PROSODIC FEATURES OF FINALITY FOR INTONATION UNITS IN FRENCH DISCOURSE

Magali Vincent , Albert Di Cristo, Daniel Hirst
Institut de Phonétique, URA CNRS 261 "Parole & Langage"
Université de Provence, 13621 Aix en Provence, France
email : {phonetic; di cristo; hirst} @univ-aix.fr

ABSTRACT

This study aims to bring to light the prosodic features which contribute to the identification of finality in discourse. Perception tests applied to spontaneous speech extracted from radio interviews allowed us to derive two indices of finality. The main acoustic correlate was the F0 parameter. We found no evidence that silent pause or duration played an effective role in the perception of finality, however there was a significant drop in intensity on the final syllable.

INTRODUCTION

It is generally assumed that one of the main linguistic functions of prosody is a structural or organisational one, which consists in segmenting the flow of speech into coherent units of different sizes [1], [2].

At some high structural level, prosody serves to indicate a binary distinction between finished and unfinished utterances. At lower levels, prosody can also be used to signal boundaries of intermediate units such as prosodic phrases and prosodic words. In recent years, a large number of studies dealing with different languages have been devoted to investigating the correspondance between prosodic parameters and boundary type and/or magnitude. While the results of these studies are consistent with the fact that some parameters (mainly F0 and pauses), are used in similar ways across languages to signal different levels of boundaries, the situation is less clear in the case of segmental lengthening. Data on this subject is often controversial. The literature on this topic suggests that the role of duration as a boundary marker is probably dependent on both language, [3], [4], [5], and mode of discourse [6], [7].

Practically all research on this topic has been concerned with carefully controlled laboratory speech. More recently, however, a series of perceptual and acoustic studies of the role of prosodic cues of finality (essentially contour type, register and range) have been carried out in discourse for Dutch [8], [9].

For French there is very little data on this subject. Crompton [10] suggested that preboundary lengthening in French is correlated with the degree of the following boundary, being greater before final boundaries than before non-final ones. This claim was only partially confirmed by Fletcher [11] in a study of isolated sentences.

In this paper we present preliminary results of a study concerning the contribution of prosodic cues : silent pause, intensity, duration and pitch-patterns, to the identification of finality in discourse, as well as the relative importance of these different cues in context and in isolation. Although there are almost certainly cues which pre-signal boundaries quite early in an utterance [12], we concentrate in this study on acoustic parameters in the immediate vicinity of the boundary.

CORPUS

Our corpus is composed of two extracts of radio broadcast interviews. For each interview we extracted and analyzed the answer to a question asked by a journalist. The first extract is uttered by a female speaker and the second one by a male speaker. Both speech turn which last approximately 1.5 minutes each correspond to a complete utterance, characterized by syntactic and semantic coherence.

TEST 1

A preliminary listening test was carried out to see how far listeners would agree on the presence of boundaries in spontaneous utterances, and how far, they would agree on the nature of the boundaries.

Experimental procedure : Ten postgraduate phonetics students listened to both recordings and were asked to add punctuation (restricted to comma, semi-colon and full-stop) to an orthographic transcription of the recording which contained no punctuation marks.

TEST 2

The aim of the second listening test was to see whether subjects were able to identify utterances as utterance final out of context.

Experimental procedure : Ten subjects with no previous experience in prosody listened first to recording 1 and then to recording 2. The recordings were broken up into 26 extracts for recording 1 and 23 extracts for recording 2. Each extract corresponded to a (final or non-final) boundary which had been identified in Test 1 by at least 50% of the subjects. The extracts were presented in random order. Each subject heard each extract three times so that in total each subject listened to 147 extracts.

Subjects were asked to key their responses to each extract as being either utterance final or utterance non-final. The responses were recorded on a microcomputer.

The results of the two tests were used to calculate two indices : a *contextual finality index* calculated as the number of subjects who had marked a full-stop at the corresponding boundary and an *isolated finality index* calculated as the number of responses designating the boundary as utterance final, divided by 3 to give a score out of 10.

Although the two indices were highly correlated ($r = 0.75$, $p < 0.0001$) they were significantly different ($t = -4.651$, $p < 0.0001$). As can be seen in Figure 1 the contextual finality index was practically always lower than the isolated index.

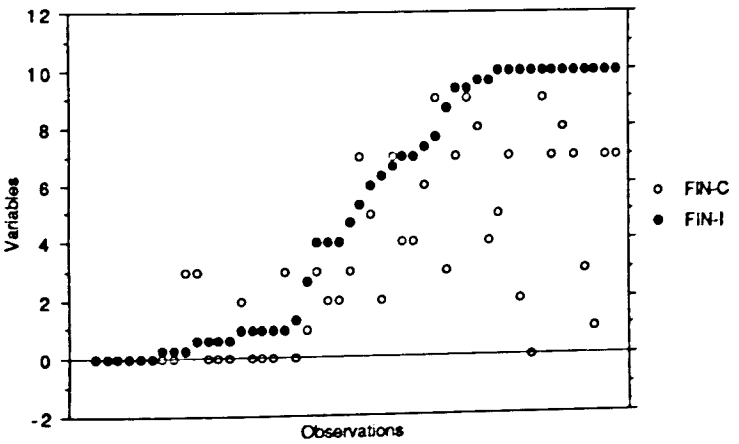

Figure 1 : isolated (FIN-I) and contextual (FIN-C) finality indices sorted by increasing value of the isolated index.

For statistical analyses the boundaries were split into three categories FINAL, NON-FINAL and INTERMEDIATE for each of the two indices, using 25% and 75% as threshold values. Thus scores between 0 and 2.5 were classified as NON-FINAL, scores between 2.5 and 7.5 as INTERMEDIATE and scores between 7.5 and 10 as FINAL. The contingency table for the nine intersecting categories thus defined is given in table 1.

	final	int.	non-final	Totals
final	5	0	0	5
int.	10	8	3	21
non-final	3	4	16	23
Totals	18	12	19	49

Table 1 : contingency table for boundaries classified as final, intermediate or non-final in context (rows) and isolated (columns).

It can be seen immediately that there is a major dissymmetry in that all 5 values coded as final by the contextual index are also coded final by the isolated index but that more than two thirds of the values coded final by the isolated index are coded as intermediate or non-final by the contextual index.

This is consistent with the idea [12] that the perception of a final boundary is carried out essentially on the basis of cues which precede the boundary. Cues which follow the boundary may weaken the perception of finality but they will not convert a non-final boundary into a final one.

Acoustic Analysis :

The corpus was digitised and analysed on a SUN-Sparc station using a MES, a signal editing environment developped by R. Espesser. Phoneme labels were placed manually. F0 was analysed and modelled as a sequence of target points defining a quadratic spline function using an automatic modelling algorithm with manual correction [13]. For each of the 49 boundaries used in the auditory tests, the following parameters were measured :

Silence : duration of silent pause following the boundary.

Intensity and duration : intensity was measured at 1/3 of the duration of the vowel. The duration of vocalic nucleus for the penultimate, final and following syllable for each boundary were measured.

Fundamental frequency : the penultimate final and following target-points were taken.

Normalisation.

The duration of the vocalic nuclei were normalised using a z-transform [14]. F0 targets were normalised with an ERB scale offset to the mean of the speaker's range [15].

Dynamic aspects

In order to capture some dynamic aspects of the prosodic features, for each triplet of values described above as penultimate, final and following we calculated successive difference values (i.e. penultimate minus final and final minus following). In all 26 acoustic features were calculated for each of the 49 boundaries analysed.

An analysis of variance was carried out for each of the 26 acoustic features on both the Isolated and Contextual classifications of the boundaries.

RESULTS

Silence : there was no significant difference in the duration of the silent pause following the boundary for either the isolated ($p = 0.114$) or the contextual classifications ($p = 0.278$).

Intensity : a significantly lower intensity value was found for the final syllable preceding terminal boundaries than for boundaries classified as either non-terminal or intermediate. This difference was significant for both contextual ($p < 0.001$) and non-contextual classifications ($p < 0.0001$).

Duration : There was no significant difference between the normalised duration of the vocalic nucleus of the final syllable preceding a final and a non-final or intermediate boundary. From the raw data the duration of this vowel appeared significantly shorter for final boundaries than for non-final or intermediate boundaries but this effect disappeared after normalisation and was probably due to the intrinsic duration of the vowels in question. There was however a difference on the penultimate vowel which was significantly shorter before a final boundary ($p < 0.05$, $p < 0.005$).

Fundamental frequency : The strongest effects ($p < 0.0001$) were found on both the absolute value of the final F0 target which was lower before a final boundary than before a non-final or intermediate boundary, and the difference between the penultimate and the final values. These differences were highly significant both before and after normalisation and for both types of classification.

DISCUSSION

Our results did not confirm the hypothesis that preboundary lengthening is greater before terminal boundaries than before non-terminal ones. Similar results were reported for English by Wightman et al. [16] who suggested that the distinction between different types of major boundaries "must be distinguished by other cues such as pausing and intonation" [p. 1716].

There was, however, no evidence from our data that the presence and/or duration of a silent pause played an effective role in the perception of terminal boundaries.

As expected, the strongest effect in our data was linked to the F0 parameter, principally that of register as expressed by the normalised pitch target preceding the boundary.

An additional finding was that a drop in intensity on the final syllable was highly correlated with the perception of finality. Despite the extemely relative nature of intensity this parameter was highly significant even though the speech analysed was not recorded under laboratory conditions.

REFERENCES

[1] Bruce, G. (1985), "Structure and functions of prosody", *Proceedings of the French-Swedish Seminar on Speech.* Guerin, B. & Carré, R. (eds.), Grenoble, France, pp. 549-559.

[2] Beckman, M. (1986), *Stress and Non-Stress Accent*, Foris : Dordrecht.

[3] Thorsen, N. (1980), "A study of the perception of sentence intonation evidence from Danish", *Journal of Acoustical Society of America*, vol. 67, pp. 1014-1030.

[4] Berkovits, R. (1984), "Duration and fundamental frequency in sentence-final intonation", *Journal of Phonetics*, vol.12, pp. 255-265.

[5] Farnetani, E. (1989), "Acoustic correlates of linguistic boundaries in Italian : a study on duration and fundamental frequency", *Eurospeech* (Paris), vol.2, pp. 332-335.

[6] Klatt, D. (1975), "Vowel lengthening is syntactically determined in a connected discourse", *Journal of Phonetics*, vol.3, pp. 332-335.

[7] Klatt, D. & Cooper, W.E. (1975), "Perception of segment duration in sentence context", *Structure and process in speech perception*, Cohen, A. & Nooteboom, S. (eds.), Springer-Verlag, pp. 69-89.

[8] Swerts, M. (1994), *Prosodic Features of Discouse Units*, Doctoral Dissertation, Eindhoven University of Technology.

[9] Swerts, M. & Bouwhuis, Don G. and Collier, R. (1994), "Melodic cues to the perceived "finality" of utterances", *Journal of Acoustical Society of America*, vol. 96, pp. 2064-2075.

[10] Crompton, A. (1980), "Timing Patterns in French", *Phonetica*, vol.37, pp. 205-234.

[11] Fletcher, J. (1991), "Rhythm and final lengthening in French", *Journal of Phonetics*, vol.19, pp. 193-212.

[12] Swerts, M. & Geluykens, R.C. (1994), "Prosody as a marker of intonation flow in spoken discourse", *Language & Speech*, vol. 37, pp. 21-43.

[13] Hirst, D.J. & Espesser, R. (1993), "Automatic modelling of fundamental frequency with a quadratic spline function." *Travaux de l'Institut de Phonétique d'Aix*, vol. 15, pp. 71-85.

[14] Campbell, W.N. (1992), *Multi-level Timing in Speech*, PhD thesis, University of Sussex.

[15] Hermes, D.J. & Van Gestel, J.C. (1991), "The frequency scale of speech intonation.", *Journal of Acoustical Society of America*, vol. 90, pp. 97-102.

[16] Wightman, C., Schattuck-Hufnagel, S., Ostendorf, M. and Price, P. (1992), "Segmental durations in the vicinity of prosodic phrase boundaries", *Journal of Acoustical Society of America*, vol. 91, pp. 1707-1717.