Pitch and Timing Cues in Speech Intelligibility: the Case of Child Language

N. Bacri
Paris, France

1. Introduction

These last years, several investigators have presented evidence for a direct link between timing and pitch cues in sentence perception. It has been shown that prosodic parameters allow listeners to disambiguate sentences (Lehiste et al., 1976), whereas deviant durations and fundamental frequency contours decrease intelligibility (Huggins,1978, 1979). Moreover, data bearing on the relative effect of temporal cues and pitch movements point to the greater efficiency of the former. In experiments where the three suprasegmental variables for one meaning of an algebraic expression are interchanged for the other meaning of the expression, Streeter (1978) demonstrated that the magnitude of the pitch effect was about 14-20%, depending on the speaker. When conflicting cues are introduced, lengthening alone gave a correct score of 90% (Nooteboom et al. 1978), but F_0 overrode the cue value of pauses as boundary markers (Larkey 1981). However, if local variations of F_0, except in the latter case, are not a crucial factor for sentence comprehension, the contribution of prosody to speech perception appears to be more important than suggested by these data. Duration and pitch are analysed as connected cues and form an integrated percept. They supply different information, and specifically, prosodic continuity seems to be essential to the listener's ability to perceive speech at all (Nooteboom et al., 1978, Carlson et al., 1979).

In reply to the question as to what kinds of perceptual cues are available to allow a listener to understand speech with a low degree of intelligibility, I have put forward the hypothesis that compatibility between timing and pitch cues support intelligibility whereas conflicting cues result in misunderstanding. In this paper, I try to answer this question by comparing temporal organization and pitch patterns of two types of utterances, judged intelligible or not.

2. Subjects, test material and procedure

Three adults (two females and one male), native speakers of Parisian French, were asked to sort out 1450 utterances of spontaneous speech, produced by a child, into three categories: intelligible (IU), unintelligible (UU), undecidable. Utterances were classified as follows: 463 IU (32%) and 859 UU (59%);

128 utterances (9%) were judged undecidable by at least one judge. The child, aged 1;10 at the beginning of the study and 2;2 at the end, was visited six times at home, every three weeks. Each session lasted 45 minutes and was recorded. All utterances were transcribed in IPA by a phonetician who then calculated the number of syllables in each sequence. An independent rater likewise computed the number of syllables by sequence. A level of agreement of 0.96 was attained between the two ratings (de Bardies et al., 1981).

The present study does not bear on phonemic distribution. However, it is worth noticing that is is significantly different from adult French distribution.

For each utterance, a pitch detector plotted variations of F_0 and amplitude of the speech signal as a function of time. The following variables were measured: total duration of utterances, initial and final syllabic durations, overall speaking rate (duration of utterances / number of syllables / sec.), articulation rate or time spent articulating (duration of utterances minus duration of pauses / number of syllables / sec.), duration and frequency of pauses, mean F_0 value for the first F_0 plot, amounts of rises and falls, average number of fluctuations per second and per syllable in the F_0 patterns. For the last two variables, a fluctuation was defined as a point corresponding to a local maximum or minimum on the F_0 curve. In order to be sure of the perceptual relevance of F_0 movements, a F_0 change was labelled fluctuation if it exceeded 35 Hz during at least 50 msec.

3. Timing analysis: durational cues

Analysis of variance was carried out on 360 utterances from 2 to 5 syllables, equally distributed over the sessions. Total duration of utterances had no significant effect on intelligibility, and neither had the number of syllables. The main effects were due to initial and final syllabic durations. Initial syllables in IU were shorter than in UU ($F(1, 356) = 6.30, p < .02$). The mean durations were 139 msec. and 156 msec., respectively. Final syllables in IU were longer than in UU ($F(1, 356) = 5.84, p < .05$). The mean durations were 185 msec. and 166 msec. respectively. Moreover: final syllables in IU increased in length with the increased length of utterances. The interaction between final syllabic duration and number of syllables was significant ($F(1, 178) = 4.98, p < .05$). However, the degree of variation for syllabic durations was considerable, as previously noted by Oller and Smith (1977) for child language. So an analysis of syllabic durations was performed session by session, and on the whole corpus, data being partitioned in short and long utterances according to the number of syllables (2+3 vs 4+5 syllables). *t*-test showed significant differences between IU and UU, but only for initial syllabic durations of long utterances. *t*-test for paired observations performed on IU and UU respectively showed systematic differences between initial and final syllabic durations for IU (short IU: $t(5)=4.25, p < .01$; long IU: $t(5)=5.69, p < .01$), but not for UU (Fig. 1).

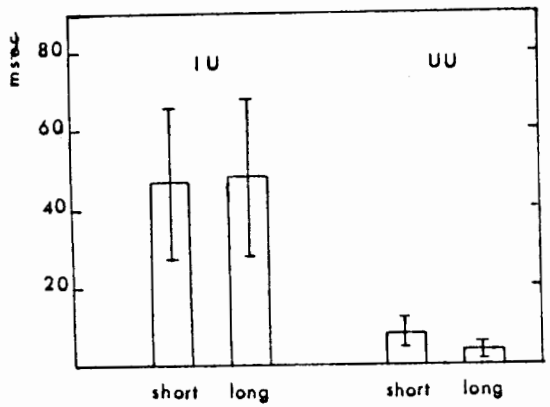

Figure 1. Mean differences and confidence intervals (in ms.) between initial and final syllabic durations, within short and long sequences, for intelligible (IU) and unintelligible (UU) utterances.

Moreover, changes in duration varied from 21% to 30% in IU and so, according to Klatt (1976), could serve as perceptual cues, whereas they varied from 1% to 8% in UU and it was not possible to discriminate them.

The main difference between IU and UU comes from the contrast between initial and final syllabic durations in IU, a contrast that does not appear in UU. IU alone present a regular temporal organization, and therefore a rhythmical structure, which can be used by the listener as a cue to the linguistic value of the utterances, mainly to their boundaries. These local variations are probably directly linked to changes in the spectral properties of speech (Miller 1981), and can facilitate the judgment of intelligibility. But it is also well known that rate of speech, and particularly articulation rate, influence temporal characteristics and the way in which intelligibility is achieved. So it is worth studying whether these local variations are linked to variations in speech rate, to some kind of extrinsic timing.

4. Timing analysis: pauses, speaking rate, articulation rate

For the two types of utterances, pause length was about 20% to 27% of the total mean duration. IU were characterized by one pause in median position, just after the first syllable (short utterances) or after the second or third one (long utterances). In all cases, pause was preceded by syllable lengthening. The mean prepausal syllabic duration was 168 msec. On the other hand, no preferential position for pauses and no lengthening were found for UU. Speaking rate and articulation rate were similar for the two types of utterances. Speaking rate was 4.23 syll./sec. for IU, 4.40 for UU; articulation rate was 5.75 syll./sec. for IU, 5.99 for UU;

Our data suggest the perceptual relevance of relative timing in the intelligibility of child language. But the role of speech rate (overall speaking rate and

articulation rate) has not been confirmed. It seems likely that intrinsic timing, i.e. variations in initial and final syllabic durations, is not related to changes in the rate of speech. It is not possible to conclude from the data that listeners adjust to speech rate, in order to judge utterances. It can be assumed that it is intrinsic timing alone which serves as a cue because it supports the rhythmical structure of intelligible utterances. It can be thought that the contrast between initial and final syllabic durations in IU is a cue both to the phonetic identity and to the identification of utterances boundaries. This last hypothesis concerning the role of timing cues in intelligibility is confirmed by the close correspondence between temporal and speech patterns.

5. Fundamental frequency analysis

Analysis of F_0 initial mean value has showed its steadiness through the whole corpus for every kind of contour, rising, flat or falling (336 to 294 Hz for IU; 340 to 322 Hz for UU). Amounts of rises and falls had similar ranges of values, about 160 Hz for the rises, 100 Hz for the falls. No difference was found between types of utterances. However, it is worth noticing that F_0 changes exceeding 100 Hz could occur on arbitrary positions for UU, whereas they tended to occur on the last syllable for IU, or on a syllable preceding a pause. In this last case, the utterance was identified as a sequence of two constituents.

Analysis of the two 'dynamic' variables indicated that the average number of fluctuations per syllable and per second was quasi constant on the whole corpus, whereas for each session the two types of utterances differed significantly. The number of fluctuations per syllable was two to three times smaller for IU than for UU. Values of t (6) varied from 3.6, $p < .02$ at 1;10 year old to 9.80, $p < .001$ at 2;2 year old. Average number of fluctuations per syllable was .666 for IU, 1.26 for UU (Fig. 2).

These results suggest that one fluctuation per syllable impairs intelligibility. Moreover, small fluctuations occurred in all syllabic positions for UU, whereas they occur for IU on the last one or the last two syllables, final rise or fall being preceded by one or two small F_0 changes. In long IU, a third fluctuation occurred only before a 200 to 300 msec. pause. Simultaneity of occurrence of durational and pitch boundary markers seems to facilitate intelligibility judgment. This explanation is confirmed by analysis of the average number of fluctuations per second; 2.78 for IU, 4.95 for UU (Fig. 2). As a consequence, IU F_0 slopes were smoother than UU ones.

The close relationship between intelligibility judgment and the rate of F_0 change suggests that abrupt slopes and frequent jumps lower intelligibility, particularly when large F_0 changes do not precede a pause and when a lot of small fluctuations disturb the establishment of pitch continuity.

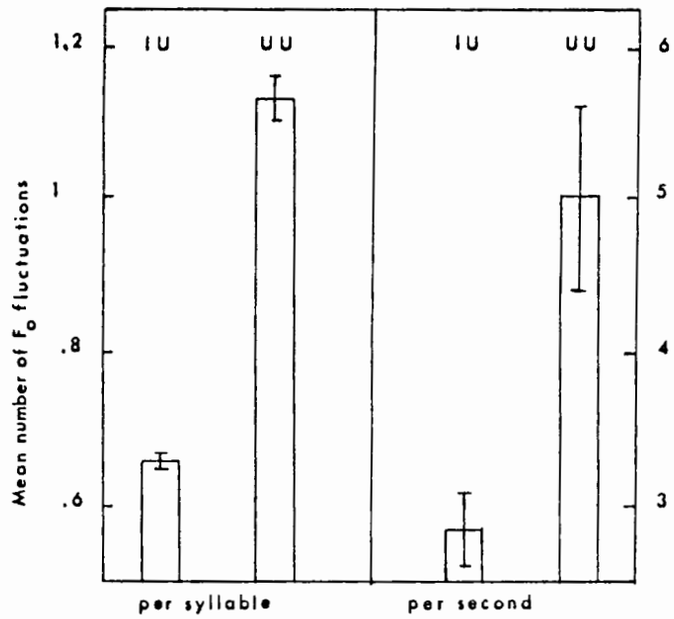

Figure 2. Mean number of F_0 fluctuations per syllable and per second for intelligible (IU) and unintelligible (UU) utterances, with confidence intervals.

6. Discussion

It is well known that the French language is characterized by a final accent marked by a final syllabic lengthening as well as by changes in F_0 (Delattre, 1966, Crompton, 1980; Rossi et al., 1981), but there is neither a rapid jump, nor a break in F_0 slope exceeding 40% or 50% of the former value (Delgutte, 1978; Vaissière, 1980). In this paper, I have attempted to show that prosodic patterns differ systematically between utterances perceived by adults as babbling or as first language. However, neither differences in overall duration, speaking or articulation rates, nor differences in F_0 initial values, amounts of rises and falls, can account for intelligibility judgments. Relative lengthening and pitch continuity seem to be the main perceptual cues.

In spite of the great variability in prosodic realizations, close correspondence between syllabic lengthening and F_0 movements preceding boundaries in IU, as well as the smoothness of F_0 slopes, confirm the importance of prosody in speech perception. Conflicting cues providing misleading information significantly impair intelligibility. I cannot specify here which is the main determinant of listeners' judgment: duration or pitch. However, the results suggest that pitch continuity contributes to intelligibility by establishing a process of backward perceptual normalization, related to the intrinsic timing of utterances, that facilitates syllabic identification and the detection of boundary markers.

Though the data show that boundary detection contributes to the listener's construction of an integrated percept, this does not exclude the role of segmental features, which I have not studied here. It is also worth noticing that this study bears on the language of a French child, and that it needs to be confirmed with other languages.

References

de Boysson-Bardies, B., Bacri, N., Sagart, L., and Poizat, M. (1981). Timing in late babbling. *J. Child Lang.* **8**, 525-539.

Carlson, R., Granström, B. and Klatt, D.H. (1979). Some notes on the perception of temporal patterns in speech. *Ninth Inter. Congress of Phonetic Sciences*, II 260-267, Copenhagen.

Crompton, A. (1980). Timing patterns in French. *Phonetica* **37**, 205-234.

Delattre, P. (1966). Les dix intonations de base du français. *French Review*, **40**(1), 1-14.

Delgutte, B. (1978). Technique for the perceptual investigation of F_0 contours with application to French. *J. Acoust. Soc. Am.* **64**(5), 1319-1332.

Huggins, A.W.F. (1978). Speech Timing and Intelligibility. J. Requin (ed.). *Attention and Performance* VII, Hillsdale, Erlbaum Ass.

Huggins, A.W.F. (1979). Some effects on intelligibility of inappropriate temporal relations within speech units. *Ninth Intern. Congress of Phonetic Sciences*, II 283-289, Copenhagen.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.*, **59**(5), 1208-1221.

Larkey, L.S. (1981). *The role of prosodic information in speech perception.* Doct. Dissertation, Univ. of Minnesota.

Lehiste, I., Olive, J.P. and Streeter, L.A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.*, **60**(5), 1199-1202.

Miller, J.L. (1981). Effects of Speaking Rate on Segmental Distinctions. P.D. Eimas and J.L. Miller (eds.) *Perspectives on the Study of Speech.* Hillsdale, Erlbaum Ass.

Nooteboom, S.G., Brokx, P.L. and de Rooij, J.J. (1978). Contributions of prosody to Speech Perception. W. Levelt and G. Flores d'Arcais (eds.) *Studies in the Perception of Language.* N.Y.: Wiley.

Oller, D.K. and Smith, B.L. (1977). Effect of final-syllable position on vowel duration in infant babbling. *J. Acoust. Soc. Am.*, **62**(4), 994-997.

Rossi, M., Di Cristo, A., Hirst, D., Martin, Ph. and Nishinuma, Y. (1981). *L'Intonation: de l'acoustique à la sémantique.* Paris: Klincksieck.

Streeter, L.A. (1978). Acoustic determinants of phrase boundary perception. *J. Acoust. Soc. Am.*, **64**(6). 1582-1592.

Vaissière, J. (1980). Structuration prosodique de la phrase française. *Les Annales de l'Ecole Normale Supérieure de Pise*, Pise.