Choosing the right rate normalization method for measurements of speech rhythm

Dellwo, Volker

CHOOSING THE RIGHT RATE NORMALIZATION METHOD FOR MEASUREMENTS OF SPEECH RHYTHM

Volker Dellwo
Division of Psychology and Language Sciences, University College London
v.dellwo@ucl.ac.uk

1. ABSTRACT

Some acoustic correlates of language rhythm are durational characteristics of consonants and vowels. The present study investigates the influence of speech rate on these acoustic correlates. In experiment I four widely applied correlates of speech rhythm (%V, ΔC, nPVI and rPVI) were correlated with the rate of consonantal and vocalic intervals using speech from five different languages (Czech, English, French, German, Italian) that was characterized by high tempo variability within each language (very slow to very fast produced speech). It was found that rhythm measures based on consonantal interval durations (ΔC, rPVI) correlate negatively with rate measures and that rhythm measures based on vocalic intervals (%V, nPVI) are not influenced by rate. In experiment II the effectiveness of rate normalization procedures on the rate dependent measures, ΔC and rPVI, was tested by correlating these measures with speech rate before and after normalization using the same speech data as in Experiment 1. ΔC was normalized by logarithmically transforming the consonantal interval durations and rPVI was normalized by previously proposed ways for the normalization of nPVI. It was found that rate effects on ΔC and rPVI could be normalized for effectively using the suggested rate normalization procedures. In Experiment III it was tested whether rate normalized measures of speech rhythm support the impression that some languages can be categorized according to their auditory rhythmic characteristics (e.g. stress- and syllable-timing). Strong support for this was only found for the rate normalized rPVI why the normalized ΔC revealed mixed results. It was concluded that ΔC is less appropriate for rhythmic measurements that aim to separate languages of different rhythmic classes.

2. INTRODUCTION

The systematic study of speech rhythm began towards the beginning of the past century. It was motivated by assumptions that rhythm plays an important role in acquiring a correct pronunciation in a foreign language and thus enhancing non native speech intelligibility (James, 1929) or simply for phonetic classification purposes (Classe, 1939; Pike, 1945; Abercrombie, 1967). More recent findings demonstrating that knowledge about the rhythm of a language is crucial for predicting word boundaries (Cutler, 1997; Cutler & Norris, 1988; Kim, Davis & Cutler, 2008) or that rhythmic cues can help infants segregating between different languages when growing up in a bilingual environment (Nazi *et al.*, 1988; Ramus *et al.*, 1999) gave rise to a growing interest in the field of speech rhythm.

Rhythmic variability in speech is manifold. Languages, for example, can posses specific auditory rhythmic characteristics (James, 1929; Classe, 1939; Pike, 1945; Abercrombie, 1967; Ramus *et al.*, 1999; Grabe & Low, 2002) but there is also rhythmic variability within a language. Native-speakers can sound rhythmically different from non-

native speakers (White & Mattys, 2007; Mok & Dellwo, 2008) and different language varieties may be characterized by different rhythmic features (e.g., Singaporean and Standard Southern British English: Low, Grabe, & Nolan, 2000; Deterding, 2001). Even speakers of the same language variety may differ in speech rhythm (Dellwo & Koreman, 2008) and rhythm may vary within the same speaker depending, for example, on emotional state (Cahn, 1990). One of the most central questions in the field of speech rhythm has been how such auditory rhythmic variability can be measured in the speech signal or in other words: What are the acoustic correlates of speech rhythm? This question turned out to be difficult to answer since unlike other perceptual prosodic phenomena like intonation, which is mainly encoded by fundamental frequency variability, it seems less clear which acoustic phenomenon is responsible for the percept of speech rhythm. It is also likely that rhythm is encoded by a number of different acoustic parameters like fundamental frequency, amplitude, and duration, and possibly our perception of rhythm results from a complex interaction between those parameters. It is further unclear whether the perception of rhythm by listener groups with varying linguistic background (e.g. different native languages) is based on the same acoustic parameters in the same way. Given that listeners of different languages, for example, make different use of prosodic stress correlates (Wang, 2008) it seems conceivable that a similar situation is true in the case of speech rhythm.

Of all the rhythmic variability in speech, the variability of rhythm between languages has, without doubt, been studied most. And it is probably for this reason that measures of speech rhythm have mostly been developed to capture between-language rhythmic variability in the speech signal. How can this between-language variability be characterized? There have been numerous attempts to categorize languages that share auditory rhythmic features (Classe, 1939; James, 1929; Pike, 1945; Abercrombie, 1967; Ramus *et al.*, 1999; Grabe & Low, 2002). Such rhythmic features were metaphorically described as sounding either more like a 'machine-gun' (e.g. French, Spanish and Yoruba) or more like a 'Morse code' (e.g. English, German, and Arabic). This comparison, introduced by James (1929) and still widely used in present times, reveals the idea that some languages appear more regularly timed than others (machine-gun vs. Morse-code respectively). This regular timing was initially assumed to be manifested in regular (or quasi-isochronous) syllabic durations in machine-gun languages and irregular (or non-isochronous) syllabic durations in Morse-code languages. Languages that were assumed to have regularly timed syllables were therefore referred to as 'syllable-timed'. The assumed lack of durational syllable regularity in Morse-Code languages led to the idea that the intervals between stressed syllables (inter-stress intervals) in such languages are timed regularly (irrespective of the number of unstressed syllables they contain). Languages revealing these assumed acoustic characteristics were therefore called 'stress-timed' languages. However, it remains unclear why stress-timed languages were (and often still are) assumed to have regularly timed inter-stress intervals. No study reports auditory support for such an assumption. It thus seems conceivable that the idea of quasi-isochronously timed inter-stress intervals in Morse-code languages was created merely as an acoustic analogy to the isochronous timing of syllables in syllable-timed languages. The early assumptions that language characteristic rhythm stands in relation with the timing of syllables or inter-stress intervals is probably the reason for the fact that most of the attempts to measure speech rhythm in the acoustic signal are based on measuring

segmental durational characteristics of speech and widely disregard other durational (e.g. the timing of fundamental frequency contours) or spectral parameters (e.g. fundamental frequency variability, dynamic variability, etc.; see Tilsen and Johnson, 2008, for an alternative approach).

It seems not surprising that the earliest attempts to measure stress- and syllable-timing in the speech signal were based on measuring the durational variability of syllables and inter-stress intervals in these languages, assuming that the variability of syllable durations should be lower and the variability of inter-stress interval durations should be higher in syllable- than in stress-timed languages. Countless approaches have been carried out from the 1960s to the end of the 1980s to find evidence for this assumption, however, no support has ever been found (see Ramus *et al.*, 1999; Grabe & Low, 2002, for reviews of the literature). It thus seems that the use of the terminology 'stress-timing' and 'syllable-timing' should be discontinued and be replaced by terminology closer reflecting the auditory impression of rhythm like 'regular' vs. 'irregular' rhythm. (The present article makes a first approach to do this. However, given the continuous wide usage of the terminology 'stress' and 'syllable' timing it will here continued to be used in parallel.)

In search for acoustic regularity and irregularity in syllable- and stress-timed languages respectively, a number of studies started reporting first success when shifting the unit of analysis from syllable and inter-stress interval durations to consonantal (C) and vocalic (V) interval durations[1] (Ramus *et al.*, 1999; Grabe & Low, 2002; Dellwo, 2006; Mattys and White, 2007). It was assumed that the durational characteristics of C and V intervals are influenced by language specific phonological features and that languages sharing such features appear rhythmically similar (Bolinger, 1981; Roach, 1982; Dauer, 1983, 1987; Ramus et al, 1999; Grabe & Low, 2002). Stress-timed languages typically share the feature of a high complexity of C intervals (i.e. allowing multiple consonants in a C interval) which leads to a high variability of C interval durations. C intervals in syllable-timed languages typically consist of only one consonant. Further, stress-timed languages typically share the feature of reducing vowels to schwa, which is believed to introduce high durational variability of V intervals. Such reduction phenomena are untypical in syllable-timed languages. Given these assumptions four measures were proposed which have been widely applied in the field of speech rhythm measurements. Two measures were proposed by Ramus *et al.* (1999). They are the standard deviation of C intervals (ΔC) and the percentage over which speech is vocalic (%V). Two further measures have been proposed by Grabe & Low (2002) which calculate the average differences between consecutive C intervals and V intervals and are known as the Pairwise Variability Index (PVI). The PVI applied to V intervals was rate normalized (nPVI, discussion follows) and the PVI for C intervals was not rate normalized (thus referred to as 'raw'; rPVI). All before mentioned measures are basically influenced by the durational variability of C and

[1] A C interval is the duration of a string of consonants between two vowels (or any combination of vowel and pause) and, likewise, a V interval is a string of vowels between two consonants (or any combination of consonant and pause) in speech. Both C and V intervals may contain a syllable, word or even sentence boundary. Mind that some studies refer to C intervals as 'inter-vocalic-intervals' (Grabe & Low, 2002), however, draw no methodological difference.

V intervals: high variability leads to high values, low variability leads to low values.[2] Numerous studies have demonstrated that such variability measurements support the distinction at least of some languages into rhythmic categories (Ramus *et al.*, 1999; Grabe & Low, 2002; Dellwo & Wagner, 2003; White & Mattys, 2007; Mok & Dellwo, 2008). It thus seems conceivable that the auditory impression of rhythmic regularity and irregularity in syllable- and stress-timed languages respectively, is the result of acoustic parameters like ΔC, %V, or the PVI measures. Support for this theory has been found for the parameters ΔC and %V (Ramus *et al.*, 1999).

All measures mentioned above are based on durational characteristics of C and V intervals. Consequently speech produced at higher rates will shorten such intervals and speech at lower rates will lengthen them. Since the techniques involved in the production of speech vary widely across different sounds, it must be assumed that rate does not affect the duration of different segment types in the same way. Thus the durational characteristics like ΔC or PVI, for example, may vary when speakers speak faster or slower and such changes are unlikely to be of a linear nature (Ramus, 2003; Grabe & Low, 2002; Dellwo & Wagner, 2003; Barry *et al.*, 2003; Dellwo, 2006). Further, speech rate can influence such measures on two levels: First, given that the overall durations of C and V intervals change when speech is uttered faster or slower, it may have an influence on the within-language variability of C and V intervals. Second, the rate of C and V intervals can vary significantly between languages (Dellwo, 2008) for the same reasons that rhythmic properties vary. After all, less complex C intervals should on average be shorter thus should be produced at higher rates. For the rhythm measures this means that, given they interact with rate systematically, it may inevitably lead to situations in which an obtained acoustic rhythmic difference between two groups under observation is actually a rate difference between these groups.

A number of studies contain evidence that rhythmic measurements are influenced by rate parameters (Grabe & Low, 2002; Dellwo & Wagner, 2003; Barry *et al.*, 2003; Ramus, 2002; Dellwo, 2006; White & Mattys, 2007; Dellwo, 2008). These studies, however, have shortfalls: They are either based on data with a very poor rate variability (Grabe & Low, 2002; White & Mattys, 2007; Ramus, 2002), they did not correlate rhythm and rate measures systematically (Grabe & Low, 2002; Dellwo & Wagner, 2003; Barry *et al.*, 2003; Ramus, 2002), or they only looked at very selected rhythmic measures (Grabe & Low, 2002; Dellwo & Wagner, 2003; Dellwo, 2006). The problem that rhythmic measures can interact with rate was addressed in other studies by suggesting rate normalization procedures for selected measures but they have not been seldom been without dispute. For example, Grabe & Low (2002) used a rate normalization method for vocalic rhythm measure (nPVI), but Barry *et al.* (2003) claims that the applied method does not fulfill its purpose. Dellwo (2006) introduced a rate normalization method for ΔC by measuring the standard deviation proportional to the mean (coefficient of variation). White & Mattys (2007) extended this technique to vocalic interval variability ΔV by measuring VarcoV.

[2] In case of %V the situation is slightly different. As a ratio measure it does not reveal the variability of V intervals. However, it is assumed to be influenced by variable consonantal complexity and vocalic reduction (high consonantal complexity may lead to a higher overall proportional content of C intervals, vocalic reductions may lead to shorter vowels and a lower overall V interval proportion; see Ramus *et al.*, 1999, for details).

Both Dellwo and White & Mattys found that this rate normalization can be of advantage, for example, as they found strong support for the rhythm class concept by the rate normalized measures. The present study, however, will reveal that the Varco-normalization can be problematic since the typical data distributions of C and V interval durations do not meet the assumptions underlying such calculations (see section 4). The same criticism applies for a rate normalization method for the PVI based on z-transformations of the raw data (Wagner & Dellwo, 2004).

Given the gap of systematic knowledge about the degree to which individual rhythm measurements are dependent on rate, and about the effectiveness of rate normalization procedures the present study has three aims:

(a) In experiment 1 (section 3) the influence of rate on the four most widely used rhythm measures %V, ΔC, consonantal PVI and vocalic PVI were studied.

(b) In experiment 2 (section 4) the effectiveness of rate normalization methods for rhythm measures was tested.

(c) In experiment 3 (section 5) it was tested how effectively rate normalized measures separate languages of different rhythm classes.

3. INFLUENCE OF RATE ON RHYTHM MEASURES

In the present experiment it was tested which of four widely used rhythm measures based on C and V interval variability (%V, ΔC, nPVI, rPVI) are influenced by rate variability of C and V data. To trigger effects of interval rate on the rhythm measures under investigation speech data was used in which speakers tried to vary their speech tempo from very slow to very fast, thus producing a wide range of interval rates.

3.1 Method

Speech Material: The speech material was taken from a database designed for speech rhythm and rate analysis at Bonn University and University College London (BonnTempo Corpus; Dellwo *et al.*, 2004; Dellwo, forthcoming). The speech material is based on a texts which 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). The languages were selected to represent both traditional rhythmic classes. 'Stress-timing' is represented by English and German, 'syllable-timing' by French and Italian and a language that is difficult to classify on an auditory basis, Czech.

The text was read by each speaker in each language under five different intended speech tempo versions. Speakers were first asked to read normal, then slow, then even slower (very slow), then fast, and then as fast as possible (see Dellwo *et al.*, 2004, for details of the recording procedure) leading to five different intended tempo categories from very slow to as fast as possible. This method resulted in a wide variety of C and V rates for each speaker and language recorded. Thus the data is highly suitable for studying effects of rate on rhythmic variability in speech.

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 1: 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 dataset of the present study

Measures and measurement units: Four measures were tested for rate influences listed below. Each measure was calculated for an interval of speech between naturally occurring pauses (inter-pause-interval) so that no pause content was included in the calculations. Typically inter-pause-intervals consisted of a clause or a sub-clause but this may vary slightly between different speakers and it may vary tremendously between different intended speech tempi (at higher intended tempi fewer pauses were produced).

The measurements of rhythm and rate were:

(a) The percentage over which speech is vocalic (%V). This measure is a ratio measure showing the proportional differences between the overall durational vocalic and consonantal content in speech (see APPENDIX I, Equation 2 for a formula).

(b) The standard deviation of C interval durations (ΔC; APPENDIX I, Equation 3).

(c) The 'raw' consonantal Pairwise Variability Index (rPVI). This measure calculates the average difference between consecutive C interval durations in each inter-pause-interval (APPENDIX I, Equation 4).

(d) The normalized vocalic Pairwise Variability Index (nPVI). This measure calculates averages of relative differences between consecutive V intervals. Relative differences are obtained by calculating each difference proportional to the overall duration of the two consecutive V intervals under observation (APPENDIX I, Equation 5).

(e) Rate was measured as the average number of C and V intervals (combined) per inter-pause-interval.

Procedure: Four widely used measures of speech rhythm (see introduction) were first cross-plotted against the rate of C and V intervals (CV interval rate) for a descriptive analysis. With a curve analysis procedure it was then tested which mathematical function best describes the relationship.

3.2 Results

Figure 1 shows %V (top left), ΔC (top right), nPVI (bottom left), and rPVI (bottom right) cross-plotted against CV-rate. Descriptively the graphs reveal clearly that there is a relationship between CV-rate and ΔC as well as rPVI. In both cases the relationship can be described as a negative correlation, i.e. an increase in CV-rate leads to a decrease in ΔC or rPVI. No relationship can be detected between either %V or nPVI and CV-rate.

A linear and a logarithmic curve have been fitted to all four data plots. As can be expected from the descriptive analysis the R square for the %V/CV-rate and the nPVI/CV-rate comparisons turned out very poorly. For %V both the linear and logarithmic curve return a value of 0.035 ($p>0.05$). A very similar situation is the case for nPVI (R square

linear: 0.03, logarithmic: 0.031; $p>0.05$). It can thus safely be concluded that there is no influence of CV-rate on either %V or nPVI.

For ΔC the returned R square value for the linear fit results in 0.535 and for the logarithmic fit in 0.63 ($p<0.005$ in both cases). The situation is very similar for rPVI (R square linear: .492, logarithmic: .577; $p<0.005$). It can thus be concluded that both ΔC and rPVI are highly dependent on variability of CV-rate and that a logarithmic model is best for describing the relationship between the two parameters.

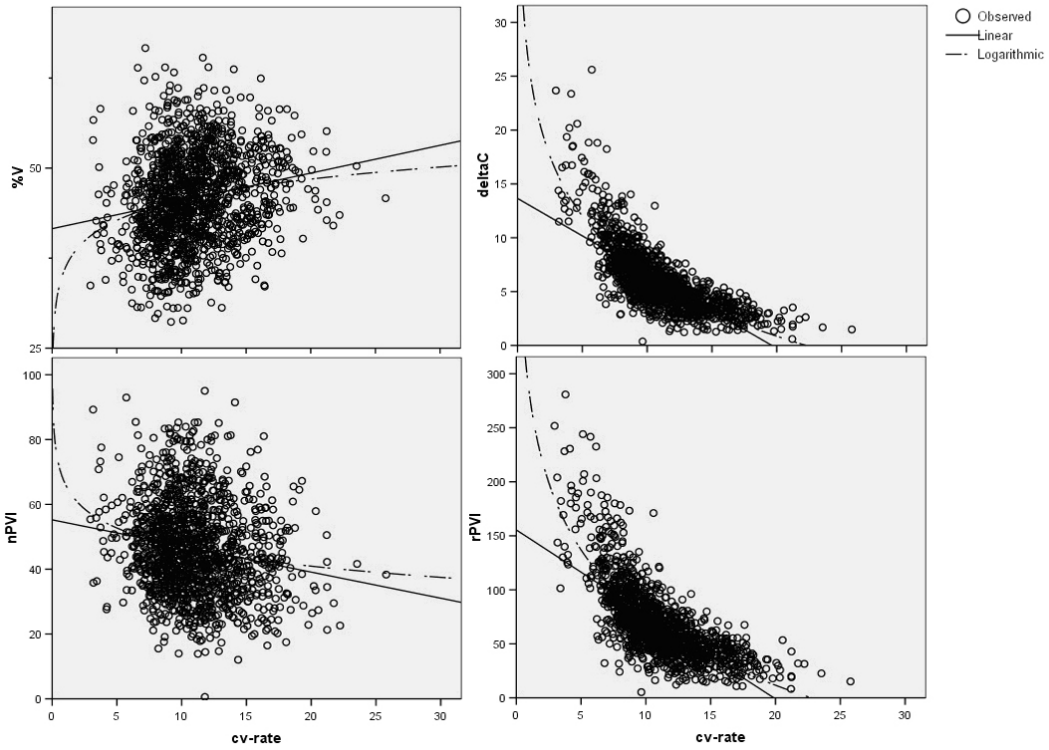

Figure 1: Scatter plot the rhythm measures %V (top left), ΔC (top right), rPVI (bottom left), and nPVI (bottom right) as a function of CV-rate with a linear and logarithmic curve fitted (each point is defined by the respective rhythm and rate values obtained from one inter-pause-interval)

3.3 Discussion

Both the V-interval variability measures %V and nPVI are clearly not dependent on CV-rate but the C interval measures rPVI and ΔC proved to be strongly affected. An explanation for this finding is straightforward: 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. The findings reveal that even in data with probably maximum CV-rate variability the measures %V and nPVI are not affected by the rate parameter. This demonstrates that the rate

normalization procedure applied for the nPVI (see Grabe & Low, 2002) is effective contrary to the concerns in Barry *et al.* (2003) where this normalization procedure was put into question (see Introduction).

The results also show clearly that a rate normalization is necessary both in case of ΔC and the rPVI. If these measures are not normalized they inevitably carry speech rate information thus it will be unclear to which extent obtained rhythmic differences between two groups may be the result of rate variability. Grabe & Low (2002) argue that rate normalization of the rPVI is problematic because rate differences in the rPVI will be a combined effect of speaking rate and between language differences in syllable structure. It remains unclear, however, why this interaction should prevent one from carrying out a normalization since both factors contribute to the same parameter: *rate*. After all it seems somehow irrelevant for measuring speech rhythm where rate influences derived from as long as they are in the signal. For this reason it is argued here that rPVI will need to be rate normalized if rhythmic measurements of c interval variability want to be obtained.

In the next section rate normalization methods will be applied for ΔC and rPVI and their effectiveness will be tested. After this (section 5) it will be tested how well the rate normalized ΔC and rPVI support the impression that languages of different rhythmic classes (stress- and syllable-timed) vary in rhythm.

4. NORMALIZING RHYTHM MEASURES FOR RATE

In this section appropriate rate normalization methods for ΔC and rPVI will be suggested and it will be tested how efficient they are by comparing the influence of speech rate on the measures before and after normalization.

4.1 Normalizing ΔC for rate

4.1.1 Data considerations for the calculation of ΔC

It was first tested whether the data assumptions are met for calculating ΔC. The calculation of a standard deviation (e.g. the standard deviation of C interval durations, ΔC) assumes the data to be normally distributed (Gaussian normal distribution). Thus, any study on speech rhythm based on the calculation of standard deviations of any interval in speech (C and V intervals, but also syllables or inter-stress intervals) needs to guarantee that the underlying data is normally distributed. However, interval durations in speech should not necessarily be assumed to be normally distributed. After all, speech segments have some threshold of maximum shortness (a segment can hardly be shorter than 5 ms) but there is probably no threshold limiting the length of a segment (especially vowels can be of very long duration, for example, as an effect of phrase final lengthening). For this reason it should be assumed that speech units of any type (consonants, C intervals, vowels, V intervals, 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 low frequency of data points of high durations. It should also be assumed that a higher proportion of values is distributed around the lower threshold of the data which leads to positive kurtosis. This assumption was tested by plotting the distributions for C, V and Syllabic (S) interval durations in the data set described in the previous section (BonnTempo Corpus; Dellwo *et al.*, 2004).

Results reveal that the assumption of non-normally distributed interval durations is supported by the data. Figure 2 (top) displays the distributions of C (left), V (right) and

Syllable (S; centre) durations. 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 (positive Skew). Furthermore the peak values of the central data are much higher than for a normal distribution (positive Kurtosis).

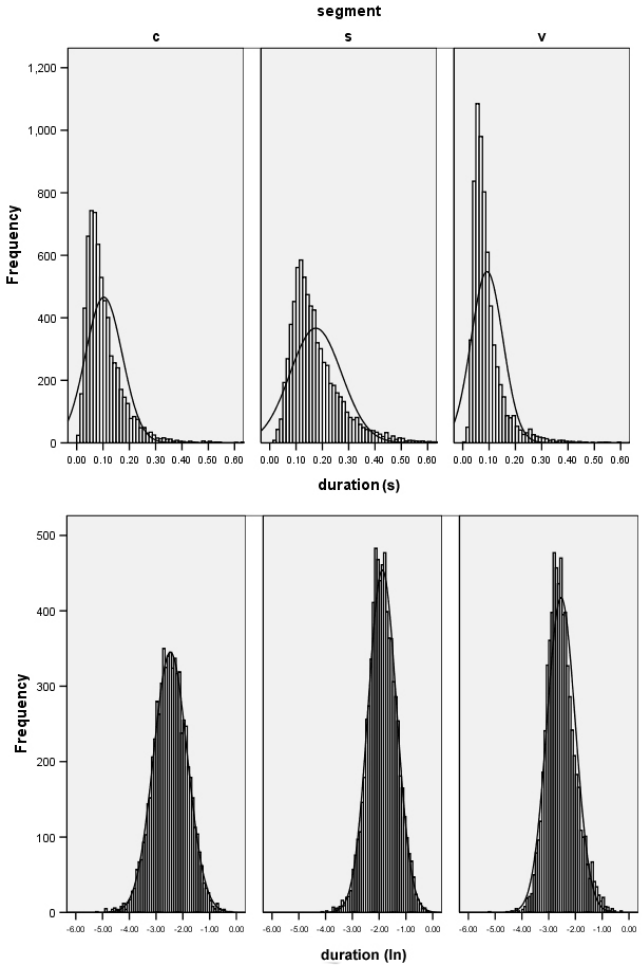

Figure 2: Histograms showing the distribution of C (left), V (right) and syllabic (centre) interval durations with superimposed Gaussian normal distribution for raw durations (top) and logarithmically transformed durations (bottom)

A common way of reducing positive Skew and Kurtosis is by calculating the logarithm for each interval duration (logarithmic transform, henceforth: ln transform), typically to the base e (Euler's number). Descriptive results of this transformation can be obtained from the histograms in Figure 2 (bottom). The figure contains the distributions for ln transformed C (left), S (centre) and V (right) intervals durations with superimposed

normal distribution lines. It is clearly visible that the interval duration distributions are now much closer to a Gaussian normal distribution and can thus be regarded as approximately normally distributed.

In addition to the descriptive analysis, Skewness and Kurtosis coefficients for the S, C and V intervals before and after the ln transformation were calculated and are obtainable from Table 2. The table displays the results for the raw (raw) and ln transformed data (ln). The table also contains the standard error for the deviation of ln transformed durations from the normal distribution. It is clear 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. However, such an amount of Skewness is still in an acceptable range. The ln transform 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 Kurtosis even came down close to 0 (0.084). The low standard errors (<0.034) for the ln data show that the ln transformed data distributions do not locally deviate much from the normal distributions.

	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 2: Values for Skewness and Kurtosis before (raw) and after (ln) ln transformation for syllabic (S), consonantal (C) and vocalic (V) intervals (standard error refers to deviation of the ln transformed data distribution from a normal distribution)

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 possibly strongly influence all analysis procedures assuming normally distributed data. This is clearly the case for ΔC as the calculation of a standard deviation assumes a normal (i.e. Gaussian) data distribution. It could be demonstrated that a logarithmic transformation of C-interval durations leads to a satisfactory normal distribution of the data.

4.1.2 Testing the influence of rate on ΔC based on ln transformed C durations

For this section ΔC was calculated based on ln transformed C interval durations (this measure is henceforth referred to as: ΔCln) and the influence of rate on this measure was tested with the same method as in section 3 (first, plotting ΔCln across CV-rate for a descriptive analysis and then correlating the two parameters to test the strength of the relationship). The results can be viewed in Figure 3 (left) which contains the cross-plot of ΔCln over CV-rate. The plot reveals that there is no obvious relationship between the two parameters. A linear regression analysis confirms this visual impression with a low and insignificant R-square value of 0.045 ($p>0.05$).

This result shows that the ln transform of the data is sufficient for normalizing CV-rate effects on ΔC. An explanation for the effect is straightforward. The ln transform leaves intervals of short durations at more or less equal duration whereas long durations are shortened drastically. In this respect, short durations from fast rates approximate long durations of slow rates. What remains is the proportional variability but no longer the high absolute interval differences.

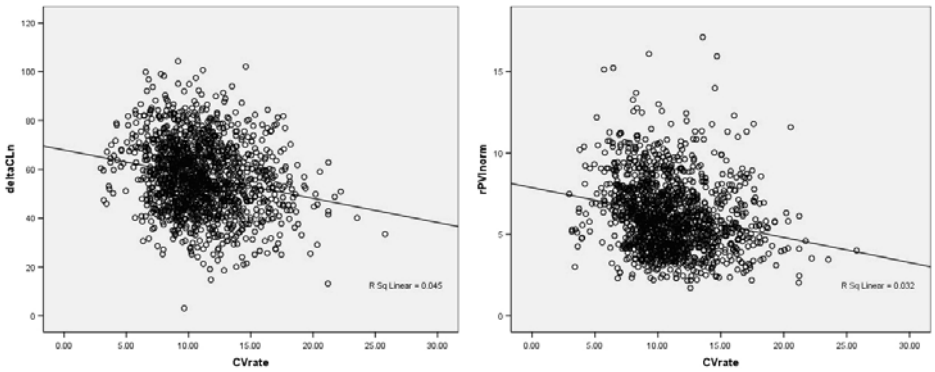

Figure 3: Scatter-plot showing ΔC based on ln transformed data (left) and rate normalized rPVInorm (right), both plotted over CV-rate

4.1.3 Discussion

In the previous two sections it was demonstrated, first, that the data distributions of C, V and S interval durations does not meet the assumptions necessary for the calculation of standard deviations (e.g. ΔC), second, that a logarithmic transform of interval durations changes the data distributions to fulfill the assumptions and, third, that ΔC based on ln transformed durations (ΔCln) is not speech rate dependent. As such, the ln transform of the data is a suitable rate normalization procedure for ΔC. Rate normalization methods based on the coefficient of variation of ΔC (varcoC =(ΔC*100)/meanC; Dellwo, 2006, White and Mattys, 2007) are based on the absolute standard deviation and it is questionable to what extent such measures are suitable for further statistical processing (e.g. for referential statistic methods assuming normally distributed data).

4.2 Normalizing rPVI for rate

4.2.1 Normalization procedure

Previous research applied rate normalization methods for rPVI by calculating the measures for z-transformed data (Wagner & Dellwo, 2004). However, also the z-transform assumes a normal distribution of the data which is not the case for C interval duration distributions (see above). A solution might be to apply the z-transform to logarithmically transformed durations. However, by performing a logarithmic transform of the data, large differences between consecutive C intervals become reduced drastically. It is probably counterproductive to reduce such differences as they are the differences that are the basis for rPVI variability. For this reason this technique was not further pursued in the present study.

In case of the PVI a rate normalization method already exists and is widely applied for the vocalic nPVI and it was demonstrated to be effective as the nPVI based on V interval durations revealed not to be dependent on rate (section 3). For this reason the same rate normalization method will be applied for rPVI. This is an easy procedure as only relative instead of absolute differences between consecutive C interval durations need to be averaged (see APPENDIX I, Equation 7). To avoid confusion with the vocalic nPVI measure the measure will be referred to as rPVInorm. This measure was correlated with CV-rate in the same way as ΔCln (above) using the same data as described in section 3.

4.2.2 Results

The results of the rate normalized consonantal rPVInorm can be viewed in Figure 3 (right plot). The graph shows the rPVInorm plotted across CV-rate and it is obvious that the normalization procedure is an effective control for speech rate in case of the consonantal variability measure. With an R square of 0.032 ($p>0.05$) as a result of a linear regression analysis it can be concluded that there is no correlation between the CV-rate and the rPVInorm.

4.2.3 Conclusions about rPVI normalization

It was demonstrated that the consonantal rPVI measure can be normalized effectively using the same rate normalization method as for the vocalic nPVI measure. Such a normalization method is probably more appropriate than existing methods using z-transform (Wagner & Dellwo, 2004) because the z-transform assumes normal distributions of the underlying data.

5. THE POWER OF RATE NORMALIZED RHYTHM MEASURES TO SEPARATE LANGUAGES OF DIFFERENT RHYTHMIC CLASSES

In section 3 it was demonstrated that the rhythmic correlates ΔC and nPVI correlate with the rate of C and V intervals. Section 4 showed how to normalize these measures for rate variability effectively. It remains to be tested whether the rate normalized measures still fulfill one of their main purposes, namely whether they support the auditory impression that languages of different rhythmic class sound different in their rhythm (more or less regularly timed; see introduction).

It is also possible that different languages react differently to rate normalizations. Languages with a low C interval complexity (e.g. French and Italian) may be able to maintain this complexity at high speeds while languages with a high C interval complexity (e.g. German and English) may reduce complex C intervals with increasing speed as a result of segment elision, for example. This may lead to a situation in which relative C interval variability measured by rate normalized measures (ΔCln and rPVInorm) changes with rate in some but not in other languages.

To compare speech rate influences on rhythm measures between languages, rates have to be made comparable across the different languages under investigation. For this reason rates in five parts of the total distributions (quintiles) were compared with each other. Quintiles were chosen because they roughly corresponded with the five intended tempo categories speakers produced. These intended tempo categories were not chosen as rate indicators as absolute speech rates within these categories may vary widely (see Dellwo, 2008)

5.1 Method

Speakers and Speech Data: The same data as in section 3 (Experiment I) was used in the present experiment.

Procedure: ΔCln and rPVInorm were calculated for each quintile of CV-rate for each of the five languages (Czech, English, French, German, Italian). Mean values of the results for each of the five languages were plotted across the five rate quintiles and ANOVAs were processed to analyze within and between language variability.

5.2 Results

Figure 4 contains the descriptive results for ΔCln (top) and rPVInorm (bottom) before (left) and after (right) normalization. Each graph contains the mean values for each respective measure and language at each quintile of CV-rate. Mean values were interpolated in each language with a line.

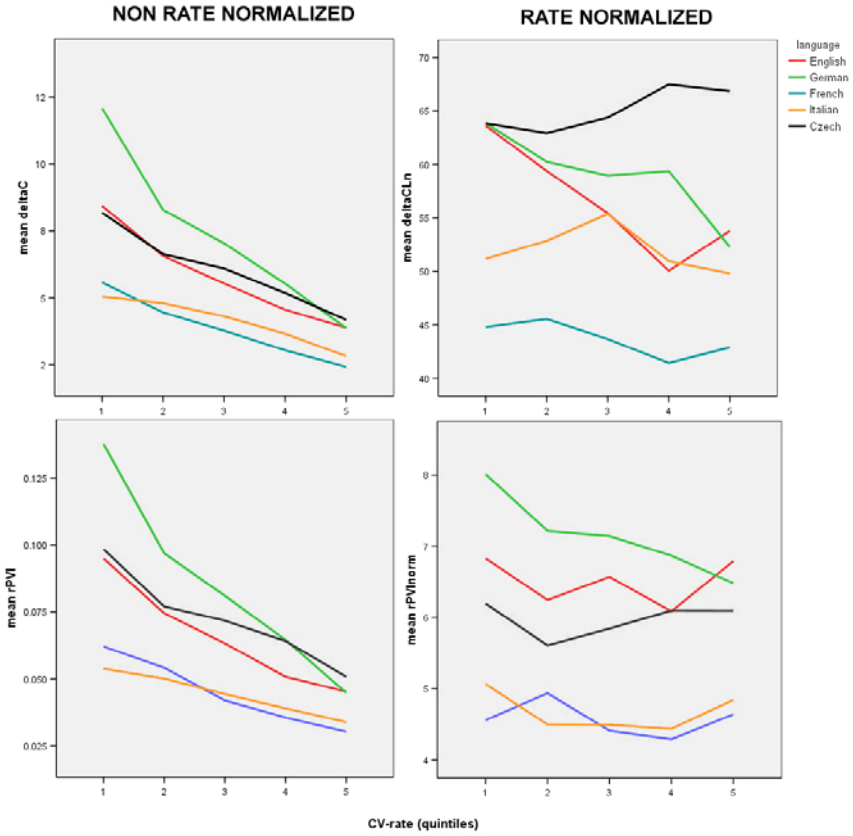

Figure 4: ΔC (top) and rPVI (bottom) before (left) and after (right) normalization. Graphs plot mean values for each quintile of CV-rate connected by a line for each of the five languages under investigation (Czech, English, French, German, Italian)

Both graphs on the left in Figure 4 reveal the effects of speech rated on the non normalized measures as it can be seen that both rhythm measures drop in each language with increasing CV-rate quintile. This should of course be expected given the analysis in section 3. However, upon visual inspection the results in both graphs look extremely similar and it is important to notice that at all CV-rate categories the rhythm class theory is well supported: the two syllable-timed languages, Italian and French, are both very similar and clearly lower than stress-timed English and German. Unclassified Czech is more similar to traditional stress-timed languages, English and German. This is true at all individual CV-rate quintiles. However, when looking at the results across the rate quintiles both graphs imply that the rhythm of slow French and Italian (1st quintile) is similar to medium speed English and Czech (3rd quintile) which again is similar to fast German (5th quintile). Such conclusions are likely to be incorrect because these similarities are a result of CV-rate variability in the data and not C interval durational variability.

After normalization (Figure 4, left graphs) the picture changes drastically and not many similarities between nPVInorm and ΔCln remain. In both cases it can be observed that speech rate normalization had an effect in that speech rate influences are either less systematic or not present any more. However, in particular in the case of ΔCln, strong differences can be observed between languages. Also, upon visual inspection the rhythm class distinction is still supported in the case of rPVInorm (French and Italian are both much lower than English and German) but not any more in the case of ΔCln where languages like Italian and English (syllable-timed and stress-timed) have about the same mean values for the 3rd and 4th quintile (which is probably the most common rate in each language). At the 5th quintile (fastest CV-rates) the languages German, English and Italian group together, suggesting that they share rhythmical features. Czech is drastically higher and French drastically lower than these languages, suggesting that they are rhythmically very different. At the 1st quintile Czech, English, and German are grouping up (again suggesting similarity in rhythm). Because of their different nature after normalization both measures, ΔCln and rPVInorm, will be discussed separately in the following.

rPVInorm: The effect of speech rate in the five different languages after the normalization has been tested with five ANOVAs (one for each language). Results show 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 rPVInorm was very effective. Speech rate differences only remain in one langauge, German, and there only between rather extreme rates. These differences, however, can probably be neglected since the extreme rate ranges in BonnTempo are unlikely to occur regularly in real speech situations (Dellwo *et al.*, 2004).

Additional five ANOVAs tested rhythm class differences at each of the five quintiles with rPVI as dependent variable and rhythm class as a 2 class factor. Czech was excluded from this analysis as its rhythmic categorization has been disputed (Dancovicova & Dellwo, 2007); English and German were attributed to the stress-timed, French and Italian to the syllable-timed group. The analysis showed that at each quintile highly significant differences were obtainable between rhythm classes ($p<0.001$). It can therefore be

concluded that the rate normalized rPVInorm is a very robust variability measure supporting the rhythm class hypothesis across a range of extreme speech rates in all languages.

ΔCln: Rate normalization in the case of ΔC creates a very different picture. Languages seem to be very unequally affected by the rate normalization and a rhythm class distinction is not very well obtainable any more. ANOVAs for testing within-language variation (quintile as a five class factor and language as the dependent variable) help interpreting the situation and show that the visual impression may be a slightly misleading: 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 5th quintile is significantly different from all other groups. Given these results, it can be concluded that rate normalization is probably not as ineffective as the visual impression of the data suggests (apart from a few exceptions in English and German). However, the support for rhythm classes in this measure is less strong. More research needs to be performed to ΔCln to produce a clearer picture.

5.3 Discussion and conclusion

It can be concluded that the rate normalization procedures applied to rPVI and ΔC lead to a more robust picture of rate. However, because of these traces rhythmic class separation of ΔC is somehow distorted after normalization. For this reason rPVI is considered the more robust measure for rhythmic class separation when rate variability is present in the data.

Given the assumption above (5) that the complexity of C intervals in stress-timed languages may favor variability of C interval durations across rates (because complex intervals are likely to be reduced in complexity at higher rates) we found support here only with ΔCln (ANOVAS for within language variability across rates revealed significant differences for English and German) but not with rPVInorm. So the raised concerns in Grabe & Low (2002) against normalizing rPVI because between language syllable complexity differences might interact with speakers' speech rate (see Introduction) seem unjustified in the light of the present results.

6. OVERALL SUMMARY AND CONCLUSIONS

In this paper it was demonstrated that the widely used rhythm measures ΔC and rPVI correlate with CV-rate variability and that %V and nPVI are unaffected by rate. It was demonstrated that there are effective ways to normalize the rate affected measures ΔC and rPVI. For the rhythm measures' power to separate languages of different rhythm classes on an acoustic level it was demonstrated that the normalized rPVI shows more consistent results than the normalized ΔC.

The main conclusions of this paper are thus that both consonantal rhythm measures, ΔC and rPVI, need to be normalized for rate when C interval variability is intended to be measured. If the aim of the rhythm analysis is to separate languages of different rhythm classes from each other then rPVInorm is probably the best choice when a rate normalized consonantal measure needs to be chosen.

7. REFERENCES

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

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

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

Cahn, J. E. (1990), The generation of affect in synthesized speech, in *Journal of the American Voice I/O Society*,
(electronic Publication: https://eprints.kfupm.edu.sa/70011/1/70011.pdf)

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

Cutler, A. (1997), The syllable's role in the segmentation of stress languages, in *Language and Cognitive Processes*, 12, 839-845.

Cutler, A. & Norris, D. G. (1988), The role of strong syllables in segmentation for lexical access, in *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.

Dancovicova, J. & Dellwo, V. (2007), Czech speech rhythm and the rhythm class hypothesis, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1241-1244.

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

Dauer, R.M. (1987), Phonetic and phonological components of language rhythm, in *Proceedings of the 11th International Congress of Phonetic Sciences*, Talinn, Estonia, 447-450.

Dellwo, V. (forthcoming), *Influences of speech rate on acoustic correlates of speech rhythm: An experimental investigation based on acoustic and perceptual evidence*, PhD thesis, Bonn University, Germany.

Dellwo, V. (2006), Rhythm and Speech Rate: A Variation Coefficient for ΔC, in *Language and Language-processing* (P. Karnowski & I. Szigeti, editors), Frankfurt am Main: Peter Lang, 231-241.

Dellwo, V. (2008), The role of speech rate in perceiving speech rhythm, in *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 375-378.

Dellwo, V. & Wagner, P. (2003), Relations between Language Rhythm and Speech Rate, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 471–474.

Dellwo, V. & Koreman, J. (2008), How speaker idiosyncratic is acoustically measurable speech rhythm?, in *Electronic Proceedings of the annual meeting of the International Association of Forensic Phonetics and Acoustics (IAFPA)*, Lausanne, Switzerland.

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

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

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

Kim, J., Davis, C., and Cutler, A. (2008), Perceptual tests of rhythmic similarity: II. Syllable rhythm, *Language and speech*, 51, 343-359.

Low, E.L., Grabe, E. & Nolan, F. (2000), Quantitative characterization of speech rhythm: Syllable-timing in Singapore English, *Language and Speech*, 43, 377-401.

Mok, P. & Dellwo, V. (2008), Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English, in *Proceedings of Speech Prosody*, Campinas, Brazil, 423-426.

Nazzi, T., Bertocini, J. & 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.

Pike, K.L. (1945), *Intonation of American English*, Ann Arbor: University of Michigan Press.

Ramus, F. (2002), Language discrimination by newborns, *Annual Review of Language Acquisition*, 2, 85-115.

Ramus, F., Hauser, M.D., Miller, C., Morris, D. & Mehler, J. (2000), Language discrimination by human newborns and cotton-top tamarin monkeys, *Science*, 288, 349-351.

Ramus, F. & Mehler, J. (1999), Language identification based on suprasegmental cues: A study based on resynthesis, *Journal of the Acoustical Society of America*, 105(1), 512-521.

Ramus, F., Nespor, M. & Mehler, J. (1999), Correlates of linguistic rhythm in the speech signal, *Cognition*, 73, 265-292.

Roach, P. (1982), On the distinction between 'stress-timed' and 'syllable-timed' languages, in *Linguistic controversies* (D. Crystal, editor), London: Edward Arnold, 73-79.

Rincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F. & Mehler, J. (2005), The role of speech rhythm in languages discrimination: further tests with a non-human primate, *Developmental Science*, 8(1), 26-35.

Tilsen, S. & Johnson, K. (2008), Low-Frequency Fourier analysis of speech rhythm, *Journal of the Acoustical Society of America*, 124(2), (Online EL Publication).

Toro, J.M., Trobalon, J.B. & Sebastian-Galles, N. (2003), The use of prosodic cues in language discrimination tasks by rats, *Animal Cognition*, 6(2), 131-136.

Wagner, P. & Dellwo, V. (2004), Introducing YARD (Yet Another Rhythm Determination) and Re-Introducing Isochrony to Rhythm Research, in *Proceedings of Speech Prosody 2004*, Nara, Japan.

Wang, Q. (2008), L2 stress-perception: The reliance on different acoustic cues, in *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 635-638.

White, L. & Mattys, S. (2007), Calibrating rhythm: first language and second language studies, *Journal of Phonetics*, 35, 501-522.

White, L., Mattys, S., Series, L. & Gage, S. (2007), Rhythm metrics predict rhythmic discrimination, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1009-1012.

APPENDIX I: LIST OF FORMULAS FOR THE MEASUREMENTS OF SPEECH RATE AND RHYTHM

Equation 1: *Combined C and V interval rate*

$$CVrate = \frac{n_C + n_V}{\sum_{i=1}^{n_C} c_i + \sum_{i=1}^{n_V} v_i}$$

n = number of sampled intervals
C = C interval
V = V interval
c = C interval duration
v = V interval duration

Equation 2: *Percentage over which speech is vocalic (%V)*

$$\%\mathrm{V} = \frac{\left(\sum_{i=1}^{n_v} v_i\right) \cdot 100}{\sum_{i=1}^{n_c} c_i + \sum_{i=1}^{n_v} v_i}$$

n_v = total number of V-interval samples
n_c = number of C-interval samples
v = V interval duration
c = C interval duration

Equation 3: *Standard deviation of C intervals (ΔC)*

$$\Delta C = 100 \cdot \sqrt{\frac{n \cdot \sum_{i=1}^{n} C_i^{\,2} - \left(\sum_{i=1}^{n} C_i \right)^2}{n \cdot (n-1)}}$$

n = number of sampled intervals
C = duration of C interval

Equation 4: *Non-normalized consonantal Pairwise Variability Index*

$$\mathrm{rPVI} = \frac{\sum_{c=1}^{n-1} \left| x_c - x_{c+1} \right|}{n-1}$$

n = number of C-intervals sampled
c = C interval duration

Equation 5: *Normalized vocalic Pairwise variability index*

$$\mathrm{nPVI} = 100 \cdot \frac{\sum_{v=1}^{n-1} \left| \frac{x_v - x_{v+1}}{(x_v + x_{v+1})/2} \right|}{n-1}$$

n = number of V-intervals sampled
v = V interval duration

Equation 6: *Coefficient of variation (varcoC) of ΔC*

$$\mathrm{varcoC} = \frac{\Delta c \cdot 100}{mean_C}$$

c = C interval duration

Equation 7: *normalized rPVI*

$$\mathrm{rPVInorm} = 100 \cdot \frac{\sum_{C=1}^{n-1} \left| \frac{x_c - x_{c+1}}{(x_c + x_{c+1})/2} \right|}{n-1}$$

n = number of C-intervals sampled
c = C interval duration