

PERCEPTION OF A PROSODIC BREAK: THE CASE OF INCIDENTAL PHRASES IN FRENCH

N. BACRI *
Lab. de Psychologie Expérimentale

A. NICAISE **
U. F. R. en Linguistique

* C.N.R.S. - E.H.E.S.S. 54 bd Raspail 75006 PARIS, France

** PARIS VII 2 pl Jussieu 75005 PARIS et Paris XII, France

An acoustic analysis of the production of incidental phrases provided a basis for a perceptual study of prosodic boundaries. A break is characterized by the introduction of a pause, a lengthening of the vowel preceding the pause and a F_0 contrast. The three types of modification were taken as potential cues and stimuli with 1, 2 or 3 cues for a break were synthesized. Subjects had to decide whether the utterance they heard contained an incidental phrase or not. Though the pause was the factor with the greatest weight, it was not in itself sufficient to induce a significant proportion of "break" responses. The conjunction of F_0 modification and pause was necessary to get that result. Interactions between the three parameters were significant. The notion of prosodic configuration based on the integration of several cues is proposed to explain these results.

Intonation has often been described as the succession of F_0 patterns and a F_0 contour as a "series of targets within an envelope specifying F_0 range" /7, p.985/. Pitch movements should be sufficient to convey information about the internal structure of the sentence and even about its linguistic meaning /2/. Some other studies have focused on the relevance of durational cues to the perception of phrase boundaries /4, 10/. However a number of recent studies have suggested that, at an auditory level, separate acoustic features, specifically durational and spectral cues, are evaluated and integrated to form a single percept /1, 3, 8, 9/. So, intonation has been studied from a multiparametric point of view and the question of the integration of several information sources into a prosodic pattern has been raised.

The first aim of the present experiment is to assess how the acoustic correlates of prosody are processed to produce a perceptual continuity or, on the contrary, a perceptual break. In order to determine the nature of integration processes, it was necessary to independently vary these acoustic features. Moreover, it is often taken for granted that the contrast between intonation contours is

primarily based on the direction and amplitude of F_0 variations /5/. This hypothesis will be confronted with results derived from the study of the perception of a prosodic break in French. If the data pertaining to perceptual integration are really of importance, parsing should be accomplished either as the summation of separate evaluations or as the weighting of the outputs of temporal and pitch processors, depending on the presence of significant interactions. A second aim of the experiment is to analyse if acoustically different prosodic configurations can be judged, under certain conditions, as linguistically equivalent. The linguistic constraints, for a given language, can help the listener to construct the semantic interpretation for an utterance. The variability of prosodic organizations for a given sentence is nonetheless well known. Stimuli have, accordingly, been chosen from the productions of female and male speakers, in order to evaluate how listener's parsing strategies are matched with speaker's production strategies, and so to assess the degree of adjustment of the listeners to the speech signal.

Preliminary experiment

In a preliminary study of the production of 14 speakers, French utterances containing an incidental phrase as their second constituent were compared with utterances with the same segmental content but no intonational break. This study revealed that the following characteristics could be used to differentiate the two cases :

- Lengthening of the vowel preceding the break (present in 62% cases, it varies between 30% and 100% of the basic duration of the vowel in the utterance that does not include a break)

- Pause preceding the parenthetical clause of phrase (present in 76% cases, its duration can reach 325 ms)

- Pause following the parenthetical constituent (present in 76% cases, its duration can reach 210 ms)

	Continuity contour		Incidental Phrase	
	F	M	F	M
/o/ duration	80 ms	100 ms	125 ms	218 ms
left pause	0 ms	0 ms	290 ms	110 ms
right pause	110 ms	70 ms	120 ms	210 ms
/i/ F_0 value	263-238 Hz	159-152 Hz	192-182 Hz	149-143 Hz
/bo/ F_0 val.	222-200 Hz	145-182 Hz	204-323 Hz	130-170 Hz
/deka/ F_0 val.	189/250-260 Hz	120/147-156Hz	170/244-270 Hz	120/120-160 Hz
Speech rate	5.08 syl/s	4.60 syl/s	3.59 syl/s	3.57 syl/s
Artic. rate	6.18 syl/s	5.17 syl/s	5.26 syl/s	4.65 syl/s

Table I - Original stimuli: Continuity contour "Thibaut de Caen est là", Incidental Phrase contour "Thibaut, de Caen, est là", as produced by a female speaker, F, and a male speaker, M.

- Highest F_0 -peak of the utterance on the final syllable of the parenthetical (86% cases).

Moreover, a comparison of the utterances containing a parenthetical clause with the corresponding "normal" utterances revealed that the strategy that was most frequently adopted by speakers was a succession Rise + Rise on the first two constituents of the utterance containing a parenthetical versus a Fall + Rise pattern in the other case (76% cases were of this type).

The experiment described attempts to :

- 1) establish the validity of those differences as potential cues
- 2) specify their role in the perception of an intonational break. Is one of the cues primary or even necessary ? When do integration processes take place ? The problem is to determine how the information from various dimensions of the signal is put together to produce an interpretation .
- 3) evaluate the stability of a prosodic configuration through two different realizations, and the extent of the adjustment of listeners to speakers.

Experiment: Stimuli

The stimuli used in the experiment were derived from two productions of the utterance: "Thibaut de Caen est là" read by two different speakers (one male, one female). Those stimuli were then sampled at 10 KHz (12 coefficients) and a LPC analysis was performed (10 ms window). The relevant characteristics of the original stimuli are summarized in Table I.

We have simulated the presence of a parenthetical on "de Caen" using the following modifications of the stimuli without a prosodic break (Table II):

	Stimulus Number											
	1	2	3	4	5	6	7	8	9	10	11	12
V.L	50	∅	50	50	25	25	25	∅	∅	50	∅	25
Pause	100	100	0	100	100	0	100	0	100	0	0	0
F_0	RR	RR	RR	FR	RR	RR	FR	RR	FR	FR	FR	FR

Table II - Characteristics of the stimuli: vowel lengthening (ms) before the potential pause, pause (ms), F_0 contour (R: rise, F: fall).

- Duration of the final vowel of the constituent preceding the potential parenthetical. Vowel duration took 3 values : original value, 25 ms and 50 ms lengthening.

- Presence or absence of a pause before the potential parenthetical (duration : 100 ms). (Those values were deliberately chosen as minimal).

- F_0 contour on the first two constituents of the utterance. F_0 contour was modified so as to reproduce the F_0 contour of two utterances containing a parenthetical taken from the production of the speakers of the two original stimuli. It should be noted that the strategies used by the two speakers are different (Table I). The three factors were combined to produce 12 stimuli per speaker presenting 0,1,2 or 3 cues favoring the break interpretation.

Subjects and procedure

Four blocks of stimuli were composed (2 x 12-item blocks per original speaker, each stimulus being presented twice). The stimuli were randomized within each block. There was a 7-sec pause between items and a 30-sec pause between blocks. 16 subjects passed the experiment. The factorial design was as follows:

$S_{16} * D_3 * P_2 * F_2 * O_2$
(S: subjects, D: vocalic lengthening, P: pause, F: F_0 pattern, O: presentation order). Subjects had to decide whether the utterance they heard contained an incidental phrase or not. They were also asked for a confidence-rating on a 3-point scale. The experiment proper was preceded by 4 stimuli derived from the original productions of the speakers. The answers of one of the subjects were not analysed because he could not identify those stimuli correctly.

Results

All the stimuli were significantly identified either with one category or the other, except the stimulus with 50 ms lengthening, F_0 contour corresponding to the presence of a parenthetical but no pause. Taking the degree of certainty into account, the answers were coded on a 0-1 scale (0=Non-parenthetical, high certainty, 1=parenthetical, high certainty). The results are summarized in Table

III (the stimuli are classified according to the number of potential cues and numbered for easy reference). An analysis of variance revealed that the order of presentation and speaker factors had no significant effect, in spite of the fact that the F_0 strategies for the two speakers were different. A second analysis was performed, grouping the scores corresponding to the non significant factors. The presence/absence of a pause ($F(1, 14) = 165.7, p < .001$), duration of the final vowel ($F(2, 28) = 34.35, p < .001$) and F_0 contour ($F(1, 14) = 41.3, p < .001$) were highly significant factors. Interactions between vowel-duration and pause ($F(2, 28) = 10.3, p < .01$), vowel-duration and F_0 contour ($F(2, 28) = 6, p < .01$) were also significant.

	Stimuli											
	1	2	3	4	5	6	7	8	9	10	11	12
m	.83	.77	.61	.60	.76	.21	.55	.18	.50	.22	.12	.09
sd	.05	.06	.15	.10	.05	.08	.06	.06	.07	.07	.03	.03

Table III - Mean scores and standard deviations as a function of the number of continuity cues (0-1 scale, cf. text).

As expected, stimuli 1 and 11 were identified to an utterance containing a break or an utterance without a break (respectively) (Figure 1). An examination of the stimuli containing only one "break-cue" (stimuli 8-10 and 12) showed that the presence of a pause was the factor with the greatest weight, though it was not in itself sufficient to induce a significant proportion of "break responses" (st.9). The influence of vowel length is limited (that factor accounts for a variation of 20 to 25% in the scores). Comparisons on that factor reveal no significant difference between 0 and 25 ms lengthening but a difference between these and 50 ms lengthening. The original F_0 continuity contour is sufficient to counterbalance the influence of the other factors (st.4) but the contour corresponding to the incidental utterances has a smaller influence than expected (st.8).

Discussion

The parameters that varied in this experiment are traditionally considered as cueing a boundary. Though they were confronted with partially incompatible indices, the subjects of the experiment managed to reconstruct a linguistically pertinent prosodic organization: the distribution of their answers was not due to chance or determined solely by one of the factors manipulated. Though we noted that the pause was the factor with the greatest weight, the conjunction of F_0 modification and pause was necessary to induce a significant proportion of "break responses". The hypothesis of F_0 pattern as the dominant cue is therefore not confirmed. F_0 contours have an important effect but only when they are accompanied by another compatible characteristic. More, the results suggest that

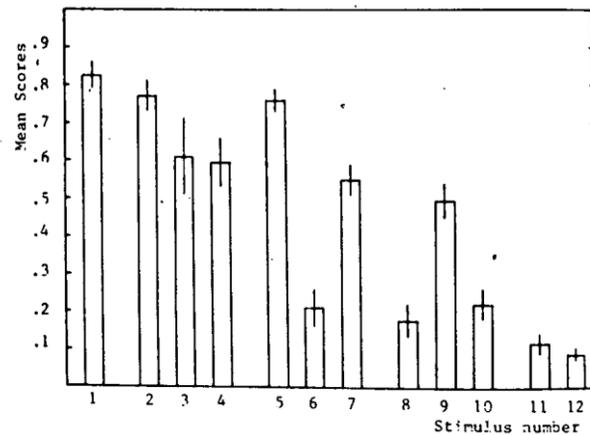


Figure 1 - Mean scores and confidence intervals as a function of the number of continuity cues. Stimulus 1: 0 continuity cue; stimulus 11: 3 continuity cues; 0-1 scale; cf. text.

there is no dominant cue but rather that we are dealing with a prosodic configuration based on the integration of several cues. Though there was no global interaction between F_0 and pause, partial comparisons reveal a significant interaction between pause and sharp rise on the first constituent. We may interpret this in terms of contrast formation: a sharp rise and a pause are integrated in one percept and the resulting configuration introduces a contrast between two successive chunks of the signal. These results suggest a parallel processing of duration and pitch followed by a weighting of their outputs. The contrastive configuration sets off the decision process.

As far as F_0 is concerned, the two speakers used two different strategies: the female speaker used a sharp rise resulting in a local contrast in the parenthetical case whereas the male speaker modified the hierarchy between two successive peaks (first peak higher than second peak in the non parenthetical case, two peaks of equivalent height in the parenthetical case). The effect of the two strategies is equivalent. But to account for the effect of the strategy used by the male speaker we must posit that a comparison is performed between two successive peaks, the two peaks being more than one second apart. Moreover, this comparison implies that a reference line (declination line) is taken into account: the parenthetical case would indicate that the break is accompanied by a resetting of the reference line.

The influence of vowel length is less easy to determine. The effect of this factor in the statistical analysis shows that it is taken into account but there is no linear effect of vowel lengthening. We propose that vowel length is not an independent cue but rather a property of the configuration: vowel

lengthening causes the stimuli that include that characteristic to be judged as better tokens of the break configuration, provided this configuration is already established through other characteristics of the signal. Of course, the present conclusion applies only to rather small values for vowel lengthening.

CONCLUSIONS

The experiment showed that a set of characteristics including vowel length, pause and F_0 movement may be perceived as a configuration. It also showed that the F_0 contours that characterize a prosodic break in production are not always sufficient to cue that break in perception. The break is constructed primarily on the conjunction of F_0 pattern and pause. The significance of the interaction between vocalic lengthening and pitch movements suggests that listeners do not proceed to a mere summation of information. Integration processes plausibly take place at the output of the duration and pitch processors, according to a weighting procedure /6/. It's worth noting that the responses to the stimuli derived from the female and male utterances were on the whole the same. Though the two speakers used a different strategy, the F_0 patterns were characterized by a difference in F_0 level before and after the left boundary of the parenthetical constituent. The greater efficiency of F_0 movement in the presence of a pause might be explained in terms of the suppression of a retro-active masking /6/, induced by the pause. The adjustment procedure seems to be automatic. The assumption is made that it is supported by the integration of the contrast between successive F_0 peaks. The experiment reported here is also a confirmation that, in order to predict the effects of a prosodic configuration, a linguistic model of prosody should take into account, not only fundamental frequency movements, but also pauses and vowel length.

REFERENCES

- /1/ A.S. Abramson, L. Lisker, "Relative power of cues: F_0 shift vs. voice timing", Haskins Lab., Status Report on Speech Research SR-77/78, 121-128, 1984.
- /2/ W.A. Ainsworth, D. Lindsay, Perception of pitch movement on tonic syllables in British English, "J. of the Acoustical Soc. of America", 79(2), 472-480, 1986.
- /3/ N. Bacri, "Fonctions de l'intonation dans l'organisation perceptive de la parole", Thèse de Doctorat d'Etat, Université Paris VIII, 1986.
- /4/ I. Gee, F. Grosjean, Performance structures: a psycholinguistic and linguistic appraisal, "Cognitive Psychology", 15, 411-458, 1983.
- /5/ Ph. Martin, Faits prosodiques et théorie

- de l'intonation, "Bulletin d'Audiophonologie", 1-2, 37-47, 1985.
- /6/ D.W. Massaro, M.M. Cohen, Voice onset time and fundamental frequency as cues to the /zi-/si/ distinction, "Perception and Psychophysics", 22, 373-382, 1977.
- /7/ J. Pierrehumbert, Synthesizing intonation, "J. of the Acoustical Soc. of America", 70(4), 985-995, 1981.
- /8/ B.H. Repp, A.M. Liberman, "Phonetic category boundaries are flexible", Haskins Lab., Status Report on Speech Research SR-77/78, 31-53, 1984.
- /9/ M. Rossi, A. Di Cristo, D. Hirst, Ph. Martin, Y. Nishinuma, "L'intonation. De l'acoustique à la sémantique", Paris: Klincksieck, 1981.
- /10/ N. Umeda, A.M. Quinn, Word duration as an acoustic measure of boundary perception, "Journal of Phonetics", 9, 19-28, 1981.