

## ON THE ANALYSIS OF SYLLABLE TIMING IN EVERYDAY SPEECH

Henrietta J. Cedergren and Hélène Perreault

Département de Linguistique

Université du Québec à Montréal

C.P. 8888, Montréal H3C 3P8, Québec, Canada

### ABSTRACT

A regression analysis is presented which investigates the effects of surface prosodic structure features and speech tempo on syllable timing in Montreal French. A cluster analysis of the 16 speakers' regression coefficient estimates allows us to distinguish between patterns of effects that are systematic across speakers, and patterns of effects that are speaker specific.

### INTRODUCTION

Explicating temporal organization in speech is a complex task which involves extricating the effects of multiple parameters. Previous experimental research has shown that observed patterns of timing may follow from properties of the linguistic text, i.e. segmental and prosodic organization features [1, 2], properties of the informational content, i.e. the flow of information in discourse [3, 4], or properties of the context, i.e. citation vs spontaneous speech forms [5]. Experimental paradigms have thus allowed researchers to investigate specific properties while presumably controlling for other dimensions of variability and have provided insights on the relevant parameters involved in the prediction of timing. However the relation between models of timing in experimental contexts and the timing of natural speech remains to be clarified. Little is known, for example, about the relations between observed properties of timing and inter-speaker variation characteristics, i.e. socio-symbolic differences among speakers [6].

This study is concerned with

addressing the issue of understanding the relation between temporal organisation and inter-speaker differences in speaking style in everyday speech. It bears on the problem of modelling timing in everyday speech. It attempts to distinguish between patterns of effects that are systematic across speakers, and patterns of effects that are speaker specific.

### METHODS

The database for this study consists of three minutes excerpts of running speech extracted from hour-long recordings of sociolinguistics interviews of Montreal French. A sample of sixteen speakers differentiated according to sex, age and social class was used in the analysis. Speaker ages correspond to two generational categories: twenty to twenty five years of age or fifty five and over; an equal number of working class and middle class speakers were selected.

Table 1. Distribution of speakers according to sex, age and social class.

	working class		middle class	
	M	W	M	W
young	2	7	25	43
	23	50	113	70
old	28	45	75	31
	37	107	81	61

The speech excerpts were sampled at a rate of 16kHz/s. Segment durations were measured by manually placed cursors on a spectrogram time-aligned with a waveform; segment durations and labels

were stored in an automatically generated file which was also coded for perceived prominence and prosodic grouping. Three levels of prosodic organisation were identified: phonetic syllables, rhythmic groups and intonational phrases. Prominence was distinguished as either demarcative, associated with the right boundary syllable of rhythmic groups or intonational phrases, or non-demarcative, secondary prominence, associated with non-final syllables.

### A duration model

As reported elsewhere [7] we have been concerned with modelling the relation between observed syllable durations in milliseconds and features of surface prosodic structure and syllable composition. In the present analysis, the factorial set codes for eight infrasyllabic, suprasyllabic and tempo effects. At the level of the syllable, onset and rhyme complexity are coded in terms of the number of segments. Suprasyllabic prosodic features distinguish positional effects within the intonational phrase and the rhythm group, and prominence. Tempo is coded as a local measure of articulation rate within the intonational phrase [7]. Finally, we include a multiplicative factor which accounts for the interaction of infrasyllabic complexity in the rhyme with intonational phrase final position.

Table 2. Factorial categories.

Onset complexity in number of segments	OC
Rhyme complexity in number of segments	RC
Intonational phrase penult position	IPP
Intonational phrase final position	IPF
Non-demarcative prominence	NDP
Rhythmic group final position	RGF
Intonational phrase tempo	IPT

### Interspeaker differences and similarities

We address the issue of modelling interspeaker differences and similarities based on the SAS complete linkage cluster analysis method. The regression coefficient estimates derived for each speaker's syllable duration model were Z-score normalized. These served as input to the cluster procedure.

### RESULTS

The linear regression analyses revealed differences in temporal effects of the factorial scheme previously described. Speakers in the sample differ both in how the factors jointly account for observed syllable duration in their speech ranging from 48.36% for Spkr 25 to 67.43% for Spkr 75, and in how the proposed effects achieve statistical significance ranging from a maximum of eight factors to a minimum of 5 core factors. Thus, OC, RC and IPT emerged as significant predictors of syllable timing for all speakers. While IPF and IPF\*RC assume different speaker specific patterns. Both predictors are significant for 6 speakers. IPF\*RC was significant for 8 speakers, while IPF was not. The reverse pattern was obtained for 1 speaker and finally neither were significant for 1 speaker.

The following equation illustrates how the factors account for 57.38% of the variance of one of the speakers, Spkr 2, a young working class male:

$$\text{Syllable duration} = \text{constant} + 51.67\text{OC} + 46.36\text{RC} + 99.05\text{IPT} + 27.56\text{IPF} + 8.20\text{RC*IPF} + 21.06\text{IPP} + 12.35\text{NDP} + 7.49\text{RGF}$$

Cluster analysis further investigated the issue of interspeaker temporal differences and similarities using as input data each speaker's coefficient estimates. The results of the complete linkage clustering procedure [8] revealed that the sample of sixteen speakers could be

divided into four groups based on their linguistic similarities: group 1 consisting of 6 speakers (25, 61, 50, 113, 2, 7), group 2 consisting of 5 speakers (23, 28, 45, 37, 70), group 3 consisting of 4 speakers (31, 75, 81, 107) and finally group 4 restricted to 1 speaker (43). The hierarchical grouping of speakers that was obtained by measuring the maximum Euclidean distance among the clusters is illustrated in Figure 1. A comparison between the grouping illustrated in Figure 1 and the pre-analytic grouping of speakers displayed in Table 1 suggests that syllable timing behaviour is not a straightforward mirror of socio-demographic grouping. However certain observation can be made about the predominant socio-demographic characteristics of the first three clusters. Two of the three groups appear to be predominantly defined by age. Five of the six speakers in group 1 are young; all of the four speakers in group 3 are older. Social class appears to be associated with speaker differentiation in groups 2 and 3. Four of the five speakers in group 2 are working class; three of the four speakers in group 3 are middle class.

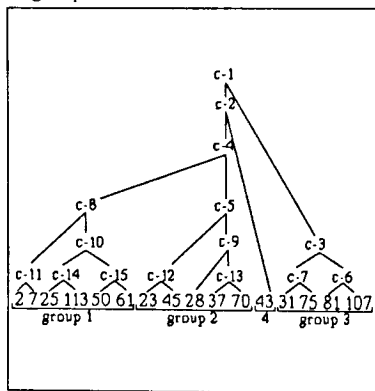


Figure 1 Hierarchical grouping of speakers.

Although the clustering procedure indicates that speakers can be allocated to four groups based on their temporal behaviour, it falls short in addressing

further questions such as: How do the particular prosodic and tempo effects contribute to characterize each group? We attempted to explore this issue by means of the least squares comparison of means option of the GLM procedure in SAS and thus evaluate the null hypothesis that the difference in means among the standardised coefficient values for each factor for each pairwise comparison of the four clusters is not significant. Table 3 summarizes the results of the tests.

It appears that intergroup discrimination is differentially associated with the prosodic and tempo parameters. At one level, these tests reveal that certain temporal characteristics are shared by all speakers in the sample and could be construed as a shared dialect feature. Thus the null hypothesis is not rejected in all pairwise comparisons of the OC, RGF and RC\*IPF effects. At the opposite extreme, these tests reveal that the most consistent indicator of group membership appears to be the local measure of tempo effect (IPT) which significantly distinguishes all pairwise comparisons of groups of speakers except groups 2 and 4. The other parameters have intermediate discriminatory effects. Some appear to be group specific; the IPP effect discriminates group 4 from all other groups of speakers. Others are more limited in the specificity of their discrimination among the groups of speakers. Thus group 1 is distinguished from groups 3 and 2 by NPD; IPF distinguishes group 2 from groups 1 and 4, and RC distinguishes group 3 from groups 1 and 2.

#### DISCUSSION

We have presented some preliminary results of an investigation which aims to understand the relation between temporal organization in everyday speech and interspeaker differences in speaking style. A factorial scheme of eight prosodic and

Table 3. Results of least squares means comparisons of speaker groups by prosodic parameters. (\* = significant comparison)

	OC	RC	IPP	IPF	NDP	RGF	IPT	IPF*RC
1 vs 2					*		*	
1 vs 3		*			*		*	
1 vs 4			*				*	
2 vs 3		*		*			*	
2 vs 4			*	*				
3 vs 4			*				*	

tempo parameters was found to account for the observed variance in measured syllable duration ranging from 48.36% to 67.43%.

The use of cluster analysis has allowed us to examine the issue of how speakers are grouped based on their timing behaviour. A post-hoc inspection of the socio-demographic characteristics of the speakers in each cluster group revealed that both age and social class appear to be related to the grouping. These results suggest that differences in syllable timing in spontaneous speech are determined not only by linguistic properties, but that they may also reflect inter-speaker socio-demographic differences.

Exploratory analysis of each cluster using a least squares means comparison revealed that prosodic and tempo parameters do not operate uniformly in discriminating among groups of speakers.

Although these results are based on the analysis of a small sample of sixteen speakers, we are confident that they provide substantive evidence that temporal organization is an important component of speaking style.

#### ACKNOWLEDGEMENT

This research was supported by the Social Sciences and Humanities Research Council of Canada under Grant No. 410-92-1840.

#### REFERENCES

- [1] Lehiste, I. (1975), The phonetic structure of paragraphs, *Structure and Process in Speech Perception*, ed. by A. Cohen & S. Nooteboom, Springer, Berlin. pp. 195-206.
- [2] Campbell, W.N. (1992), *Multi-level Timing in Speech*, Unpublished PhD Thesis, Sussex University, Department of Experimental Psychology.
- [3] Bruce, G. & P. Touati (1992), On the analysis of prosody in spontaneous speech with exemplifications from Swedish and French, *Speech Communication* 11, pp. 453-458.
- [4] Van Santen, J.P.H. (1992), Contextual Effects on Vowel Duration, *Speech Communication*, February 1992.
- [5] Lindblom, B. (1990), Explaining phonetic variation: A sketch of the H & H theory, in *Speech Production and Speech Modelling*, ed. by W. Hardcastle and A. Marchal, Kluwer Academic Publishers, Dordrecht, pp. 403-439.
- [6] Labov, W. (1994), *Principles of linguistic change: internal factors*. Blackwell, Oxford.
- [7] Cedergren, H. J. & H. Perreault (1994), Speech Rate and Syllable Timing in Spontaneous Speech, *Proceedings of the International Conference on Spoken Language Processing*, vol.3, pp. 1087-90.
- [8] Anderberg, M. R. (1973), *Cluster Analysis for Applications*, New York: Academic Press inc.