The role of speech rate in perceiving speech rhythm

Volker Dellwo

Department of Phonetics and Linguistics
University College London, UK
v.dellwo@ucl.ac.uk

Abstract

Human listeners can distinguish between languages of different rhythmic classes (e.g. stress- and syllable-timed languages). The present study investigated the role of speech rate in this process. Acoustic data suggests (experiment I) that speech rate can distinguishes as reliable between stress- and syllable-timed languages as previously proposed correlates of speech rhythm (%V, VarcoC and nPVI). Behavioral data showed (experiment II) that listeners make use of rate differences when asked to assess rhythmic characteristics of stress- and syllable-timed languages in delexicalized speech. Results imply that speech rate is an important acoustic correlate for cross-language speech rhythm.

1. Introduction

Some languages can be classified into distinct rhythmic types of which the two most prominent are the stress-timed and syllable-timed rhythm classes ([17], [11]). Behavioral experiments have shown that adult human listeners ([16]), as well as newborns ([13], [14]), monkeys ([15], [18]), and rats ([19]) 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 widely been likened to a 'machine gun' sound, stress-timed languages to a 'Morse-code' signal. 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 argue that this regularity percept is mainly triggered by the variability of consonantal (C) and vocalic (V) intervals in connected speech ([17], [11]). In terms of acoustic measurements, [17] 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 ([17], [9]), 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; [11]) as is the rate-normalized standard deviation of V interval durations (VarcoV; [20], in press).

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'-'Morse code' metaphor was evoked by such characteristics. However, this metaphor contains a hitherto neglected detail. 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. Psychoacoustic research (see [10] 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 and V intervals) could contribute to how listeners perceive their durational regularity, hence how they perceive speech rhythm.

Do units of speech rhythm vary in speech rate between languages? So far, the rate of C and V intervals has not received much attention in speech rhythm research. In fact, all of the studies ([13], [17], [15], [14], [19], and [18]) that have found behavioral effects of speech rhythm in both humans and non-humans use stimuli selected from a corpus by [13] 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 and V interval durations being less variable in syllable-timed languages is that these languages typically have phonologically less complex syllable structures ([1], [5], [6], [4], [17], [11]). Subsequently it is possible that mean C and 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 (i.e. whether speech sounds more or less regular). The aim of the current study was (a) to analyze how speech rate varies naturally between stress- and syllable-timed languages (experiment I) and (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 (experiment II).

2. Experiment I: Acoustic measurements of speech rate between rhythm classes

In the present experiment the hypothesis was tested that languages traditionally classified as stress-timed (here: English and German) and syllable-timed (here: French and Italian) vary in speech rate in addition to measurable speech rhythm.

2.1. Method

Subjects: 7 English, 15 German, 5 French, and 3 Italian speakers took part in the experiment. All subjects were native speakers of the respective languages without speech pathologies. Subjects were paid for their participation.

Procedure: Speakers were recorded reading a short text (translated from German into each other language) in their native language with similar numbers of syllables (English: 77, German: 76, French: 93, Italian: 108) and an equal number of sentences (4 sentences consisting of 7 sub-clauses). Speakers in each language were highly consistent in producing each of the 7 sub-clauses as an intonation phrase (intonational unit typically proceeded and followed by a pause). The measurements described below were calculated for each intonation phrase. There were 217 samples summed over languages (49 [English: 7 speakers * 7 intonation phrases] + 105 [German: 15*7] + 42 [French: 6*7] + 21 [Italian: 3*7]). All speech material was labeled according to C and V interval durations ([17]) using Praat speech processing software ([3]).

Speech rhythm was measured with three parameters: (a) durational V:C ratio was measured by the percentage over which speech is vocalic (%V; [17]). (b) C interval variability was measured as the variation coefficient of the standard deviation of C intervals (VarcoC; [8], [20], [12]). ΔC has been demonstrated to be dependent on mean C interval durations ([2], [7]) which in return is dependent on speech rate (e.g. syllables/second). VarcoC is a derivative of ΔC which was developed to normalize for this variability. (c) V variability was measured with the speech rate normalized Pairwise Variability Index (nPVI; [11]).

Speech rate was measured in terms of C and V intervals per second (CV rate). C and V intervals were chosen for speech rate measurements over the more common unit of syllables because the syllable:CV interval ratio has been demonstrated to vary significantly between stress- and syllable-timed languages ([7]). In addition to the global CV rate, the rate of V intervals/second (V rate) and of C intervals/second (C rate) was calculated. This was done because varying syllable complexity between stress- and syllable-timed languages is commonly understood to reflect the number of consonants in a syllable but not necessarily the number of vowels ([1], [5], [4]). It is thus possible that V rates do not vary between stress- and syllable-timed languages. On the contrary: since stress-timed languages typically allow vowel reduction ([1], [5], [11], [17]) mean V intervals in these languages may even be shorter (thus of higher rate).

2.2. Results & Discussion

As displayed in Fig 1, the V:C ratio (measured in %V) is lower in stress- than in syllable-timed languages. Both C and V interval variability (represented by VarcoC and nPVI respectively) are higher in stress- than in syllable-timed languages. This was confirmed through an independent samples t-test with Bonferroni correction (%V: t[215]= -8.7; VarcoC: t[215]= 7.0; nPVI: t[215]= 3.8; p<0.001 for all comparisons). These results replicate findings by [17] and [11] which showed that speech rhythm classes can be distinguished along these dimensions.

How does speech vary between stress- and syllable-timed languages? Fig 1 displays that speech rate differs between languages of different rhythmic classes. The combined CV rate is higher in syllable- than in stress-timed languages. Equal patterns could be obtained for the individual C and V rates (not displayed). All differences were significant (independent samples t-test with Bonferroni correction; CV rate: t[215]= -7.5; C rate: t[215]= -5.9; V rate: t[215]= -8.7; p<0.001 always).

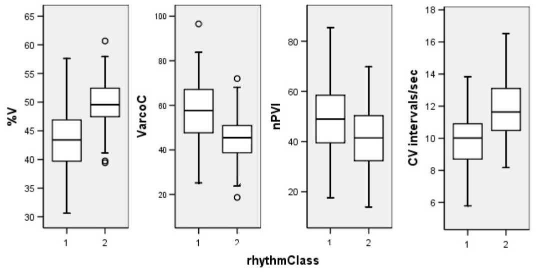

Figure 1: Box plots displaying the distribution of %V (top left), VarcoC (top centre), nPVI (top left), CV rate (bottom left), V rate (bottom centre) and C rate (bottom right) for stress-timed (1) and syllable-timed (2) languages respectively.

The important novel information of the present experiment is that C and V interval rates differ significantly between stress- and syllable-timed languages. The results therefore suggest that C and V rates may be as reliable indicators for rhythm class as rhythm measures capturing the variability (ΔC, nPVI) and ratios (%V) of C and V interval durations.

3. Experiment II: Perceptual measurements of durational C and V interval characteristics

Experiment I showed that both C and V interval rates are higher in syllable- than in stress-timed languages. Given this result, it is just as possible that listeners make use of C and V interval rate as cues to rhythm class judgments, in particular when linguistic content is meaningless to them (e.g. newborns, monkeys, or rats). This possibility was investigated in a perception experiment with adults who listened to stimuli from stress-timed French and syllable-timed German (German and French were chosen because they showed maximum differences in respect to all four parameters under investigation [%V, VarcoC, nPVI, CV rate]). In order to simulate a situation in which listeners were linguistically naive (similar to newborns, monkeys, or rats), language stimuli were delexicalized to contain durational C and V interval variability only and listeners were not told that the stimuli were derived from speech. Listeners were asked to rate the stimuli according to whether they sounded 'regular' or 'irregular'. The aim of the experiment was to test (a) whether listeners would rate German as less regular than French and (b) to investigate which of the durational characteristics of C and V intervals, %V, VarcoC, nPVI or CV rate, listeners make most use of for their rating of regularity.

3.1. Method

Subjects: 18 listeners without hearing problems took part in the experiment (10 English and 8 German, mean age: 29, range: 18-47 years).

Stimulus and Apparatus: 24 intonation phrases (12 German and 12 French representing stress- and syllable-timed languages respectively) from the speech material described above were selected according to the following criteria: In order to ensure that sufficient variability for each parameter was present, the intonation phrases from the highest and lowest quartile of the %V, VarcoC, nPVI, and CV rate spaces in each language were chosen (4 highest + 4 lowest = 8 stimuli). In addition 4 intonation phrases were chosen randomly from each language (8 controlled + 4 random = 12). Only intonation phrases that were preceded and followed by a clear pause were considered. This ensured that typical final phrase lengthening, which may have a large influence on the perception of speech rhythm, was included. The number of V intervals in each phrase varied between 7 and 12. All stimuli were evaluated by expert phonetic listeners and were considered natural examples of speech in each language.

Stimuli were delexicalized by turning V intervals into a complex periodic waveform of three harmonics (fundamental of 230 Hz + 2nd and 3rd harmonics) and C intervals into white noise. The noise intensity was 15 dB below tone intensity. This difference was chosen because it was found to reflect common intensity differences between vocalic and fricative content in real speech. Stimuli were presented to subjects via headphones using standard PCs.

Procedure: 3 repetitions of the 24 stimuli were put into random order for a total of 72. Subjects were asked to rate stimuli according to whether they sounded 'regular' or 'irregular' on an equally spaced 13 point scale which was labeled 'pretty irregular' (leftmost point), 'rather irregular' (scale point 5 from the left), 'rather regular' (point 9 from the left), 'pretty regular' (rightmost scale point). Subjects were instructed that ideal regularity means that the noise and tone intervals were of equal duration but that none of the stimuli would come close to this. To illustrate the concept of ideal regularity, subjects were played an example of a truly isochronous series of alternating tone and white noise intervals as an illustration.

Subjects controlled presentation with a 'play' button on the computer screen above the scale. Unlimited numbers of replays were possible; however, subjects were instructed to make their choices 'quickly and intuitively' (they typically listened to stimuli no more than two times).

To familiarize subjects with the full range of stimuli they heard 20 examples of the stimulus set before the start of the experiment. This test series included the 8 most extreme trials in each language (16 trials) plus 4 randomly chosen trials (2 from each language).

3.2. Results and Discussion

The box plot in Fig 2a displays the distribution of the listener ratings for each language. It can be seen that stimuli derived from German were rated as less regular than stimuli from French. The effect is highly significant (independent samples t-test, language * rating: t[22]= -5.3, p<0.001). This means that C and V interval durational information typical for stress-timed German is perceived as less regular than in syllable-timed French. This is in accordance with results from similar experiments in which listeners distinguished between languages of different rhythmic classes given durational C and V cues only ([16], [17], [18], [19]). In these studies, however, more speech like stimuli were used (e.g. *sasasa* delexicalization; [17]) and, in case of human adults, subjects were typically told that they judged languages they did not know. The fact that subjects in the present experiment were not aware that the stimuli derived from speech (but unconsciously distinguished between languages on the basis of their regularity rating) is further support for the assumption that rhythmical duration differences between languages is a type of information not exclusive to the language domain ([15], [18], [19]).

Which acoustic cues did listeners base their regularity decision on? Fig 2b displays listener ratings plotted as a function of %V, VarcoC, nPVI and CV rate. Linear regression showed that listener ratings could not be predicted by %V (R^2=0.006, p=0.72), nor by nPVI (R^2=0.13, p=0.084) and only poorly by VarcoC (R^2=0.26, p=0.006). However, CV rate predicted listener ratings of regularity well (R^2=0.655, p<0.001). This

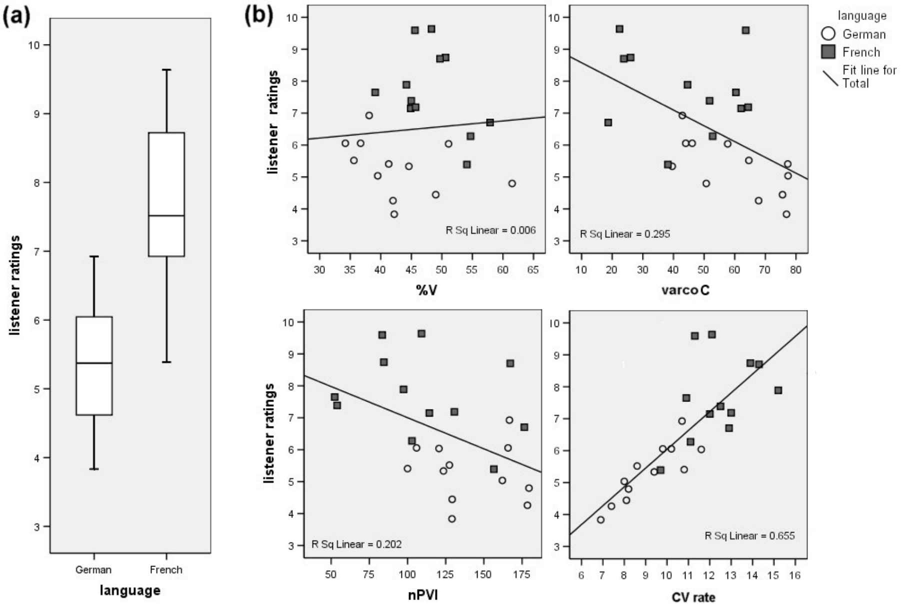

Figure 2: (a) Box-plot displaying the distribution of listener ratings of regularity as a function of language (German and French); (b) Cross plots of listener ratings as a function of %V (top left), VarcoC (top right), nPVI (bottom left) and CV rate (bottom right) with superimposed lines of best fit for a linear correlation and R^2 for a linear correlation.

suggests that listeners weighted interval rate highly when deciding whether an alternating tone/noise sequence was more or less regular.

Since the number of tone sequences (thus the total duration) varied considerably across the stimuli (see Method) this could have had an effect on the results (it may be that short stimuli for some reason tended to be faster than longer ones, etc.). However, listener ratings were not related significantly to the number of tones in a stimulus (linear regression: $R^2 = 0.011$, p=0.63).

4. Summary and conclusions

On an acoustic level the present research identified CV rate as an additional rhythm class distinction factor alongside widely used rhythm measures ΔC (or here the rate normalized version VarcoC), %V, and nPVI. In the behavioral domain it was shown that the perception of regularity in stimuli reflecting C and V interval variability of stress-timed German and syllable-timed French is highly influenced by CV rate differences between these languages.

Given these results it is possible that the percept of 'regularity' and 'irregularity' in syllable- and stress-timed languages respectively is to some degree caused by rate differences between these languages. In the present experiment adult listeners were treated as linguistically naive by presenting them non-speech stimuli and a non-speech task. It would be interesting to see whether the results can be replicated by listeners which are linguistically naive by nature (e.g. human newborns, monkeys, or rats). If such listeners would make use of rate information in order to distinguish between stimuli from different rhythmic classes they would be likely to do that in real life environments (e.g. newborns in a bilingual environment; [17]) when rates are bound to vary across rhythm classes. It would further be interesting to see whether rate information plays a role in discriminating speech rhythm when listeners are not linguistically naive (e.g. phonetic experts). If rate should be a cue to the perception of variability in real speech then it is also possible that within language rate variability would contribute to speech from the same language to appear more or less syllable- or stress-timed.

5. Acknowledgements

This project was partly supported by the German Research Council (DFG). The author wishes to thank Patti Adank, Paul Iverson, Stuart Rosen, Bronwen Evans, Yi Xu, and Andy Faulkner for helpful comments and discussions of the draft version, Petra Wagner and Wolfgang Hess for collaborations with the data collection for experiment I,

6. References

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

[2] 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 ICPhS, Barcelona, Spain, 2693–2696.

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

[4] Dankovičová, J. and Dellwo, V. (2007). "Czech speech rhythm and the rhythm class hypothesis," Proceedings of the 16th ICPhS, Saarbruecken, 2007.

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

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

[7] Dellwo, V. (submitted). "Influences of speech rate on acoustic correlates of speech rhythm: An experimental investigation based on acoustic and perceptual evidence." PhD thesis, Bonn University, Germany.

[8] Dellwo, V. (2006). "Rhythm and Speech Rate: A Variation Coefficient for ΔC," Pawel Karnowski & Imre Szigeti (eds.) Language and Language-processing. Frankfurt am Main: Peter Lang, 231-241.

[9] Dellwo, V., and Wagner, P. (2003). "Relations between Language Rhythm and Speech Rate," D. Recasens, M.J.Solé and J. Romero (eds.), Proceedings of the 15th ICPhS, Barcelona, Spain, 471–474.

[10] Friberg, A. and Sundberg, J. (1995). "Time discrimination in a monotonic, isochronous sequence". J. Acoust. Soc. Am. 98 (5), 2524-2531.

[11] 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.

[12] Lee, C. S., and McAngus Todd, N. P. (2004). "Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora," Cognition 93, 225-254.

[13] Nazzi, T., Bertoncini, J., and Mehler, J. (1998). "Language discrimination by newborns: Toward an understanding of the role of rhythm," Experimental Psychology 24(3), 756-766.

[14] Ramus, F. (2002). "Language discrimination by newborns," Annual Review of Language Acquisition 2, 85-115.

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

[16] Ramus, F., and Mehler, J. (1999). "Language identification based on suprasegmental cues: A study based on resynthesis," J. Acoust. Soc. Am. 105(1), 512-521.

[17] Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," Cognition 73, 265-292.

[18] 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.

[19] 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.

[20] White, L. and Mattys, S. (in press) "Calibrating rhythm. First language and second language studies," J. Phonetics.