ARTICULATION RATE VARIATION WITHIN THE INTONATION PHRASE IN CZECH AND ENGLISH

Jana Dankovičová
University College London, UK

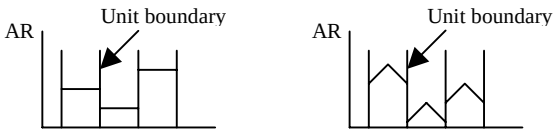
ABSTRACT

A cross-linguistic study was conducted on Czech and English reading to test the hypothesis that the intonation phrase (IP) is the domain of a systematic pattern in articulation rate variation, namely, of a slowing down throughout the IP (rallentando). The articulation rate was measured across (phonological) words. The statistical analysis showed a significant tendency towards the rallentando pattern in both Czech and English. Moreover, in both languages the word's articulation rate was also significantly affected by its size in syllables, with the rate increasing with the number of syllables. Word class was a significant factor as well, with phonological words containing only function words being faster than those containing only content words. In both languages word size seems to be the strongest factor, position within the IP somewhat weaker and word class the weakest factor.

1. INTRODUCTION

Articulation rate (AR) is a prosodic feature, defined as a measure of rate of speaking in which all pauses are excluded from the calculation [1]. Although there is a general agreement that AR is not constant but varies throughout speech, our knowledge in this area is still very limited. This may partly be due to the lack of a theory of AR and, arising from that, the uncertainty with respect to methodology for investigation. Yet it is believed that introducing the parameter of variation in AR would contribute to improving both the naturalness of the signal in speech synthesis and the accuracy of automatic speech recognition performance.

The issue of AR variation was addressed specifically in the author's previous experiments on Czech [2, 3], where the question of the domain of AR variation was seen as central. In the first experiment [1] two basic hypothetical scenarios were examined. Either AR would be constant within the unit and any change in rate would be restricted only to the unit boundaries (scenario represented schematically in (a) below), or articulation may follow a certain pattern within the unit, a pattern whose shape would be recurrent in all units. In (b) a pattern of an increase followed by decrease is shown as a possibility, although the pattern could have any shape.



Three candidate domains were examined – the syntactic clause, intonation phrase and the interpause stretch. AR within these units was measured across phonological words. The analysis demonstrated a considerable rate variation within these units but, in the case of the intonation phrase (IP), also a regular patterning. The pattern consisted of a slowing down throughout the phrase. It was termed 'rallentando' and formally defined as tendency towards an overall decrease in AR within the phrase.

In the second experiment [2] the rallentando pattern within the IP was tested on a much larger database, including four speakers and three tasks – reading a story, paraphrasing the story and spontaneous speech. The rallentando pattern was confirmed both across the speakers and across the three tasks.

The position of phonological words within the IP is obviously not the only factor affecting their AR. Among those mentioned in the literature are (i) the number of syllables and phones in the word [4, 5], (ii) whether the word is a function or content word, e.g., [5], (iii) syntactic position of the word [6] and (iv) whether the word represents a new or given information [5, 7]. The number of syllables in the phonological word was found a significant factor also in the two experiments described above. Moreover, the distinction between function and content words was tested in [2] and found significant as well.

This paper represents a comparative study of AR variation in Czech and English reading. In the case of Czech, the data were extracted from the database analyzed in [2] so that comparisons between the two languages would be possible. In the case of English, the experiment is a pilot study testing whether the rallentando pattern applies also to English and it is planned to lead in the future to a larger study involving also paraphrasing and spontaneous speech. Apart from position of the word within the IP, the number of syllables it contained and its word class category were considered as possible determinants of its AR.

2. EXPERIMENTAL DESIGN

2.1. Material

2.1.1. Czech. Four native Czech speakers participated in the study, two male and two female, all with university education, aged between 23 and 34. None of them knew the exact purpose of the experiment. The whole recording session involved three tasks: reading a story aloud, paraphrasing the story and elicited spontaneous speech. Reading, the task reported on in the present paper, included about 7.5 minutes in total.

2.1.2. English. A male native southern British English speaker, age 29, with university education, participated in the experiment. The tasks were the same as for Czech speakers: reading a story, paraphrasing the story and elicited spontaneous speech. Only reading has been analyzed so far; it consists of 6 minutes of speech.

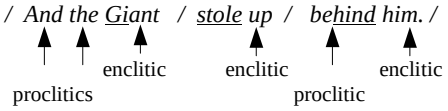
2.2. Methodology

2.2.1. The intonation phrase. The intonation phrase is defined as a stretch of speech determined by the intonation contour. It is a well-defined unit in the Czech phonetic literature, e.g., [8, 9], and with respect to English, it corresponds to the tone unit in British intonational analyses [10] and the intonation phrase in the Beckman and Pierrehumbert's work [11].

2.2.2. The unit for AR measurement. Articulation rate variation can be measured across various units (for discussion see [3]). Following the methodology used in [1] and [2], in the present study the phonological word was chosen. It is defined as a string of syllables containing a single stress and respecting lexical word boundaries (for English, e.g., [11]; for Czech [9]). It can thus be considered a rhythmical unit.

In Czech the stress falls regularly on the first syllable and a phonological word usually coincides with a lexical word. Sometimes, however, it includes one or more enclitics (for examples, see [1]) and in the phrase-initial position it may contain also proclitics.

In English, stress placement varies and thus when stress falls on other than the first syllable of the lexical word, the phonological word contains one or more proclitics, as well as possible enclitics. The following example illustrates this (slash stands for the phonological word boundaries, stressed syllables are underlined):



The analysis showed that, in fact, even in English the majority of phonological words were headed by stressed syllables and thus, when proclitics occurred, they were mostly found in the phrase-initial position.

The reasons for selecting the phonological word are related to the question about at what level of speech processing articulation rate is planned and for what unit it is planned. My hypothesis is that articulation rate, as a prosodic feature, is likely to be planned at phonological or phonetic levels of speech production. The phonological word has been recognized by some theories of speech production [13] as an important unit at the phonological level of the processing of connected speech. Moreover, in the Czech phonetic literature, e.g., [9], it is traditionally considered to be the smallest rhythmical unit and rhythm and articulation rate are related concepts. There is also some evidence for the word as a unit of tempo in the sense that accentuation seems to affect the duration of a single word rather than all words within a larger domain (e.g., constituents); moreover, the word as a whole (i.e., all segments and syllables) is affected [14]. Finally, the phonological word is a small enough unit to make it possible to observe the variation in articulation rate in reasonable detail.

2.2.3. Measurement technique. IPs and phonological words were identified impressionistically. To avoid idiosyncrasy, the identification was performed independently by two phoneticians for Czech - the author and another Czech phonetician, and for English by the author and two British phoneticians. The agreement was very high both for intonation phrases and phonological words.

Speech samples were digitized at a resolution of 16 bits and a sampling rate of 16 kHz. In order to make measurements of the duration of phonological words as accurate as possible, a combination of impressionistic listening and reference to waveform displays and wideband spectrograms was used.

Articulation rate was expressed in terms of the number of syllables per second.

2.2.4. Determinants of articulation rate variation. Three possible determinants were considered in the study: the position of the phonological word within the IP, its size in syllables and its word class. The word class relates to content-function word dichotomy. As phonological words in some cases included more than one lexical word (when they contained clitics), it was impossible to employ a binary category for the word class. Therefore, three categories were used for phonological words: (i) those containing one or more content words (category 'C'); (ii) those containing a content word and one or more (cliticizing) function words (category 'C+F'); (iii) those containing function word(s) only (category 'F').

2.2.5. Statistical analysis. Statistical processing involved analysis of variance on AR of phonological words (henceforth 'words'). Generalized Linear Modelling (GLM): General Factorial was used, in combination with a post-hoc Tukey test, which shows what pairs of means were significantly different ($p < 0.05$ was used). The factors were the word position within the phrase, its size, word class and, for Czech also speaker. Only IPs containing 2, 3 or 4 words were examined, since IPs with 5 or more words were very rare both in English and Czech. The coding of word position within the phrase made it necessary to perform the analysis for these groups of phrases separately. Adjusted R^2 provided information about the success of the model (the proportion of variance in articulation accounted for) and η^2, a measure of the strength of association (ratio of the sum of squares for a factor and the total sum of squares, [15]) gave a rough estimate of the relative strength of individual factors.

3. RESULTS

3.1. Czech

3.1.1. Inter-speaker correlation of AR of phonological words. Having available recordings of the same text read by different speakers made it possible to carry out cross-speaker comparisons of the AR of individual words. Obviously, only those IPs could be used in the analysis in which the location of both phrase boundaries and word boundaries was the same across speakers. The analysis showed that the speakers coincided to a considerable extent in this respect and so 81% of the total number of words could be used (i.e., 150 words per speaker).

Table 1 shows the results in the form of the matrix of Pearson coefficients of correlation between the four speakers (AV, LG, OU and PK). The degree of correlation between the subjects is strikingly high (and significant at 0.1% level). It indicates strongly that the AR of words is not arbitrary. It follows that the patterns of rate variation cannot be arbitrary either.

Speakers	AV	LG	OU
LG	0.915		
OU	0.917	0.903	
PK	0.902	0.905	0.876

Table 1. Inter-speaker correlations of the AR of phonological words in reading.

3.1.2. Statistical results. A total number of 563 words were used in the analysis (182 in 2-word IPs, 273 in 3-word IPs and 108 in 4-word IPs). The total number of IPs examined was 209 (91 in 2-word IPs, 91 in 3-word IPs and 27 in 4-word IPs).

Table 1 shows that all main factors, apart from 'speaker' in 4-word IPs, were highly significant. Significance of the 2-way interaction 'position'*'word class' in 3-word and 4-word IPs suggested that there may be differences in the AR patterns across the categories of word class. A similar interpretation applies for 'size' in the interaction 'size'*'word class'. However, further investigation suggested that the differences may be spurious due to the unbalanced number of observations within categories.

2-wd IPs	*df*	*F*	*p*	*Adj. R²*	η^2
Model	14,167	35.9	<0.001	73%	
Position	1,167	41.9	<0.001		0.063
Wd size	4,167	36.4	<0.001		0.218
Wd class	2,167	4.8	=0.009		0.014
Speaker	3,167	4.6	=0.004		0.021
Size*Wd class	4,167	9.9	<0.001		
3-wd IPs					
Model	15,257	27.2	<0.001	59%	
Position	2,257	35.0	<0.001		0.105
Wd size	4,257	14.7	<0.001		0.088
Wd class	2,257	18.1	<0.001		0.054
Speaker	3,257	5.0	=0.002		0.023
Pos*Wd class	4,257	2.8	=0.025		
4-wd IPs					
Model	11,96	19.9	<0.001	66%	
Position	3,96	10.4	<0.001		0.099
Wd size	3,96	20.0	<0.001		0.191
Wd class	2,96	5.7	=0.005		0.034
Pos*Wd class	3,96	5.5	=0.002		

Table 1. Results for the GLM analysis for 2-word, 3-word and 4-word IPs (Czech).

The models for 2-word, 3-word and 4-word IPs were reasonably successful, accounting for 73%, 59% and 66% of variance in AR respectively. The strongest factor was generally 'size', followed by 'position'; the weakest was 'word class' and 'speaker'.

Figure 1 shows boxplots for AR in different word positions within the IP, separately for 2-word, 3-word and 4-word IPs. In 2-word IPs the first word was significantly faster than the second word. For 3-word IPs the pattern of slowing down throughout the IP was also found; however, the Tukey test showed that the AR means in positions 1 and 2 were not significantly different. In 4-word IPs, although the mean in position 3 was higher than in position 2, the overall pattern still falls into the category of rallentando. Moreover, the means in positions 2 and 3 were not significantly different. All the other positions were significantly different from each other.

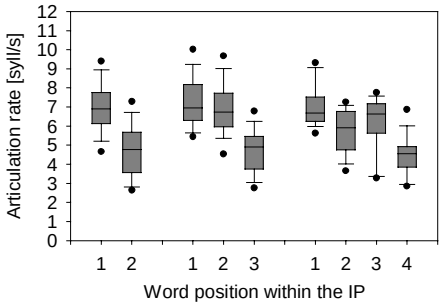


Figure 1. Word AR in different positions within the IP (Czech).

With respect to the word size, the AR of a word increased with the number of syllables it contained (cf. stress-timing). While in 2-word IPs all the categories (1-, 2-, 3- and 4-syllable words) were significantly different from each other and in 4-words all but 3- and 4-syllable words, in 3-word IPs only monosyllables were significantly different from all the other categories.

The examination of word class showed that in all three groups of IPs phonological words classified as category 'C' tended to be significantly slower than category 'C+F' and these, in turn, were significantly slower than phonological words in category 'F'.

Although the overall results for the factor of speaker were significant, it turned out that only one subject was responsible for this by speaking significantly faster than the other subjects.

3.2. English

3.2.1. Statistical results. For English 406 words were measured in total (150 in 2-word IPs, 180 in 3-word IPs and 76 in 4-word IPs) and a total of 154 IPs were examined (75 in 2-word IPs, 60 in 3-word IPs and 19 in 4-word IPs).

Table 2 and Figure 2 summarize the results of the statistical analysis for English. Like in Czech, 'position' was a highly significant factor in all three groups of IPs, showing a slowing down throughout the IP. In 3-word IPs the AR means in all positions were significantly different from each other. In 4-word IPs the mean AR in position 2 was slightly higher than in position 1 but the difference was not significant; all the other positions differed significantly from each other.

The effect of the number of syllables in the word was also highly significant in all three groups of IPs, the tendency being the same as in Czech – AR of the word increasing non-linearly with increasing number of syllables. In 2-word and 3-word IPs monosyllables and disyllables were significantly different from the words containing 3, 4, 5 and 6 syllables but the latter categories were not significantly different from each other. In 4-word IPs, monosyllables were significantly slower than the words with a higher number of syllables but disyllables were significantly different only from trisyllables.

The results for word class were inconsistent across the three groups of IPs. Firstly, it was a significant factor only in 2-word and 3-word IPs. Secondly, while 'C' category was always the slowest, in 3-word IPs it differed significantly from the fastest 'F' but not from 'C+F'. Thirdly, in 4-word IPs it was 'C+F' which was the fastest category, followed by 'F' and all the

categories differed significantly from each other.

Size of words was the major determinant of their AR, followed by their position within the IP. The word class proved the weakest factor. Neither group of IPs showed any significant interactions.

2-wd IPs	*df*	*F*	*p*	*Adj. R²*	η^2
Model	11,138	50.6	<0.001	78.6%	
Position	1,138	50.3	<0.001		0.072
Wd size	8,138	16.0	<0.001		0.184
Wd class	2,138	6.7	=0.002		0.019
3-wd IPs					
Model	11,168	42.3	<0.001	71.7%	
Position	2,168	23.9	<0.001		0.075
Wd size	7,168	13.4	<0.001		0.148
Wd class	2,168	4.2	=0.017		0.014
4-wd IPs					
Model	7,68	19.0	<0.001	62.7%	
Position	3,68	10.8	<0.001		0.066
Wd size	4,68	10.1	<0.001		0.170

Table 2. Results for the GLM analysis for 2-word, 3-word and 4-word IPs (English).

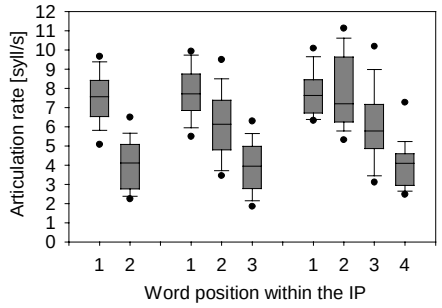


Figure 2. Word AR in different positions within the IP (English).

4. DISCUSSION

The study presented here demonstrates that AR does not vary in an arbitrary way but its variation is structured. It is structured both in terms of having a domain – the intonation phrase, and in terms of a recurrent pattern of slowing down throughout the phrase – rallentando. Moreover, the study suggests that this pattern applies cross-linguistically, at least for reading. As mentioned above, it was found previously also for Czech semi-spontaneous and spontaneous speech [3]. A larger scale study is planned to find out whether rallentando applies across different tasks also in English.

The question of why AR varies in this particular way is open to speculation. However, it is very likely that phrase-final lengthening is a partial contributor, as the final word within the IP was always significantly slower than all the non-final words.

Apart from their position within the IP, the AR of words was also significantly affected by their size in syllables and their word class. In both languages a tendency was found for the word AR to increase with increasing number of syllables. The tendency towards isochrony in Czech provides further evidence against the traditional definition of stress-timing (English) and syllable-timing (Czech). With respect to word class, in both languages phonological words containing only content words tended to be slower than those containing function words. This is in accord with findings reported in other studies, e.g., [5].

The finding of a regular patterning of AR suggests that the IP, in addition to its role in planning intonation contours, is further involved in the generation of temporal structure, in particular, in structuring AR variation. An interesting parallel emerges between rallentando in AR and declination in the sense that both involve a progressive decline, and both seem to have the IP as their domain. A question arises whether they stem from similar causes, and, related to that, whether they might be planned at the same stage in speech production. Further research is needed to address these issues.

ACKNOWLEDGMENTS

I would like to thank Francis Nolan and Jill House for their valuable comments and suggestions and also for participating in identification of IPs and phonological words in the English material. For the help with Czech material my thanks go to Zdena Palková.

REFERENCES

[1] Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in spontaneous speech.* London and New York: Academic Press.

[2] Dankovičová, J. 1997. The domain of articulation rate variation in Czech. *Journal of Phonetics*, 25, 287-312.

[3] Dankovičová, J. 1998. *The linguistic basis of articulation rate variation in Czech.* Unpublished Ph.D. thesis. Oxford University.

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

[5] O'Shaughnessy, D. 1995. Timing patterns in fluent and disfluent spontaneous speech. In *Proceedings of the International Conference on Acoustics, Speech and Signal processing*, 600-603. Detroit, Michigan, USA.

[6] Carlson, R. 1991. Duration models in use. In *Proceedings of 12th International Congress of Phonetic Sciences*, vol. 1, 243-246. Aix-en-Provence.

[7] Fowler, C. A. and Housum, J. 1987. Talkers' signalling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, 489-504.

[8] Daneš, F. 1957. *Intonace a věta ve spisovné češtině.* Praha.

[9] Palková, Z. 1994. *Fonetika a fonologie češtiny*. Praha: Karolinum.

[10] Crystal, D. 1969. *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

[11] Beckman, M. and Pierrehumbert, J. 1986. Intonational structure in English and Japanese. *Phonology Yearbook*, 3, 255-310.

[12] Grosjean, F. and Gee, J. P. 1987. Prosodic structure and spoken word recognition. In Frauenfelder, U. H. and Tyler, L. K. (eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.

[13] Levelt, W. J. M. 1989. *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

[14] Eefting, W. 1991. The effect of 'information value' and 'accentuation' on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America*, 89 (1), 412-424.

[15] Linton, M. and Gallo, P. S. 1975. *The practical statistician: Simplified handbook of statistics*. Monterey, CA: Brooks/Cole.