Differential lengthening of syllabic constituents in French: the effect of accent type and speaking style.

Daniel Hirst, Corine Astésano, Albert Di Cristo

CNRS Parole et Langage, Université de Provence, France

ABSTRACT

This paper presents results from the analysis of segmental duration in French in three different speaking styles, (a) 'reading'; (b) 'news broadcast' ; (c) 'spontaneous interview'. A ten minute corpus was hand segmented and labelled using six accent categories (i) unstressed (ii) word-initial (iii) emphatic word-initial (iv) word-final (v) intonation unit final before non-terminal boundary (vi) intonation unit final before terminal boundary.

Results show that when different accent types and different speaking styles are taken into account, there is no uniform lengthening of prosodic constituents in French, whether the syllable or some other higher-level unit is taken as the domain for lengthening. Instead differential lengthening is observed consistently with essentially greater lengthening towards the beginning of the syllable for initial prominence and greater lengthening towards the end of the syllable for final prominence.

The degree of the different types of lengthening was significantly different across speaking styles.

1. INTRODUCTION AND BACKGROUND

1.1 Segmental duration

It is well known that one of the major problems for the development of Text-to-Speech systems is that of supplying appropriate values for segmental duration in different contexts. The number of factors potentially influencing segmental duration is very large. Indeed it has been argued [10, 11] that the combinatorics of natural language is such that even an extremely large training corpus will only contain a limited proportion of the possible combinations of factors.

Campbell [3] argued that an efficient way of overcoming the sparcity of data problem was by first predicting the duration of a higher-level unit, the syllable, and then predicting the duration of the individual segments by an "accomodation principle" which in its strongest form states that each segment in a given syllable is stretched or squashed to the same degree. Campbell showed that when z-score normalised log durations were used, the accomodation principle gave quite good predictions of segmental durations. These predictions were even better when a 'backstep' procedure was added, allowing for bottom-up influence of the individual segments on the syllable's duration.

Later work [4; 5] showed that the accommodation principle in its strongest form, needed to be supplemented by differential lengthening of syllabic constituents: onset, peak and coda, in order to account for emphatic lengthening on the one hand, which affects primarily the early part of the syllable, and pre-boundary lengthening, which is more pronounced towards the end of the syllable.

For French, it has been argued [2; 9] that segmental duration is better modelled using a different higher-level prosodic unit than the syllable, the IPCG ("inter-perceptual centre group"), where the perceptual centre is assumed to coincide with the syllable onset after a pause but with the vowel onset elsewhere. Applying the accommodation principle at this level was found to give better results than the syllable in predicting the variability of the data analysed.

Fant et al. [8] had observed that post-vocalic consonants in French were lengthened only in pre-pausal positions. In our own work [1], preliminary results analysing two extracts of spontaneous speech showed that while the different constituents of the rime (nucleus and coda) seemed to behave consistently, there was a definite tendancy for differential lengthening of syllabic constituents with the onset being lengthened more with initial prominence and the rime being more influenced by final prominence, just as Campbell had found for English.

1.2 Rhythm and prominence in French.

Traditionally, French is described as having exceptionless fixed word-final stress. Quite considerable work in recent years (for an overview see [6, 7]) has brought to light a much more complex pattern of lengthening and pitch prominence, in particular in more spontaneous types of speech, with in particular word-initial prominences being much more frequent than was previously thought and not restricted to emphatic or focalised words. This suggests that the type of data analysed may be a very important factor and that models developped on the basis of laboratory speech may need serious revision for applications where more spontaneous styles of speech are required.

This study extends our earlier work by comparing the differential lengthening of syllabic constituents (onset, nucleus and coda) of six different accent types in six recordings, chosen to illustrate three different speaking styles.

2. MATERIAL AND PROCEDURE

2.1. Material

For our study, we analysed segmental durations in six recordings of connected speech (a total of approximately 10 minutes of speech), selected as fairly representative samples of three different speaking styles: reading, radio news broadcast and interview, each style being illustrated by two recordings.

Each recording involved a different speaker each of whom was a native speaker of educated standard French.

- The reading material consisted of a connected paragraph extracted from a story, and was read by two non-professional readers.
- The radio news broadcasts are coherent excerpts produced by professional journalists from one of the French national broadcast stations.
- Finally, the interview material involved two personalities discussing their work on a French national radio station. We selected a coherent passage of the interview for each speaker, consisting of the answer to a question asked by the journalist.

Our entire corpus is thus not controlled phonetically or semantically. The three speaking styles will henceforth be referred to as 'Reading', 'News' and 'Interview' respectively.

The choice of 3 different speaking styles was intended to bring to light any potential invariants of duration as well as the degree of variability between the different styles. Our hypothesis was that these three styles would illustrate three degrees along a continuum from the more careful prepared reading style to the completely impromptu interview style with the radio news broadcast being intermediate between the other two styles.

The recordings were transcribed without punctuation. An expert was asked to locate all perceived prominences, to mark emphatic accents and non-terminal and terminal Intonation Unit boundaries. The syllables were classified into 6 categories, (following [7]):

- word-final (WF)
- final in a non-terminal Intonation Unit (IU-NT)
- final in a terminal Intonation Unit (IU-T)
- word-initial (WI)
- emphatic word-initial (EWI)
- unstressed (US) (all remaining syllables)

The corpus contained a total of approximately 2600 syllables.

2.2. Experimental procedure

The excerpts were digitized at 16 kHz and phonemic and syllabic labels were aligned by hand with the speech signal. Each constituent of the syllable was then coded as Onset, Nucleus or Coda.

The duration of the different syllable constituents was measured and the raw data was normalised using the z-transform [3, 5] with the phonemic means and standard deviations calculated seperately for each speaker.

Our experimental design aimed to answer two questions :

1. does lengthening spread differently across syllabic constituents, according to accentual categories [4, 5, 8] ?
2. does each speaking style involve specific lengthening strategies ?

The results were computed through an analysis of variance (Anova), with the 3 speaking styles, syllabic constituents (ONC) and accentual categories as independent variables, and normalised duration as dependent variable. Further tests were carried out for each accentual category independently, in order to bring to light more explicitly specific distinctions or similarities of lengthening strategies across speaking styles.

It is evident that in a corpus of this type, occurrences of prominence are not homogeneously distributed among the categories, word-final prominence for example being far more frequent compared to terminal Intonation Unit final prominence or emphatic word initial prominence for example. The number of each type of syllabic constituents is also not equal, the Coda population being less numerous due to the specific nature of French syllable structure which largely favours open syllables (CV). The smaller number of Coda elements accounts for a larger Confidence Interval, as can be seen in the figures. For one combination of factors (Coda for emphatic word initial prominence in the reading style) there were less than 5 tokens (in fact only one occurrence). The value of this was ignored in the interpretation of the results. For the remaining 53 combinations 5 combinations were represented by less than ten values each. These combinations were not used in the interpretation of the results but were left in the figures since the corresponding values were consistent with the trends observed from the rest of the data.

3. RESULTS

The Anova shows that the lengthening of the three syllabic constituents onset, nucleus and coda are significantly different ($F(2;5750) = 11.186$ $p < 0.0001$). The accentual categories are also significantly different ($F(5;5750) = 160.587$, $p < 0.0001$), thus providing evidence for the relevance of the functional distinction between accent types. The interaction between syllabic constituents and accentual categories is also highly significant ($F(10;5750) = 18.228$, $p < 0.0001$), thus answering our first question: segmental lengthening does apply differently depending on the accent categories.

The speaking styles are not globally significantly distinct on the basis of lengthening ($F(2;5750) = 2.244$, $p = 0.1062$), showing that the three speaking styles have fairly similar overall strategies. The interaction between syllabic constituents and speaking styles is also non significantly different ($F(4;5750) = 1.638$, $p = 0.1618$). On the other hand, the interaction between speaking styles and accentual categories is highly significantly different ($F(10;5750) = 4.979$, $p < 0.0001$), indicating that the strategies vary across speaking styles when considering global lengthening throughout the different accent categories. The interaction between the 3 independent variables is just significant at the 5% level ($F(20;5750) = 1.598$, $p = 0.0442$), indicating that each speaking style tends to have specific differential lengthening strategies across the accent categories. This last remark answers our second question: there are differences of differential lengthening across speaking styles. More detailed results are presented below on the differences and similarities of the global and differential lengthening between the three speaking styles.

Figures 1, 2 and 3 show the distribution of lengthening throughout the syllabic constituents according to accent categories for each speaking style.

Two major results can be observed :

- word-initial accents, whether rhythmic (WI) or emphatic (EWI), show greater lengthening of the onset. This tendency is significant for the 3 speaking styles concerning the word initial prominence. As far as the emphatic word initial prominence is concerned, it is significant only for the Interview and News styles.
- word-final accents show greater lengthening of the nucleus and coda, especially at major prosodic boundaries (intonation units). However, there are some differences in speaking style strategies, as can be seen when comparing the three figures. Concerning the word-final accent (WF), the 3 speaking styles exhibit the same strategy, i.e. the nucleus is significantly longer than the coda, but not significantly longer than the onset, except in the 'News' style. As far as the syllables at the end of non-terminal intonation units (IU-NT) are concerned, the nucleus is significantly longer than the onset and coda for the three speaking styles. The major difference between the 3 speaking styles concerning differential lengthening lies at the end of terminal Intonation Units (IUT). The coda is significantly longer than the nucleus and onset for the 'Reading' style, whereas the lengthening concerns fairly equally the group nucleus and coda (i.e. the rime) for the 'News' style, the coda being significantly longer than the onset only. Finally, lengthening is homogeneously distributed over the syllabic constituents for the 'Interview' style.

In the following figures, O stands for Onset, N for Nucleus and C for Coda.

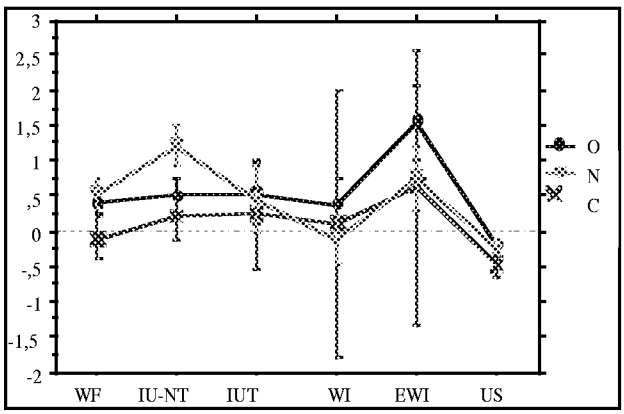

Figure 1 : Differential lengthening of syllabic constituents across accent categories for the 'Interview' style.

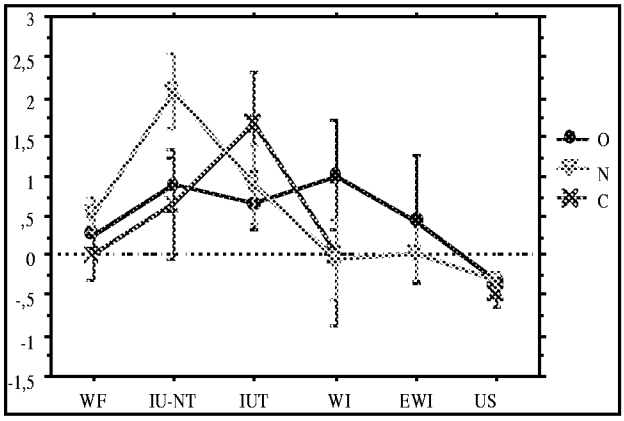

Figure 2 : Differential lengthening of syllabic constituents across accent categories for the 'Reading' style.

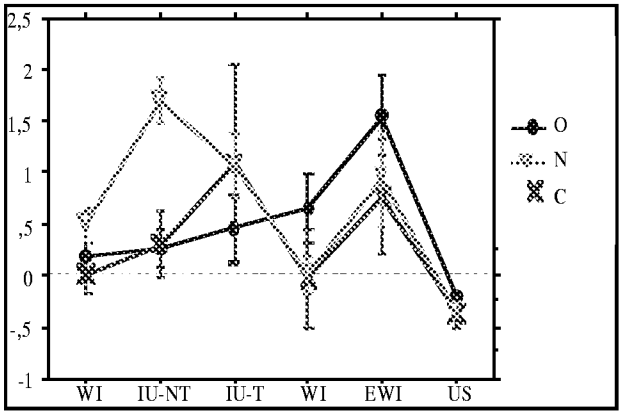

Figure 3 : Differential lengthening of syllabic constituents across accent categories for the 'News' style.

4. DISCUSSION

Our results suggest that the strong accommodation principle is not sufficient to account for segmental timing in French whether the syllable or the IPCG is taken as the relevant domain. Instead, earlier results by Fant et al. [8] , Campbell [4; 5] and Astésano et al. [1] were replicated, showing that, as in English or Swedish, differential lengthening of syllabic constituents in French needs to be taken into account. The distribution of accents thus seems to be an important factor in the three languages: indeed, our results confirmed that lengthening of the Onset is characteristic of word-initial accents, whereas lengthening of the Nucleus and Coda elements is more characteristic of word final accents, particularly before intonation boundaries.

This result suggests that the reason Barbosa and Bailly's IPCG model gave better results for French than Campbell's syllable based model stems from the nature of the corpus analysed which consisted of isolated sentences read in a very neutral style with little or no initial rhythmic or emphatic prominence. In this case only final lengthening was observed which as we have seen affects essentially the coda and the nucleus. The IPCG model would be clearly incapable of accounting for the lengthening of the syllable onset we observed in all three styles for both rhythmic and emphatic word-initial prominences.

One originality of our work lies in the comparison between different speaking styles, whereas the studies by Campbell and Fant et al. are primarily concerned with read speech. We were thus able to observe a variability in differential lengthening in French, which might also be found in English or Swedish.

The three speaking styles were significantly different on the nature of their respective differential lengthening strategies across accent categories. Indeed, the 'Reading' style is globally characterised by a greater lengthening of the coda, and optimally marks final prominence types. On the contrary, the 'Interview' style favours the lengthening of the onset element, marking initial accents more distinctively than final accents. The 'News' style, as we predicted, falls somewhere between the two extremes and seems to combine the two strategies, with a clear marking of both initial and final accent types.

Despite the fairly limited nature of the corpus, the effects we observed appear fairly robust across both speakers and speaking styles. Further investigation will however be necessary to see whether the three styles should really be considered as exhibiting distinct strategies or whether they should be considered as representing three points on a continuum stretching from the more prepared extreme to the more spontaneous. Work in progress extending the analysis to fully spontaneous dialogue material may be hoped to provide an answer to this.

5. REFERENCES

1. Astésano, C.; Di Cristo, A.; Hirst, D.J., «Discourse-based empirical evidence for a multi-class accent system in French», *Proc. of 13th ICPhS, Vol. 4, pp. 630-633, Stockholm*, 1995.
2. Barbosa, P.; Bailly, G. «Characterisation of rhythmic patterns for text-to-speech synthesis», *Speech Communication, Vol. 15 (1-2)* , 127-137, 1994.
3. Campbell, W.N. «Syllable-based segmental duration», in *Talking Machines: Theories, Models and Designs*, G.Bailly and C.Benoit (eds.), North-Holland, Amsterdam, 211-224, 1992.
4. Campbell, W.N. «Automatic detection of prosodic boundaries in speech. *Speech Communication* vol. 13, 343-354, 1993
5. Campbell, W.N. «Timing in Speech: a multi-level process», *in Prosody: Theory and Experiment, M. Horne (ed.), Dordrecht, Kluwer Academic Publishers*, in press.
6. Di Cristo, A. «Intonation in French», in *Intonation Systems: a Survey of Twenty Languages.* D.J.Hirst and A. Di Cristo (eds.) Cambridge University Press, Cambridge, 1998.
7. Di Cristo, A.; Hirst, D.J. «L'accent en français : stratégies et paramètres», *in J. Perrot (ed.) Polyphonie pour Ivan Fónagy*, 1997.
8. Fant, G.; Kruckenberg, A.; Nord, L. «Durational correlates of stress in Swedish, French and English», *Journal of Phonetics, Vol. 19*, 351-365, 1991.
9. Morlec, Y. *Génération multiparamétrique de la prosodie du français par apprentissage automatique.* Doctoral dissertation, INPG Grenoble, 1997.
10. van Santen, J.P.H. «Prosodic modelling in Text-to-Speech synthesis», *Proc. of Eurospeech-97*, 19-28, Rhodes, September 1997.
11. van Santen, J.P.H. «Segmental duration and speech timing», in *Computing Prosody: Computational Models for Processing Spontaneous Speech.* Y.Sagisaka, N.Campbell and N.Higuchi (eds.), Springer, New York, 225-249.