Relations between language rhythm and speech rate

Volker Dellwo, Petra Wagner
Institut für Kommunikationsforschung und Phonetik, Bonn
E-mail: vde@ikp.uni-bonn.de, pwa@ikp.uni-bonn.de

ABSTRACT

The interaction between speech rate and rhythm is a topic that has hardly been studied in respective models of language rhythm though its potential significance has recently been addressed: since both of these prosodic parameters are to a great extent dependent on speech timing they are suspected to interact to a great degree. The present research studies the influence of speech rate on the vocalic and intervocalic measures %V and ΔC that have been discussed widely in the recent past as measures by which language rhythm classes may be distinguished. Results show that ΔC is extremely speech rate dependent while %V remains rather stabile across different speech rates. Different results may be obtained according to the way speech rate is controlled across languages. An alternative model for a cross language rate control will be proposed in this context.

1 INTRODUCTION

Acoustic correlates in the speech signal that support the well known rhythmical distinction between stress-timed languages (StLgs) and syllable-timed languages (SyLgs) [cf. 1, 4] have very often been searched for; usually with little success. Most recently Ramus et al. [6] proposed such an acoustic correlate that is based on the percentage of vocalic intervals (%V) as well as the standard deviation of consonantal intervals (ΔC) in the speech signal. According to these two dimensions StLgs and SyLgs cluster around different areas (henceforth: the cluster hypothesis).

To what extent %V and ΔC really represent language rhythm is currently a question of heavy debate [cf. 2, 3]. Grabe and Low [3], for example, who tried to replicate the findings of Ramus et al. [6] for their data, come to significantly different results that do not support the cluster hypothesis. Thus they conclude that the measure proposed by Ramus et al. [6] is not reliable in respect to distinguishing language rhythm classes.

In a reply to Grabe and Low [4], Ramus [5] suspects that amongst other factors (e.g. speaker typical influences) the non-existence of a control method for speech rate may have led to the different results in [4]. In contrast to [4] who normalise speech rate in their data with a Pairwise Variability Index, Ramus et al. [6] control speech rate by averaging sentence duration (3 sec.) and the number of syllables per sentence (15 to 19). Ramus [5] suspects that %V and ΔC may be affected by speech rate in great degree and thus he argues that "[t]he usefulness of variables such as […] ΔC may well be limited to corpora where speech rate is strictly controlled" (p. 117). He further suspects that ΔC and %V may by affected by speech rate in different proportions, while these proportions may vary across different languages.

Our research makes an attempt to study the proportions of variance for %V and ΔC in relation to speech rate within and across languages. If the different results of Grabe and Low [4] are really based on the fact that they have not controlled speech rate across languages, we would expect that in a controlled experiment with different languages, where different speech rates are simulated, the clusters obtained by Ramus et al. [6] should be visible for between language versions of comparable syllable rate and they should break up when syllable rate varies significantly.

2 EXPERIMENT

Our speech production experiment involves the manipulation of speakers' (Ss) reading tempo by encouraging them to read a small text at different speeds.

Speakers: 16 Ss took part in the experiment, 5 native speakers of English (E) (2 from the Edinburgh region of Scotland and 3 from Mid-West America), 4 native speakers of French (F) from the south-western area of France and 7 native speakers of German (G) from the mid-western area of Germany. Mean Ss age is 28.4 years (SD = 5.6). Ss took part in the experiment voluntarily and were paid a small expense allowance. None of the Ss reported any sort of language impairment, nor could this be detected in any of the Ss during the course of the experiment.

Experimental material: A German text from a novel by B. Schlink (*Selbs Betrug*, 1994, p. 242) of 76 syllables in 3 sentences (4 main and 3 sub-clauses) was used as reading material for the current experiment. The text was translated by philologically educated native speakers into English (76 syllables, 3 sentences: 4 main and three sub-clauses) and French (93 syllables, 4 sentences: 4 main and 4 sub-clauses).

Reading instructions & recording procedure: All Ss were recorded in a sound proof booth at the Institute for Communication Research and Phonetics (University of Bonn). Recordings were carried out with a condenser microphone directly on PC. Ss were given the text in their native language and were asked to familiarise with it by reading it aloud several times. After familiarisation Ss were recorded performing the task to read the text in a way they

considered 'normal reading'. After that Ss were recorded twice, the first time being instructed to read the text 'slowly' and the second time to read the text 'even slower'. In a third step Ss were recorded under the instruction to read the text 'fast' and were consecutively encouraged to read the text 'faster' until they considered themselves having reached a maximum reading speed or until reading performance became so poor that recordings were terminated. Acceleration steps varied according to S from 3 to 8.

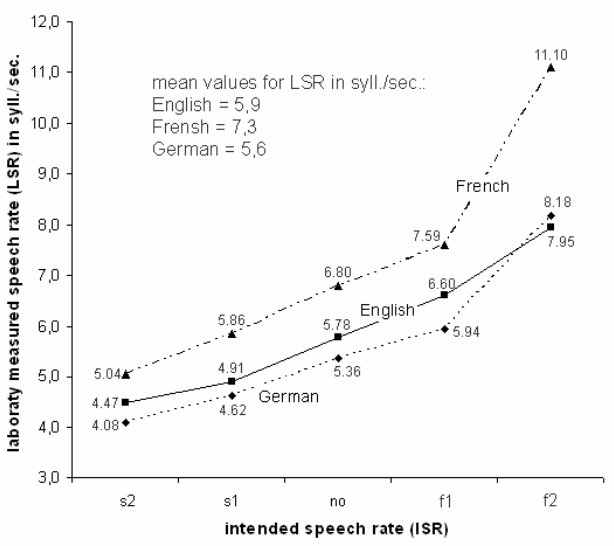

Diagram 1: Laboratory measured speech rate (LSR; exact values next to each respective entry) in syllables/second vs. intended speech rate (ISR).

Labelling procedure: In order to receive values for the durations of vocalic and consonantal intervals as well as syllable durations the version recorded at 'normal' reading speed (*no* or *the norm*) the first deceleration attempt (s1; s = slow), the second deceleration attempt (s2) as well as the first acceleration attempt (f1; f = fast) and the last acceleration attempt (f2) were labelled using both auditory and visual cues according to the criteria set up in [6]. Labelling was performed by both authors. Half-automatic label correction software programmed by the first author as well as a final control procedure was carried out to minimize individual influences of the labellers on the labelling process.

3 RESULTS

Two types of speech rate will be distinguished in the following: Intended speech rate (ISR) which refers to the reading speed that Ss intend to reach according to the experimental instructions (see above) and laboratory measured speech rate (LSR) which refers to the number of syllables that Ss produce per second (syl/sec). ISR is henceforth represented by the five labelled versions for each S (see above). LSR, %V, ΔC have been calculated for all ISR versions (note that for presentation reasons all values for ΔC have been multiplied by 100 in the following thus absolute ΔC figures will be in centi-seconds (csec)).

3.1 SPEECH RATE

Values for LSR (diagram 1) show a strong positive correlation with ISR for each language which means that Ss intention to speed up or slow down their reading speed is realized by a respective change in syllable rate. According to mean values of LSR in diagram 1 (superimposed) F reaches the highest value. G and E are rather equally below this value with E slightly above G. This pattern is also valid for all ISR versions from s2 to f1. The fact that the connection lines between languages from s2 to f1 run nearly parallel indicates that the proportional changes in speech rate between these ISR versions are rather equal for each language. Between s1 and s2 this parallel pattern breaks up, thus there is a difference between the proportions to which Ss of a language are able to increase syllable rate. Ratios calculated for no:s2 of each language show this proportion in detail: E = 1: 1.38; F = 1: 1.63; G = 1:1.52. This means that Ss of F are most able to increase their syllable rate from the norm while speakers of E are least able to do so. Ss of G lie well in the middle between these two extremes.

3.2 %V AND ΔC IN RELATION TO ISR

According to diagram 2 values for StLgs E and G cluster around an area in the upper left part of the diagram and values for SyLg F cluster in an area in the lower right of the diagram. Since %V and ΔC vary in complex fashion they will be treated separately in the following. To make within language variation of %V and ΔC according to speech rate comparable across different languages, ratios s2:s1:no:f1:f2 have been calculated (cf. table 1 and 2) with the norm being set to 1.

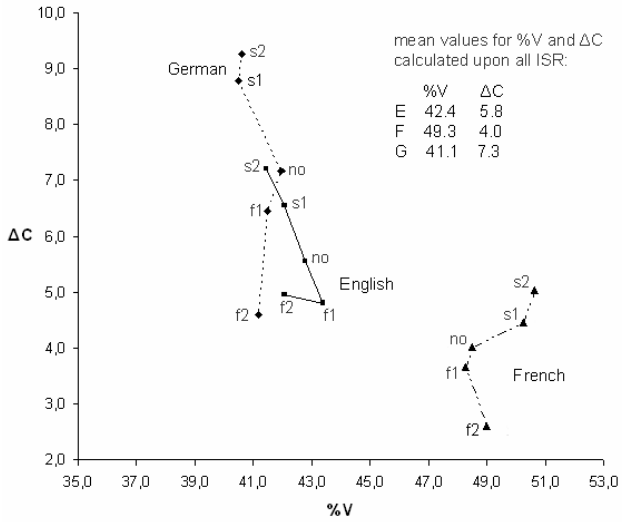

Diagram 2: %V (in %) vs. ΔC (in csec) at five ISR versions.

%V: Ratios for %V (table 1) show that values for this measure are rather stabile across all ISR versions for all

languages. For E and G the deceleration and acceleration attempts seem to correlate with a slightly lower %V (apart from E at f1). For F this tendency seems to be reversed (apart from f1). Absolute mean values for %V are higher for F (5.7 %-points) and E (2.5 %-points) than the ones obtained by Ramus et al. [6]. Range values for E (1.3 %-points) and for F (2.3 %-points) reveal that within language variation is sometimes higher for our results than it is in Ramus et al. [6] for between language variation (e.g. 1.5 %-points between Dutch and Spanish), though this is not true for the languages under investigation, E and F, where difference in Ramus et al. [6] are 3.5 %-points.

group	value	s2	s1	no	f1	f2
G	%V	40.6	40.5	42.0	41.5	41.2
	ratio	0.97	0.96	1	0.99	0.98
E	%V	41.5	42.1	42.8	43.4	42.1
	ratio	0.97	0.98	1	1.01	0.98
F	%V	50.6	50.5	48.5	48.3	49.0
	ratio	1.04	1.04	1	1.00	1.01

Table 2: %V in % for G, E and F at all different ISR versions with respective ratios.

ΔC: Ratios for ΔC values show a negative correlation with ISR (cf. table 2) apart from E at f2 where there is a small relative increase in ΔC compared to f1. This means that ΔC in the fastest ISR version of E does not vary much in relation to the first acceleration step. Absolute ΔC mean values are slightly lower for F (- 0.43 csec) and higher for E (+ 0.47 csec) than mean values for these languages in Ramus et al. [6].

group	value	s2	s1	no	f1	f2
G	ΔC	9.28	8.77	7.17	6.46	4.61
	ratio	1.29	1.22	1	0.90	0.64
E	ΔC	7.20	6.57	5.57	4.81	4,95
	ratio	1.29	1.18	1	0.86	0.89
F	ΔC	5.04	4.47	4.01	3.68	2.61
	ratio	1.26	1.12	1	0.92	0.65

Table 2: ΔC in csec for E, F, and G at all different ISR versions with respective ratios.

4 DISCUSSION

4.1 SPEECH RATE

According to the relations between ISR and LSR we interpret our findings as revealing language characteristics on the one hand and speaker universals on the other:

Characteristic for each language is the number of syllables that speakers are able to produce per second on an average basis. This value will be highly influenced by the language individual phonetic, phonologic and phonotactic syllable structures. Characteristic for languages in our data is also the way in which they allow their Ss to increase the syllable rate. In this respect F seems to provide the greatest freedom, Ss of E seem to be most restricted in syllable rate increase while G lies somewhere in the middle between these two extremes.

The fact that LSR values change in proportionally the same way for each language from s2 to f1 may reveal a speaker universal feature. The finding leads us to assume that Ss of all languages under investigation in the current experiment have a notion of a normal speech tempo in their language and a common notion of what it means to speak slowly (s1), slower (s2) and faster (f1) than the norm when they are requested to do so (cf. reading instructions above).

4.2 %V and ΔC

The obtained values for %V and ΔC are on a general basis well in accordance with the ones obtained by Ramus et al. since values for E are in the upper left corner of the diagram while F lies further below to the right of E. Since G, which has traditionally been categorised as a StLg, clusters with E in our data, we additionally found a supporting example for the hypothesis that the vocalic and intervocalic measures %V and ΔC do distinguish linguistic rhythm classes. Absolute ΔC mean values for E and F agree well with the ones obtained by Ramus et al. [6]. The fact that our %V values for F and E are higher than in Ramus et al. [6] and that within language variation in respect to speech rate is sometimes higher for our values than it is for some cases of between language variation in Ramus et al. [6], may be related to differences in our experimental material[1] and will not be regarded as evidence that could undermine the general pattern.

So our results do support the findings in Ramus et al. [6] and more importantly: our findings do support Ramus et al. [6] at all speech rates. Even if we compared ISR versions for F with versions of E or G with different LSR we would find that the clusters are not mixed up. Thus we can conclude that speech rate seems to have an influence on the values proposed by Ramus et al. [6] (especially on ΔC) but it may not be so strong that it could undermine the cluster hypothesis according to which StLgs and StLgs cluster around different areas (cf. introduction).

According to ΔC two values in our results for E and F match almost exactly with values obtained by Ramus. It is ΔC for E at the norm (here: 5.57; Ramus: 5.39) and for F at s1 (here: 4.47, Ramus: 4.39). As it has been pointed out in the introduction, LSR in Ramus et al. [6] has been matched across languages by choosing sentences of roughly 3 second durations consisting of 15 to 19 syllables (n = 17) which results in an LSR of 5.67 syl/sec (17 syllables /3 seconds). This value again matches almost exactly with the LSR value obtained by us where ΔC matches, i.e. E at the norm (5.78 syl/sec) and F at s1 (5.86 syl/sec) (cf. diagram 1). In other words, according to our data, Ramus et al. [6] compared a syllable rate of English speakers that would be considered as normal by its Ss with a syllable rate of French that its Ss would consider as being slow speech.

[1] Sentence variability is for example controlled in our material while Ss uttered different sentences in Ramus et al.

4.3 AN ALTERNATIVE SPEECH RATE CONTROL

Since our data and discussion on speech rate has revealed that Ss may have a notion about a normal, slow or fast speech tempo in their language and since we showed that in this respect Ramus based his analysis for E and F on two unequal ISR versions for these languages, we want to introduce a model in which speech tempo is not controlled on the basis of LSR but on ISR. Since LSR and ΔC for E at the norm and F at s1 match so well with results for these languages in Ramus et al. [6] we want to hypothesize (regarding our data) what would have happened if Ramus et al. [6] had controlled speech rate on the basis of ISR[2]: The main change in the results in Ramus et al. [6] had been that ΔC in the case of F had even been lower (4.01 csec) thus it had moved even clearer away from the cluster of StLgs. If this was the case for all other SyLgs then the separation of the clusters would possibly have become even clearer, provided that the LSR of 5.67 syl/sec matches a slower ISR for the other SyLgs as well. Provisional data that we analyzed for one Italian speaker supports this view. For StLgs a tendency of a clearer separation of the clusters is visible in the case of G. If we controlled syllable rate for G with 5.67 syl/sec we would choose ΔC rather at the 'fast' G version (f1) where syllable rate is 5.94 syl/sec and ΔC 6.46 csec. A tempo control on the basis of matching the ISR norm version would lead to a ΔC of 7.17 csec which is a move into a direction away from the area where SyLgs cluster.

Regarding these findings, the areas where StLgs and SyLgs cluster along the two dimensions %V and ΔC may be differentiable more clearly on the basis of an ISR control across languages. But of course, there are more than three languages to study. The tendency may well prove to be incorrect once more languages will be added to the model. The authors are currently working on enlarging their data in this respect.

4.4 PERSPECTIVES

A model of speech rate control based on matching ISR across languages of course requires that we have an extensive knowledge of the ISR characteristics for all the languages under investigation. But apart from the fact that it is questionable that a text of approximately 80 syllables read at different tempo versions by 4 to 7 speakers may represent ISR norms for a given language, another theoretical issue about Δ C arises: The fact that ΔC is speech rate dependent may be given the following rational explanation: Since consonantal intervals will be longer on an average basis in slow speech and shorter in fast speech the standard deviation of consonantal intervals, i.e. ΔC, will vary proportionally to this as well. In other words: when we compare F and G or E at the norm ISR version, consonantal intervals in F will be shorter (since syllables are shorter) than in G or E, which should have an influence on the fact that ΔC in F is lower at this ISR than in G or E. So ΔC has to be made comparable across different syllable rates in relation to the mean value of consonantal intervals for each language. A proposal that the authors are currently working on is to calculate a variation coefficient (varco) for ΔC that we define as the percentage of ΔC of the mean value for consonantal intervals (varkoΔC = (ΔC * 100)/ meanC). Provisional results for varkoΔC calculated upon our data seem to reveal new interesting results. While varkoΔC stays constant across all syllable rates for F it varies strongly in complex fashion for G and E. We are currently working on interpretations for these findings.

[2] We do not consider %V at this point since we found this value to be rather constant across ISR versions.

5 CONCLUSION

Results from the current experiments show tendencies rather than trends but still the tendencies support the hypothesis that StLgs and SyLgs are distinguishable by %V and ΔC. More data will show, whether the proposed model of speech rate control will be a stabile support for the cluster hypothesis in the future.

ACKNOWLEDGEMENTS

We wish to thank Judith Adrien and Stacy Dellwo for the translation of the experimental material into their native languages (French and English), Eva-Maria Orth and Margaret Thompson for helpful comments on the draft version, as well as all the speakers who participated in the experiment.

REFERENCES

[1] D. Abercrombie, *Elements of general phonetics*, Chicago: Aldine, 1967.

[2] F. Cummins, "Speech rhythm and rhythmic taxonomy", in *Proceedings of speech prosody 2002*, B. Bel and I. Marlin, Eds., Aix-en-Provence: Laboratoire Parole et Langage, pp. 121-136, 2002.

[3] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis", *Papers in laboratory phonology*, vol. 7, pp. 515-546, 2002.

[4] K. L. Pike, *The intonation of American English*. Michigan: University Press, 1945.

[5] F. Ramus, "Acoustic correlates of linguistic rhythm: Perspectives", *Proceedings of speech prosody 2002*, B. Bel and I. Marlin, Eds., Aix-en-Provence: Laboratoire Parole et Langage, pp. 115-120, 2002.

[6] F. Ramus, M. Nespor, J. Mehler, "Correlates of linguistic rhythm in the speech signal", *Cognition*, vol. 73, pp. 265-292, 1999.