represent results from 5 different studies of various authorship (see Friberg & Sundberg for details).

In absolute terms it is difficult to draw parallels between Abel's data and the data of the present work. None of the intervals in the present work is as small as 3 ms. In terms of C- and V-intervals per second it was demonstrated in chapter 5 that rates between 8 and 12 C- or V-intervals per second are typical for all languages under investigation at normal reading speed. This is equivalent to interval durations between about 125 and 85 ms. However, work on the perception of isochronously rhythmic sequences reveals that in longer sequences of rhythmic chains, Weber's law breaks down at far higher intervals durations. Friberg & Sundberg (1995) summarize a number of studies (see Figure 10-3) that have worked with distorted isochronous rhythmical sequences. These sequences typically consist of a series of tones or empty gaps while one of the intervals is distorted in a non-isochronous way. Such a

distortion can for example be a durational increase of one interval of the chain, a delay in the onset or offset of one interval, etc. All studies summarized in Friberg & Sundberg show that the just noticeable difference for the distortion increment increases drastically in respect to the average interval duration of the rhythmic sequence when this average interval is below 200 ms.

10.5. Implications and final remarks

The main results of the perceptual studies are clear: when speech rate is not normalized for, it is the preferred salient cue over interval variability. Results from studies on the perception of duration discussed in the previous section support this view.

In real language situations speech rate is never normalized for. As discussed briefly in chapter 1, Ramus et al. (1999) claim that infants growing up in a bilingual environment may make use of durational interval variability cues in order to separate between languages (provided they belong to different rhythmic classes) before they have acquainted any linguistic knowledge of that language. With respect to the findings from chapter 5 it can be assumed that CV-rates vary if the languages to be distinguished belong to different rhythmic classes. By reducing speech stimuli from French and German to durational information of C- and V-intervals only (delexicalization procedure), grown-ups are put into a similar situation like infants in which they have to rely on durational cues only. It was shown that grown-ups prefer speech rate to interval variability cues. For this reason it may be assumed that speech rate is an equally important cue for infants to distinguish between languages of different rhythmic classes. In order to test such a hypothesis, infant experiments will be necessary in the future.

The rhythm class hypothesis, however, was not formulated by infants but by phonetic experts (see discussions in chapter 1 and 2) and phonetic experts claimed that they are able to distinguish on an auditory basis between stress- and syllable-timed languages. Miller (1984) showed that experts are able to distinguish between languages from different rhythmic classes reliably and that such knowledge can be acquired by phonetically naive subjects after some training. The question remains

whether expert listeners are able to distinguish between rate and variability cues in the experiment in chapter 8. With respect to the findings discussed in the previous section, doubt should be put on that.

Summing up, the present study presented evidence suggesting that speech rate may be an important driving factor that may contribute significantly to the perception of what phoneticians refer to as speech rhythm. In the past the value of speech rate has probably been underestimated throughout in the field. Future experiments will show whether these findings will be replicated by further studies on rate influences on the perception of rhythm.

APPENDIX I: List of Abbreviations

The following is an alphabetical list of abbreviations that are frequently used in the text. Each abbreviation is followed by the unabbreviated use of the term and a short definition. In the unabbreviated term the letters in bold are the letters that have been used to create the abbreviation.
Most abbreviations used in the formulas in chapter 3.3.2 have only been used in this particular circumstance. Such abbreviations are not listed again in the following list.

ANOVA **An**alysis **o**f **va**riance
A procedure to test whether the mean values of the distributions of three or more related groups are the same.

BTC **B**onn**T**empo **C**orpus
A database for the analysis of speech rhythm and rate (see section 3.1 and Dellwo et al., 2004).

BTT **B**onn**T**empo **T**ools
Software tools for the analysis of the BonnTempo Corpus (BTC, see section 3.1). The BTT collection is written in the Praat scripting language (www.praat.org).

c-segment **c**onsonantal **segment**
A single consonant.

C-interval **c**onsonantal **interval**
In an utterance, a series of one or more consonants preceded and followed by a vowel (or a pause).

CV-rate **C**- or **V**-interval **rate**
A speech rate measure calculated by the sum of the number of C- and V-intervals per second in speech (see section 3.3.2 for a formula).

ipi **i**nter-**p**ause-**i**nterval
A pause-free interval of speech preceded and followed by a pause.

ln-transfom natural logarithmic transformation
A method by which numeric data points are expressed as logarithms to the base e.

lsr **l**aboratory measurable **s**peech **r**ate
The rate of speech as obtainable from the acoustic signal by measuring the number of certain units per second. The units used in the present study are typically C- or V-intervals. In some rare cases also syllables have been applied.

nPVI **n**ormalized **P**airwise **V**ariability **I**ndex
Average durational difference between consecutive V-interval durations in an utterance. This measure is normalized for speech rate influences on the average differences (see 3.3.2. for a formula).

nPVI-C **n**ormalized **P**airwise **V**ariability **I**ndex for **C**-Interval durations
Equal measurement procedure like the nPVI but applied to C-interval durations (see 3.3.2 for a formula).

rPVI **r**aw **P**airwise **V**ariability **I**ndex
Average durational difference between consecutive C-intervals in an utterance (see 3.3.2 for a formula).

varcoC **co**efficient of **var**iation of **C**-interval durations
Standard deviation of C-interval durations expressed as a percentage of the mean C-interval duration.

v-segment **v**ocalic **segment**
A single vowel.

V-interval **v**ocalic **interval**
In an utterance, a series of one or more vowels preceded and followed by a consonant (or a pause).

ΔC delta C-Interval
The standard deviation of C-Interval durations in an utterance (see 3.3.2 for a formula).

%V Percent Vocalic
The percentage over which speech is vocalic (see 3.3.2 for a formula).

APPENDIX II:

Recording texts for BonnTempo

In the following the recording texts for the BonnTempo database are listed for the languages under investigation (English and German as stress-timed examples, French and Italian as syllable-timed examples and Czech as a not easily classifiable language). The vertical lines in the text identify the stretches for which correlates of speech rate (CV-rate) and rhythm (%V, ΔC, nPVI, rPVI, etc.) have been calculated in the speech signal of each speaker and intended tempo version.

German (original Version):

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

Bernhard Schlink (1994) *Selbs Betrug: Diogenes* (p. 242)

English:

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

(Translation: Stacy Dellwo)

APPENDIX

French:

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

(Translation: Judith Adrien)

Italian:

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

(Translation: Franco Ruina)

Czech:

Následující den jsem jel do Zlína. | Je to cesta na konec světa; | za Kladnem začnou být kopce a lesy jednotvárné, | za Blanskem začnou být města nudná | a u Vyškova začne být krajina rovinatá a neutěšená. | Kdyby byli u nás disidenti vyhošťováni, | byli by vyhoštěni do Boršovské Vsi.

(Translation: Jana Dankovičová)

APPENDIX III: BonnTempo-Tools

The BonnTempo-Tools (BTT) is a collection of Praat based software (Praat scripts) to facilitate access and analysis of the BTC. The BTT can be obtained from the author by contacting him via his webpage (www.phonetiklabor.de). An installation of the Praat speech analysis software (downloadable under a GUI license from www.praat.org) is necessary to use BTT. The tools are independent of the corpus and are not necessarily needed to perform analysis. The following pages give an instruction about the installation of the tools and their basic functions.

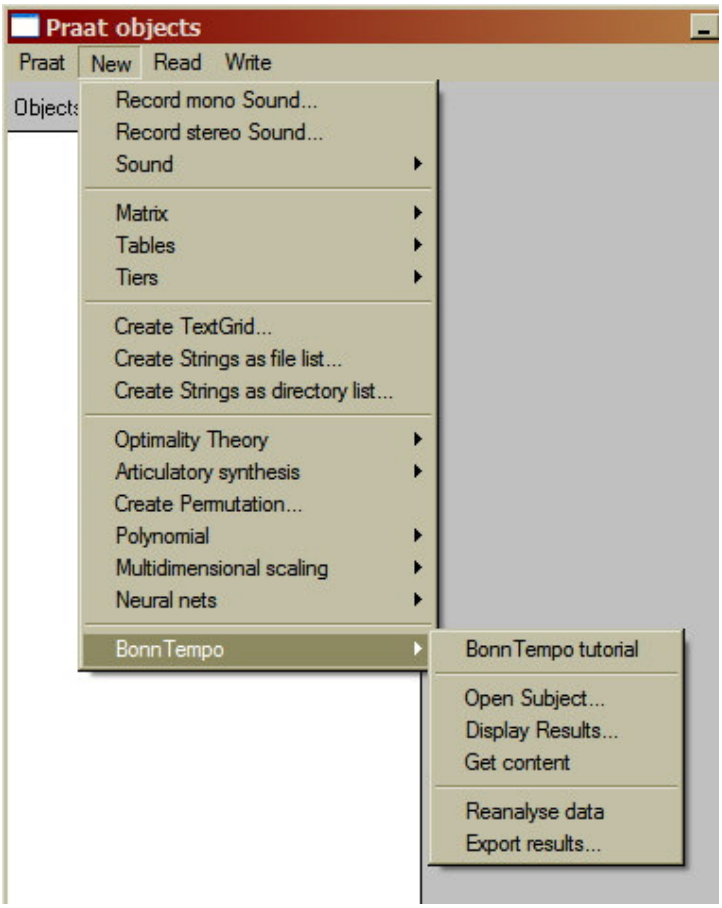

Figure APPENDIX III-1: Menu of BonnTempo after installation in Praat.

APPENDIX

Installation

The BTT can be installed via the file ‘installation.praat’ in the directory BonnTempo/tools (follow the installation instructions in the file readMe.txt in the same directory.) After the installation an additional menu point ‘BonnTempo’ can be found under the new menu in Praat (see Figure APPENDIX III-1).
A detailed interactive tutorial about the BTC and the BTT can be obtained in this menu. The remaining points, ‘Open Subject…’, ‘Display Results…’, ‘Get content’, ‘Re-analyze data’, and ‘Export results’ is described in the following.

File format of BonnTempo-Corpus

All five intended speech tempo versions (very slow, slow, normal, fast, and fastest possible) of each speaker have been saved in wav format in separate files. The file names contain information about the native language of the speaker in capital letter (e.g. ‘E’ for English), followed by the language the speaker used for reading the text in small letters (e.g. ‘f’ for French), followed by a speaker number (e.g. ‘08’), followed by the intended speech tempo version as numbers (very slow = 1, slow = 2, normal = 3, fast = 4, fastest possible = 5). Language, speaker’s name, and intended tempo information are separated by an underscore (e.g.: Ef_08_03.wav = English native speaker [E] number 8 [08] reading the French text version [f] in normal intended tempo [03]). The labelling work for each speaker was saved in Praat label files of the type ‘TextGrid’.

Tool: Get Content

This tool displays the actual content of the BTC. Since the corpus is not static but continuously growing as a result of numerous research projects (see 3.1.2) the tool informs the user about the actual content of the database (e.g. total number of

languages, the number of speakers for the whole database as well as for each language, number of syllables, number of C- and V-intervals, etc). Get Content prints all values of the current state of the database in the Praat info window from where it can for example be saved as a text file.

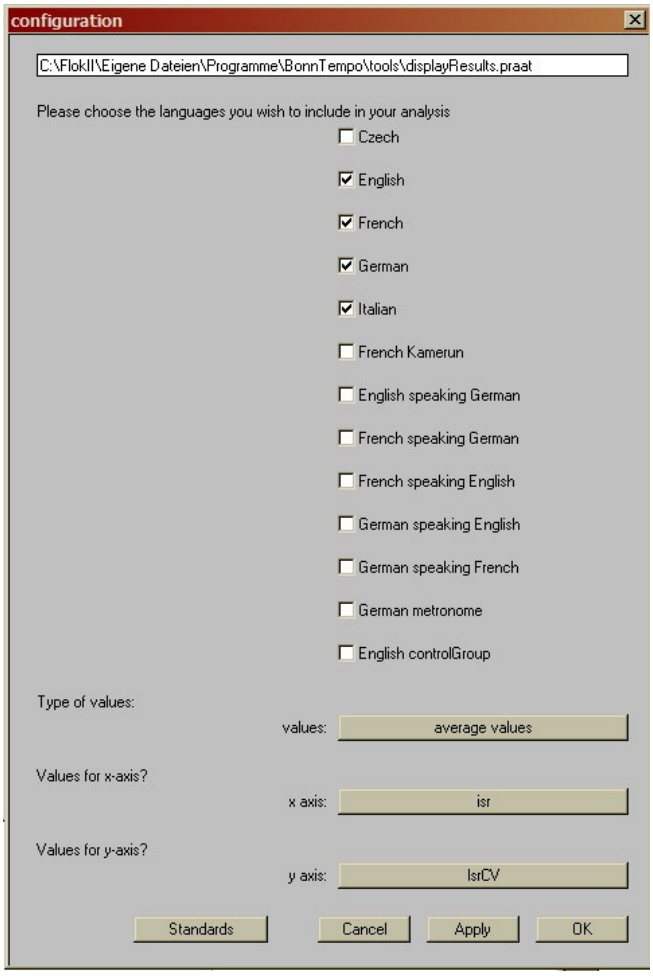

Figure APPENDIX III-2: Interface of the tool ‚display results'. The user can choose languages of the BTC individually and cross plot each combination of parameters along the x and y axis.

Tool: Re-analyze data

To speed up the analysis of the data in BonnTempo the raw label files are not directly processed each time an analysis is performed. Instead a general analysis of all values is performed in an initial stage and saved as a text file of the Praat format 'TableOfReal'. One file is produced for each analysis parameter for each language separate. Each file contains the subject of this language in each line and the five intended speech tempi in columns. Any further analysis is based on this file thus any time data is added to the corpus the file has to be renewed by executing the 'Re-analyse data' command. In the following a description of all analysis parameters is given.

The analysis parameters consist of all mean values (mean), standard deviation of mean values (stdev) and the variation coefficient of the stdevs (varco) of respective segment durations (syllables, C- and V-intervals, and also pauses). In addition to theses values the percentage of V-intervals in a speech signal and the laboratory measurable speech rate (syllables/second without pauses) as well as the nPVI and rPVI are available. Each value was calculated individually for each of the five intended tempo versions for each speaker. Calculations for individual values are described in more detail in chapter 3.

Tool: Display results

With this tool analysis of the results can be presented in graphical form. It allows the user to cross plot any pair of speech rate and/or rhythm parameters of the type described in chapter 3. It is further possible to plot any the analysis parameters along an axis containing the five intended tempo versions (very slow, slow, normal, fast, and fastest possible). Figure APPENDIX III-2 shows the interface. In the figure it is also visible that each language (or L2 variety) of the BTC can be chosen individually (by ticking the box in front of the language in the upper part of the dialogue window). The user also has the choice to plot either a) values averaged across all

speakers in each language, b) the same average values with standard deviations plotted as arrows, and c) individual values for each speaker individually.

Figure APPENDIX III-3 presents two examples of a typical cross-plot output of the 'display analysis' tool. Here the measure ΔC was plotted across %V. The data was averaged across the speakers for each of the four languages presented (English = green, French = blue, German = pink, Italian = purple). Mean values for each intended tempo version (very slow, slow, normal, fast, very fast) in each language are consecutively connected from the slowest version (very slow) to the fastest version (very fast), while the slowest version is accompanied by the language label (here: Ee, Ff, Dd, and Ii; see 3.2.3 for an explanation of the labels). The left graph of the figure presents mean values only while the right graph includes x and y standard deviations of the respective mean values displayed as arrows surrounding the means. Numeric results for all x and y values are additionally printed into the Praat info window.

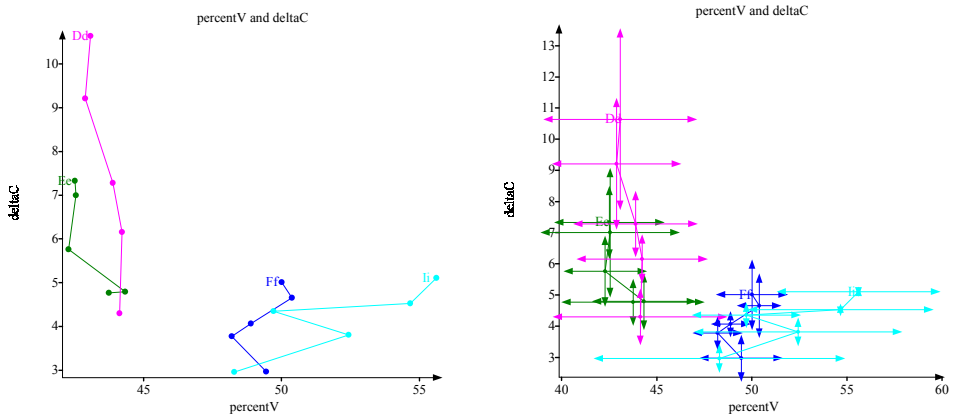

Figure APPENDIX III-3: Example of a typical box-plot data created by the tool 'display results'. The graphs display the five intended tempo versions (very slow, slow, normal, fast, very fast) of four languages (Ee, Ff, Dd, Ii; see 3.2.3) with (left) and without (right) standard deviations.

APPENDIX

Tool: Export

To allow more complex statistical analysis using standard data processing tools like SPSS or R, the BTT have a command to export the data either in raw format or semi-processed in tab separated text files which can be read in by almost all software products. Raw format exporting provides three tab-separated text files which contain a) durational data for each individual syllable in the database, b) durational data for each individual C-interval, and c) durational data for each individual V-interval. The semi processed export tool produced a tab separated text file containing all (and others) in chapter 3 described parameters for each of the five intended speech tempo versions of each speaker.

Tool: Open Subject

This is a tool with which the labelled files (TextGrid) and sound files (wav) of a particular speaker in the BTC can be added as an object in Praat list of objects. It facilitates a quick and easy access to individual files or groups of files of each speaker in the BTC. An interface lets the user choose the speaker, the type of file (all, wav, TextGrid) and the language (all, particular L1s, particular L2s) that is to be opened.

Bibliography

Abel, S. M. (1972) "Discrimination of temporal gaps" *Journal of the Acoustical Society of America* (52,2), 519-524.

Abercrombie, D. (1967) *Elements of General Phonetics*. Edinburgh: University Press.

Adams, C. (1979) "English Speech Rhythm and the Foreign Learner". The Hague: Mouton.

Allen, G. D. (1972a) "The location of rhythmic stress beats in English: an experimental study I" *Language and Speech* (15), 72-100.

Allen, G. D. (1972b) "The location of rhythmic stress beats in English: an experimental study II" *Language and Speech* (15), 179-195.

Allen, G. D. (1975) "Speech rhythm: its relation to performance universals and articulatory timing" *Journal of Phonetics* (3), 75-86.

Allen, G. D. (1978) "Vowel duration measurement: A reliability study" *Journal of the Acoustical Society of America* (63, 4), 1176-1185.

Allen, G. D. and Ladefoged, P. (1971) "Syllable Structure and Sentence Rhythm - A Cross Language Study" *Journal of the Acoustical Society of America* (50), 116.

Antipova, A. (1987) "Speech rhythm (main approaches and definitions)" *Proceedings of the 11th ICPhS*, 443-446.

Antoniadis, Z. and Strube, H. W. (1984) "Untersuchung zur spezifischen Dauer deutscher Vokale" *Phonetica* (41), 72-87.

Arvaniti, A. (1991) "Rhythmic categories: a critical evaluation on the basis of Greek." *Proceedings of the 12th ICPhS*, 298-301.

Arvaniti, A. (1994) "Acoustic features of Greek rhythmic structure" *Journal of Phonetic* (22), 239-268.

Barbosa, P. and Bailly, G. (1993) "Generation and evaluation of rhythmic patterns for text-to-speech synthesis" *Proceedings of ESCA Workshop on Prosody; Lund University Working Papers in Phonetics*, 66-69.

Barbosa, P. and Bailly, G. (1994) "Characterisation of rhythmic patterns for text-to-speech synthesis" *Speech Communication* (15), 127-137.

BIBLIOGRAPHY

Barik, H. C. (1977) "Cross-linguistic study of temporal characteristics of different types of speech materials" *Language and Speech* (20), 116-126.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., and Kostadinova, T. (2003) "Do rhythm measures tell us anything about language type?" D. Recasens, M. J. Solé and J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain: 2693–2696.

Bartkova, K. (1991) "Speaking rate modelization in French: application to speech synthesis" *Proceedings of the 12th ICPhS* (3), 482-485.

Beckman, M. E. (1992) "Evidence for Speech Rhythms across Languages" Y. Thokura, Vatikiotis-Bateson and Y. Sagisaka (eds.) *Speech Perception, Production and Linguistic Structure*. Amsterdam: IOS, 457-563.

Bell-Berti, F. (1991) "Comments on 'some observations on the organisation and rhythm of speech" *Proceedings of the 12th ICPhS* (2), 238-242.

Berinstein, A. (1978) "A cross linguistic study on the perception of stress" *Journal of the Acoustical Society of America* (63), 55-56.

Boersma, Paul (2001) "Praat, a system for doing phonetics by computer" *Glot International* (5, 9/10), 341-345.

Bolinger, D. L. (1965) "Pitch accent and sentence rhythm" I. Abe and J. Kanekio (eds.) *Forms of English: Accent, Morpheme, Order*. Cambridge Mass.: HUP, 139-180.

Bolinger, D. L. (1975) *Aspects of Language*. London: Harcourt Brace Jovanovich.

Bolinger, D. L. (1981) *Two kinds of vowels, two kinds of rhythm*. Bloomington, Indiana: Indiana University Linguistics Club.

Bond, Z. S. and Fokes, J. (1985) "Non-native patterns of English syllable timing" *Journal of Phonetics* (13), 407-420.

Borzone de Manrique, A. M. and Signorini, A. (1983) "Segmental duration and rhythm in Spanish" *Journal of Phonetics* (11), 117-128.

Brown, A. (1988) "The staccato effect in the pronunciation of English in Malaysia and Singapore" J. Foley (ed.) *New Englishes: The Case of Singapore*. Singapore: Singapore University Press, 115-147.

Brown, W. (1911) "Temporal and accentual rhythm" *Psychological Review* (13), 336-346.

BIBLIOGRAPHY

Bröggelwirth, J. (2003) "Zum aktuellen Stand der Sprechrhythmusforschung" IKP Arbeitsberichte, NF 05 (http://www.ikp.uni-bonn.de/ikpab/)

Brubaker, R. S. (1972) "Rate and pause characteristics of oral reading" *Journal of Psycholinguistic Research* (1), 141-147.

Butcher (1981) "Phonetic correlates of perceived tempo in reading and spontaneous speech" *Phon. Lab. Univ. Reading Work in Progress* (3), 105-117.

Buxton, H. (1983a) "Temporal predictability in the perception of English speech" A. Cutler and D. R. Ladd (eds.) *Prosody: Models and Measurements*. Berlin: Springer, 111-121.

Buxton, H. (1983b) *Rhythm and stress in speech*. Cambridge: CUP.

Byrd, D. and Tan, C. C. (1996) "Saying consonant cluster quickly" *Journal of Phonetics* (24), 263-282.

Carlson, R. (1991) "Duration models in use" *Proceedings of the 12th ICPhS*, 243-245.

Carlson, R. and Granstrom, B. (1986) "A search for durational rules in a real-speech data base" *Phonetica* (43), 140-154.

Caspers, J., und V. J. van Heuven (1995) "Effects of time pressure on the choice of accent-lending and boundary-marking pitch configuration in Dutch" *Proceedings of Eurospeech* (2), 1001-1004.

Cauldwell, R. (1996) "Stress-timing: observations, beliefs and evidence" *Eger Journal of English Studies* (1), 33-48.

Chatman, S. (1965) *A Theory of Meter*. Berlin: Mouton.

Chela Flores, B. (1970) "On the acquistion of English rhythm: theoretical and practical issues" *International Review of Applied Linguistics* (32, 3), 232-242.

Chen, M. (1970) "Vowel length variation as a function of the voicing of the consonant environment" *Phonetica* (22), 129-159.

Classe, A. (1939) *The rhythm of English prose*. Oxford: Blackwell.

Couper-Kuhlen, E. (1993) *English Speech Rhythm*. London: Benjamins.

Cowley, S. (1994) "Conversational functions of rhythmical patterning" *Language and Communication* (14, 4), 353-376.

BIBLIOGRAPHY

Crompton, A. (1980) "Timing patterns in French" *Phonetica* (37), 205-234.

Crystal, D. (1969) Prosodic systems and intonation in English. Cambridge: CUP.

Crystal, D. (1995) "Documenting rhythmical change" J. W. Lewis (ed.) *Studies in General and English Phonetics*. London: Routledge, 174-179.

Crystal, T. H. and House, A. S. (1982) "Segmental durations in connected speech signals: preliminary results" *Journal of the Acoustical Society of America* (72), 705-716.

Crystal, T. H. and House, A. S. (1988) "Segmental durations in connected speech signals: current results" *Journal of the Acoustical Society of America* (83, 4), 1553-1573.

Crystal, T. H. and House, A. S. (1990) "Articulation rate and the duration of syllables and stress groups in connected speech" *Journal of the Acoustical Society of America* (88, 1), 101-112.

Cummins, F. (1999) "Some lengthening factors in English combine additively at most rates" *Journal of the Acoustical Society of America* (105, 1), 476-480.

Cummins, F. and Port, R. (1998) "Rhythmic constraints on stress timing in English" *Journal of Phonetics* (26), 145-171.

Cutler, A. (1991) "Linguistic rhythm and speech segmentation" J. Sundeberg, L. Nord and R. Carlson (eds.) *Music, Languages, Speech and Brain (Wenner-Gren Symposium Series 59)*. London: Macmillan, 157-166.

Cutler, A. (1994) "Segmentation problems, rhythmic solutions" *Lingua* (92), 81-104.

Dankovičová, J. (2001) The linguistic basis of articulation rate variation in Czech. Frankfurt: Hector.

Dankovičová, J and Dellwo, V. (2007) "Czech speech rhythm and the rhythm class hypothesis" *Electronic Proceedings of the 16th ICPhS.*

Darwin, C. J. and Donovan, A. (1980) "Perceptual studies of speech rhythm: isochrony and intonation" J. C. Simon (ed.) *Spoken Language Generation and Understanding*. Dordrecht: Reidel, 77-85.

Dasher, R. and Bolinger, D. L. (1982) "On pre-accentual lengthening" *Journal of the International Phonetic Association* (12), 58-69.

BIBLIOGRAPHY

Dauer, R. (1980) *Stress and rhythm in modern Greek*. PhD Dissertation - University of Edinburgh.

Dauer, R. M. (1983) "Stress-timing and syllable-timing reanalysed". *Journal of Phonetics* (11), 51-69.

Dauer, R. M. (1987) "Phonetic and phonological components of language rhythm" *Proceedings of the 11th ICPhS*, 447-450.

Dawson, L. O. (1929) "A study of the development of the rate of articulation" *Elementary School Journal* (29), 610-615.

Dellwo, Volker (2006) "Rhythm and Speech Rate: A Variation Coefficient for ΔC" Pawel Karnowski & Imre Szigeti (eds.) *Language and Language-processing*. *Procedings of the 38th Linguistics Colloquium*, Piliscsaba 2003. Frankfurt am Main: Peter Lang: 231-241.

Dellwo, V., Ferragne, E. and Pellegrino, F. (2006) "The perception of intended speech rate in English, French, and German by French listeners" *Electronic Proceedings of Speech Prosody*, Dresden, Germany.

Dellwo, V., Fourcin, A. and Abberton, E. (2007) "The rhythmicity of voice" *Electronic Proceedings of the 16th ICPhS*.

Dellwo, V., Steiner, I., Bianca Aschenberner, Jana Dankovičová, Petra Wagner (2004) "The BonnTempo-Corpus and BonnTempo-Tools. A database for the combined study of speech rhythm and rate" *Electronic Proceedings of the 8th ICSLP*.

Dellwo, V. and Wagner, P. (2003) "Relations between Language Rhythm and Speech Rate" *Proceedings of the 15th ICPhS*, 471–474.

Denes, P. (1955) "Effects of duration on the perception of voicing" *Journal of the Acoustical Society of America* (27), 761-764.

Deterding, D. (2001) "The measurement of rhythm: a comparison of Singapore and British English" *Journal of Phonetics* (29), 217–230.

Dimitrova, S. (1997) "Bulgarian Speech Rhythm: Stress-Timed or Syllable-Timed?" *Journal of the International Phonetic Association* (27, 1), 27-33.

Donovan, A and Darwin, C. J. (1979) "The perceived rhythm of speech" *Proceedings of the 9th ICPhS* (2), Copenhagen, 268-274.

Donovan, J. (1889) Music and Action; or, the elective affinity between rhythm and pitch. A psychological essay, etc. London: Kegan Paul.

Duarte, D., Galves, A, Lopes, N. and Maronna, R. (2001) "The statistical analysis of acoustic correlates of speech rhythm" *Rhythmic Patterns, Parameter Setting and Language Change, ZiF, Universtiy of Bielefeld*.

Edmonds, E. M. (1883) *Hesperas: Rhythm and Rhyme*. London: Kegan Paul.

Eefting, W. (1988) "Temporal variation in natural speech: Some explorations" *Proceedings Speech'88: 7th FASE Symposium*, Edinburgh, 503-507.

Eefting, W. and Rietveld, A. C. M. (1988) "Just noticeable differences of articulation rate at sentence level" *Proceedings* (12) Department of Language and Speech, Phonetics Section, University of Nijmegen.

Eriksson, A. (1991) "Aspects of Swedish Speech Rhythm" *Gothenburg Monographs in Linguistics* (9) Gothenburg University: Department of Linguistics.

Faber, D. (1986) "Teaching the rhythms of English: a new theoretical base" A. Brown (ed.) *Teaching English Pronunciation. A book of readings*. London: Routledge, 245-258.

Farinas, J. and Pellegrino, F. (2001) "Automatic rhythm modeling for language identification" *Proceedings of Eurospeech* (1), 2539-2542.

Farnetani, E. and Kori, S. (1990) "Rhythmic structure in Italian noun phrases: A study on vowel durations" *Phonetica* (47), 50-65.

Faure, G. Hirst, D. J. and Chafcouloff (1980) "Rhythm in English: Isochronism, Pitch, and Perceived Stress" L. R. Waugh and C. H. v. Schooneveld (eds.) *The Meldoy of Language*. Baltimore: University Park, 71-79.

Flecher, J. (1987) "Some micro and macro effects of tempo change on timing in French" *Linguistics* (25), 951-967.

Fletcher, J. (1991) "Rhythm and final lengthening in French" *Journal of Phonetics* (19), 193-212.

Fougeron, C. and Jun, S. A. (1998) "Rate effects on French intonation: prosodic organisation and phonetic realization" *Journal of Phonetics* (26), 45-69.

Fourakis, M. (1991) "Tempo, stress, and vowel reduction in American English" *Journal of the Acoustical Society of America* (90, 4), 1816-1827.

Fowler, C. A. (1983) "Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet" *Journal of Experimental Psychology* (112, 3), 386-412.

Frackowak-Richter, L. (1987) "Modeling the rhythmic structure of utterances in Polish" *Studia Phonetica Posnaniensia* (1), 91-125.

Fraisse, P. (1963) *The Psychology of Time*. New York: Harper and Row.

Fraisse, P. (1978) "Time and rhythm in perception". E. C. Carterette and M. P. Friedman (eds.) *Handbook of Perception* (8). New York: Academic Press.

Friberg, and Sundberg (1995) "Time discrimination in a monotonic, isochronous sequence" *Journal of the Acoustical Society of America* (98, 5), 2524-2531.

Frota, S and Vigàrio, M. (2001) "On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case" *Probus* (13), 247-275.

Garcia, J. and Galves, A. (2001) "Automatic identification of vocalic intervals in the speech signal." Rhythmic Patterns, Parameter Setting and Language Change, ZiF, Universtiy of Bielefeld.

Gay, T. (1978) "Effects of speaking rate on vowel formant movements" *Journal of the Acoustical Society of America* (63), 223-230.

Gay, T. (1981) "Mechanisms in the control of speech rate" *Phonetica* (38), 148-158.

Gay, T., Ushijima, T. Hirose, H. and Cooper, F. S. (1974) "Effects of speaking rate on labial consonant-vowel articulation" *Journal of Phonetics* (2), 47-63.

Gilbert, J. H. and Burk, K. W. (1969) "Rate alterations in oral reading" *Language and Speech* (12), 192-201.

Giles, H. (1992) "Speech tempo" W. Bright (ed.) *The Oxford International Encyclopedia of Linguistics*. Oxford: OUP.

Goldman-Eisler, F. (1961) "The significance of changes in the rate of articulation" *Language and Speech* (4), 171-174.

Grabe, E. Warren, P. and Nolan, F. (1993) "Resolving category ambiguities-evidence from stress-shift" *Working Papers in Phonetics*, Lund University, 24-27.

Grabe, E. and Low, E. L. (2002) "Durational variability in speech and the rhythm class hypothesis" C. Gussenhoven and N. Warner (eds.) *Papers in Laboratory Phonology 7*, Berlin, New York: Mouton de Gruyter.

Grover, C. N. and Terken, J. M. B. (1995) "The role of stress and accent in the perception of speech rhythm" *IPO Annual Progress Report* (30), 30-37.

Grønnum, N. (1993) "Rhythm in regional variants of Standard Danish" Proceedings of the ESCA Workshop on Prosody Lund University, Working Papers in Phonetics, 20-24.

Gustafson, K. (1988) "The graphical representation of rhythm" *PROPH, Progress reports from Oxford Phonetics Laboratory* (3).

Garding, E (1975) "The influence of tempo on rhythmic and tonal patterns in three Swedish dialects" *Working Papers, Phonectis Lab Lund* (12), 71-83.

Garding, E. and Zhang, J. (1997) "Tempo effects in Chinese prosodic patterns" *Proceedings of ESCA Intonation Workshop Athens*, 145-148.

Harding, D. W. (1976) Words in rhythm: English speech rhythm in verse and prose. Cambridge: CUP.

Haselager, G. J. T., Sils, I. H. and Rietveld, A. C. M. (1991) "An alternative method of studying the development of speech rate" *Clinical Linguistics and Phonetics* (5), 53-63.

Hawkins, S. and Warren, P. (1994) "Phonetic influences on the intelligibility of converational speech" *Journal of Phonetics* (22), 493-511.

Hayes, B. (1984) "The phonology of rhythm in English" *Linguistic Inquiry* (15. Jan), 33-74.

Heilmann, H. (1939) Akzent und Wortkörper in der Sprache des Prof. A. Lloyd James. Berlin: de Gruyter.

Hensel, Walter (1940) Intonation und Lautgebung psychologisch verschiedenartige Vortragsstuecke bei Lloyd James. Berlin: de Gruyter.

Hetrich, I. and Ackermann, H. (1995) "The influence of slowed speech rate on coarticulation: acoustic analysis of durational and spectral parameters" *Proceedings of the 13th ICPhS*, 590-593.

Heuft, B. (1999) *Eine Prominenzbasierte Methode zur Prosodieanalyse- und synthese* (Dissertation, Institut fuer Kommunikationsforschung und Phonetik, University of Bonn). Frankfurt a.M.: Peter Lang.

Hoequist, C. (1983a) "Parameters for speech rate perception" *Aipuk (Arbeitsberichte Institut für Phonetik Kiel)* (20), 99-124.

Hoequist, C. (1983b) "Syllable duration in stress-, syllable- and mora-timed languages" *Phonetica* (40), 203-237.

Hoequist, C. (1983c) "Durational correlates of linguistic rhythm categories" *Phonetica* (40), 19-31.

Hoequist, C. (1983d) "The Perceptual Center and Rhythm Categories" *Language and Speech* (26, 4), 367-376.

Huggins, A. W. F. (1972a) "On the perception of temporal phenomena in speech" *Journal of the Acoustical Society of America* (51), 1279-1290.

Huggins, A. W. F. (1972b) "Just noticeable differences for segment duration in natural speech" *Journal of the Acoustical Society of America* (51), 1270-1278.

Huggins, A. W. F. (1975) "On isochrony and syntax" G. Fant and M. A. A. Tatham (eds.) *Auditory Analysis and Perception in Speech.* New York: Academic Press, 455-464.

Huggins, A. W. F. (1978) "Speech timing and intelligibility" J. Requin (ed.) *Attention and Performance* (VII). Hillsdale, N.J.: Erlbaum.

Hulbert, H. H. (1925) Rhythm in Speaking. Poetry in Speech: Speech in Poetry. London: Erskine MacDonald.

Janse, E. (2003) Production and Perception of fast speech. Utrecht: Lot.

Jokisch, O., Hirschfeld, D., Eichner, M., and Hoffmann, R. (1998) Creating an individual speech rhythm: a data driven approach. . *3rd ESCA/COCOSDA Synthesis Workshop*, 115-119.

Kaltenbacher, E. (1997) "German speech rhythm in L2 acquisition" J. Leather and A. James (eds.) New Sounds '97. *Proceedings of the Third International Symposium on the Acquistion of Second-Language Speech*. Klagenfurth: University of Klagenfurth, 158-166.

Keller, E. and Zellner, B. (1995) "A statistical timing model for French" *Proceedings of the 13th ICPhS* (3), 302-305.

Keller, E. and Zellner, B. (1996) "A timing model for fast French" *York Papers in Linguistics* (17), 53-75.

Kessinger, R. H. and Blumstein, S. E. (1998) "Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies" *Journal of Phonetics* (26), 117-128.

Kim, J., Davis, C. and Cutler, A. (forthcoming) Perceptual tests of rhythmic similarity: II. Syllable Rhythm. To appear in *Language and Speech.*

Klatt, D. H. (1973) "Interaction between two factors that influence vowel duration" *Journal of the Acoustical Society of America* (54), 1102-1104.

Kohler, K. J. (1982) "Rhythmus im Deutschen" *Arbeitsberichte Institut fuer Phonetik Kiel* (19), 89-105.

Kohler, K. J. (1983) "Stress-timing and speech rate in German: A production model" *Arbeitsberichte Institut für Phonetik Kiel* (20), 7-53.

Kohler, K. J. (1991) "Isochrony, units of rhythmic organization and speech rate" *Proceedings of the 12th ICPhS* (1), 257-261.

Kohler, K. J. (1995) *Einführung in die Phonetik des Deutschen.* Berlin: Erich Schmidt.

Krishnan, G. and Ward, W. (1998) "Temporal organisation of speech for normal and fast rates" *Proceeding of ICSLP* (3), 617-618.

Ladefoged, P. (1967) Three areas of experimental phonetics. Oxford: OUP.

Ladefoged, P. (1975) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Laeufer, C. (1995) "Effects of tempo and stress on German syllable structure" *Journal of Linguistics* (31), 227-266.

Lane, H. and Grosjean, F. (1973) "Perception of reading rate by speakers and listeners" *Journal of Experimental Psychology* (97, 2), 141-147.

Laver, J. (1994) *Principles of Phonetics*. Cambridge: CUP.

Lee, H. B. (1982a) "An acoustic study of rhythm in Korean" *Phonetics, The Phonetic Society of Korea* (4), 31-48.

Lee, H. B. (1982b) "Phonetic variation of Korean speech sounds as conditioned by tempo and rhythm" *Language Research* (18, 1), 115-130.

Lehiste, I. (1987) "Stress-timing and syllable-timing: evidence from foreign accents" *The Nordic Languages and Modern Linguistics* (6), 225-233.

Lehiste, I. (1972) "The timing of utterances and linguistic boundaries" *Journal of the Acoustical Society of America* (51), 2018-2024.

Lehiste, I. (1973) "Rhythmic units and syntactic units in production and perception" *Journal of the Acoustical Society of America* (54), 1228-1234.

Lehiste, I. (1977) "Isochrony reconsidered" *Journal of Phonetics* (5), 253-263.

Lehiste, I. (1979) "Temporal relations within speech units" *Proceedings of the 9th ICPhS* (2), 241-244.

Liberman, M. and Prince, A. (1977) "On stress and linguistic rhythm" *Linguistic Inquiry* (8, 2), 249-336.

Likert, Rensis (1932). "A Technique for the Measurement of Attitudes" *Archives of Psychology* (140), 1–55.

Lloyd James, A. (no year) *Bible Readings*. London: Linguaphone.

Lloyd James, A. (1929) Historical introduction to French Phonetics. London: ULP.

Lloyd James, A. (1935) *The broadcast word*. London: Kegan Paul, Trench, Trubner.

Lloyd James, A. (1937) Talks on English Speech: A Short Gramophone Course of English Pronunciation Place and publisher unknown.

Lloyd James, A. (1938) *Our Spoken Language*. London: Nelson.

Lloyd James, A. (1940) *Speech Signals in Telephony.* London: Pitman.

Low, E. L. (2000) "Is lexical stress placement different in Singapore English and British English?" A. Brown, D. Deterding, and E. L. Low (eds.) *The English Language in Singapore: Research on Pronunciation*. Singapore: Singapore Association of Applied Linguistics, 22-34.

Low, E. L. & Grabe, E. (1995) "Prosodic patterns in Singapore English" *Proceedings of the Intonational Congress of Phonetic Sciences* (3), 636-639.

Low, E. L., Grabe, E., and Nolan, F. (2000) "Quantitative characterisation of speech rhythm: syllable-timing in Singapore English" *Language and Speech* (43, 4), 377-402.

Maccoll, D. S. (1914) *Rhythm in English Verse, Prose, and Speech*. London: English Association. Essays and Studies, vol. 5.

Major, R. C. (1981) "Stress-timing in Brazilian Portuguese" *Journal of Phonetics* (9), 343-351.

Martin, J. G. (1975) "Rhythmic Expectancy in Continuous Speech Perception" A. Cohen and S. Nooteboom (eds.) *Structure and Process in Speech Perception*. Heidelberg: Springer, 161-177.

Massaro, D. W. (1986) "A new perspective and old problems" *Journal of Phonetics* (14), 69-74.

Meinhold, G. (1967) "Quantität und Häufigkeit von Pausen in gelesenen deutschen Texten im Zusammenhang mit dem Sprechtempo" *Gesellschafts- u. Sprachwissenschaft* (16, 1), 107-111.

Miles, D. W. (1937) "Preferred rates in rhythmic responses" *Journal of Genetic Psychology* (16), 427-469.

Milne, B. L. (1955) *Modern Speech Rhythm Exercises*. London: Macmillan.

Miller, J. L. and Grosjean, F. (1981) "How components of speaking rate influence perception of phonetic segments" *Journal of Experimental Psychology* (7, 1), 208-215.

Miller, J. L., Grosjean, F. and Lomanto, C. (1984) "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications" *Phonetica* (41), 215-225.

Miller, M. (1984) "On the perception of rhythm" *Journal of Phonetics* (12), 75-83.

Miner, J. B. (1903) "Motor, Visual and applied rhythms: An experimental study and a revised explanation" *The Psychological Review* (5/4), no page numbers.

Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (1981) "Prosodic aspects of American English speech rhythm" *Phonetica* (38), 84-106.

Nazzi, T., Bertocini, J., and Mehler, J. (1998) "Language discrimination by newborns: Towards an understanding of the role of rhythm" *Journal of Experimental Psychology: Human Perception and Performance* (24), 756-766.

Nespor, M. (1990) "On the rhythm parameter in phonology" I. M. Roca (ed.) *Logical Issues in Language Acquisition*, 157-175.

Nooteboom, S. G. (1973) "The perceptual reality of some prosodic durations" *Journal of Phonetics* (1), 25-45.

Ogden, C. K. (1940) Basic English ABC: Mit einer Darstellung der Laute des Basic English von A. LLoyd James. Cambridge: Orthological Institute.

O'Connor, J. D. (1965) "The perception of time intervals" *UCL Working Papers in Phonetics* (2).

O'Connor, J. D. (1968) "The duration of the foot in relation to the number of component sound-segments" *UCL Working Papers in Phonetics* (3).

Ohala, J. J. (1975) "The temporal regulation of speech" G. Fant and M. A. A. Tatham (eds.) *Auditory Analysis and Perception in Speech.* New York: Academic, 431-453.

Oller, D. K. (1979) "Syllable timing in Spanish, English, and Finish" H. Hollien and P. Hollien (eds.) *Current Issues in the Phonetic Sciences*. Amsterdam: Benjamins, 331-343.

den Os, E. (1988) Rhythm and tempo in Dutch and Italian. Utrecht: Elinkwijk.

den Os, E. (1985) "Perception of speech rate of Dutch and Italian Utterances" *Phonetica* (42), 124-134.

Osser, H. and Peng, F. (1964) "A cross cultural study of speech rate" *Language and Speech* (7), 120-125.

Patel, A. D., Löfqvist, A., and Naito, W. (1999) "The acoustics and kinematics of regularly-timed speech: A database and method for the study of the P-center problem" *Proceedings of the 14th ICPhS* (1), 405-408.

Pfitzinger, H. R. (1998) "Local speech rate as a combination of syllable and phone rate" *Proceedings of the ICSLP, Sydney,* 1087-1090.

Pfitzinger, H. R. (1999) "Local speech rate perception in German speech" *Proceedings of the 14th ICPhS*, 893-896.

Pijper, J. R. de und A. A. Sanderman (1994) "On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues" *Journal of the Acoustical Society of America* (96/4) 2037-2047.

Pike, K. L. (1946) *Intonation of American English*. Ann Arbor: University of Michigan.

Pointon, G. E. (1980) "Is Spanish really syllable-timed?" *Journal of Phonetics* (8), 293-304.

Pompino-Marschall, B. (1990) *Die Silbenprosodie*. Tubingen : Niemeyer.

Port, R. F., Tajima, K., and Cummins, F. (1998) "Speech and Rhythmic Behaviour" G. J. P. Savelsbergh, H. v. d. Maas and P. C. L. v. Geert (eds.) *The Non-linear Analyses of Developmental Processes*. Amsterdam: Royal Dutch Academy of Arts and Sciences.

Povel, D. J. (1980) *The role of rhythm in speech perception*. Nijmegen: Katholieke Universiteit.

Rammsayer, T. H. and S. D. Lima (1991) "Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors" *Perception & Psychophysics* (6) 565-574.

Ramus, F. and Mehler, J. (1999) "Language identification with suprasegmental cues: A study based on speech resynthesis" *Journal of the acoustical society of America* (105, 1), 512-521.

Ramus, F., Nespor, M., and Mehler, J. (1999) "Correlates of linguistic rhythm in the speech signal" *Cognition* (73), 265-292.

Rietveld, A. C. M., and Gussenhoven, C. (1987) "Perceived speech rate and intonation" *Journal of Phonetics* (15), 273-285.

Rincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., and Mehler, J. **(2005)**. "The role of speech rhythm in languages discrimination: further tests with a non-human primate" *Developmental Science* (8, 1), 26-35.

Roach, P. (1982) "On the distinction between 'stress-timed' and 'syllable-timed' languages" D. Crystal (ed.) *Linguistic controversies*. London: Edward Arnold, 73-79.

Roach, P. (1998) "Some Languages are Spoken More Quickly Than Others" P. Trudgill and L. Bauer (eds.) *Language Myths.* Penguin, 150-158.

Roach, P., Miller, D. and Sergeant, P. (1992) "Syllabic consonants at different speaking rates: A problem for automatic speech recognition" *Speech Communication* (11), 475-479.

Roe, R. B. (1823) The Principles of Rhythm, both in Speech and Music, Especially as Exhibited in the Mechanism of English Verse. Dublin: (Publisher unknown).

de Rodriquez, B. C. (1983) "Recognizing and producing English rhythmic patterns" A. Brown (ed.) *Teaching English Pronunciation: A book of readings*. London: Routledge, 350-356.

Scott, D. R., Isard, S. D., and De Boysson-Bardies, B. (1985) "Perceptual isochrony in English and in French" *Journal of Phonetics* (13), 155-162.

Seton, J. C. (1989) A psychophysical investigation of auditory rhythmic beat perception. York: University of York.

Shockey, L., Gregorski, R. and Lehiste, I. (1971) "Word-Unit Temporal Organization" *Journal of the Acoustical Society of America* (50), 117.

Singleton, R. C. (1855) Virgilus Maro: The works of Virgil, closely rendered into English rhythm, and illustrated from British poets of the 16th, 17th , and 18th centuries. By R. C. Singleton. London: (Publisher unknown).

Smiljanic, R. and Bradlow, A. (forthcoming) Temporal organization of English clear and plain speech. To appear in *Journal of the Acoustical Society of America.*

Smith, R. M. and Fowler, C. A. (1985) "The effecs of speaking rate on the isochrony of monosyllables" *Journal of the Acoustical Society of America* (1, 77), 52-53.

Sonnenschein, E. A. (1925) What is Rhythm?An essay accompanied by an appendix on experimental syllable measurements. Oxford: Basil Blackwell.

Sonntag, G. (1999) *Evaluation von Prosodie* (Dissertation, Institut für Kommunikationsforschung und Phonetik, Universität Bonn). Aachen: Shaker.

Steiner, I. (2004) *Zur Rhythmusanalyse mittels akustischer Parameter*. Unpublished MA-Thesis, Institute of Communication Research and Phonetics, University of Bonn, Germany.

Strangert, E. (1985) *Swedish speech rhythm in a cross-language perspective.* (PhD dissertation) Umea: Universitetet i Umea.

Strangert, E. (1987) "Major determinants of speech rhythm: a preliminary model and some data" *Proceedings of the 11th ICPhS* (2), 149-152.

Tarui, T. (1983) Effect of production practice on acquisition of English rhythm. Sophia: Sophia University.

Tauroza, S. and Allison, D. (1990) "Speech rates in British English" *Applied Linguistics* (11), 90-105.

Thomson, W. (1904) *The Basis of English Rhythm*. Glasgow: Holmes.

Thomson, W. (1916) *Laws of Speech-Rhythm*. Glasgow: Maclehose & Sons.

Thomson, W. (1923) *The Rhythm of Speech.* Glasgow: Maclehose, Jackson.

Toro, J. M., Trobalon, J. B., and Sebastian-Galles, N. **(2003)**. "The use of prosodic cues in language discrimination tasks by rats" *Animal Cognition* (6, 2), 131-136.

Triplett, N. and Sanford, E. C. (1901) "Studies of rhythm and metre" *American Journal of Psychology* (12), 361-387.

Trouvain, J., and Grice, M. (1999) "The effect of tempo on prosodic structure" *Proceedings of the 14th ICPhS* (2), 1067-1070.

Tuller, B., and Fowler, C. (1980) "Some articulatory correlates of perceptual isochrony" *Perception and Psychophysics* (27, 4), 277-283.

Uldall, E. T. (1972) "Relative durations of syllables in two-syllable rhythmic feet in RP" *Work in Progress, Department of Linguistics*, Edinburgh University (5), 110-111.

Van Son, R. J. J. H. (2002) "Can standard analysis tools be used on decompressed speech?" *Electronic Proceedings of the COCOSDA2002 meeting*, Denver. (URL:http://www.cocosda.org/meet/denver/COCOSDA2002-Rob.pdf).

Van Son, R. J. J. H., and Pols, L. C. W. (1989) "Comparing formant movements in fast and normal rate speech" *Proceedings of Eurospeech*, Paris (2), 665-668.

Van Son, R. J. J. H., and Pols, L. C. W. (1992) "Formant movements of Dutch vowels in a text, read at normal and fast rate" *Journal of the Acoustical Society of America* (92, 1), 121-127.

Vaissière, J. (1983) "Language-independent prosodic features" A. Cutler und R. Ladd (Hg.) *Prosody: models and measurements*. Berlin: Springer, 53-66.

Wagner, P. (2002) *Vorhersage und Wahrnehmung deutscher Wortbetonung*. PhD Dissertation, Institut für Kommunikationsforschung und Phonetik, Bonn Universtity.

Wallin, J. E. W. (1901) "Researches on the Rhythm of Speech" *Studies from the Yale Psychological Laboratory* (9), 1-142.

Wenk, B., and Wioland, F. (1982) "Is French really syllable-timed?" *Journal of Phonetics* (10), 193-216.

Wenk, B. J. (1985) "Speech rhythms in second language acquisition" *Language and Speech* (28, 2), 157-175.

White, L. and Mattys, S. (in print) Calibrating rhythm: first language and second language studies. To appear in *Journal of Phonetics.*

Whiteside, S. (1996) "Temporal-based acoustic-phonetic patterns in read speech: some evidence for speaker sex differences" *Journal of the International Phonetic Association* (26, 1), 23-40.

Wiik, K. (1991) "On a third type of speech rhythm: foot timing" *Proceedings of the 12th ICPhS* (3), 298-301.

Wish, M. and Carroll, J. D. (1974) "Perception of rhythm and accent in words and phrases" E. C. Carterette and M. P. Friedman (eds.) *Handbook of Perception* (2). New York: Academic, 462-470.

Wong, R. (1987) Teaching Pronunciation: Focus on English rhythm and intonation Englewood Cliffs, NJ: Prentice Hall Regents.

Zellner, B. (1994) "Pauses and the temporal structure of speech" E. Keller (ed.) *Fundamentals of speech synthesis and speech recognition.* London: John Wiley and Sons Ltd., 41-62.

Zellner, B. (1998a) "Fast and slow speech rate: a characterization for French" *Proceedings of ICSLP* (7), 3159-3163.

Zellner, B. (1998b) "Temporal structures for fast and slow speech rate" *ESCA/COCOSDA, Third international workshop on speech synthesis*, Jenolan Caves (Australia), 143-146.

Zellner Keller, B. (forthcoming) Prediction of Temporal Structures for Various Speech Rates. In: N. Campbell (ed.) Volume on Seech Synthesis, Springer.

Zellner Keller, B. and Keller, E. (forthcoming a) "The primordial nature of speech rhythm" P. Delcloque and V. M. Holland (eds.) *Integrating Speech Technology in Language Learning,* Swets and Zeitlinger.

Zellner Keller, B. and Keller, E. (forthcoming b) "Representing Speech Rhythm" *Working papers of COST 258*, University of Lausanne.

Xu, Y., Wang, M. (forthcoming) "The nature of syllable organization – Evidence from f0 and duration patterns in Mandarin".