

# Introduction to Part II

Anne Cutler

## 5.1 Prosody and the Structure of the Message

Most of the speech that speakers produce and listeners hear is spontaneous, and intended for the purpose of communicating. Apart from the occasional monologue, or the mutterings of a deranged passerby, the greater part of the speech we experience assumes the availability of a listener. Speakers use prosodic means (among others) to communicate to listeners the structure of the message that they wish to impart.

Intonational prominence, in particular, is a prosodic device which does principal service as an indicator of message structure. Moreover, it is a phenomenon of considerable cross-linguistic generality. The analyses reported in the contributions in this section are based on several different languages—American English, Dutch, Japanese. One finding which appears across languages (e.g., in Nakatani's analysis of English, and Nakajima and Tsukada's analysis of Japanese—is that a shift in discourse topic is accompanied by raised FO. The same result has been reported for Scottish English, by Brown, Currie, and Kenworthy [BCK80] and for speech of American parents to children, by Menn and Boyce [MB82]. Indeed, Bolinger [Bol78] listed obtrusions for prominence as (along with the expression of closure) a truly language-universal use of prosody.

The relationship between intonational prominence and message is exploited by listeners during the processing of spoken language. Thus listeners accord a high priority to the task of detecting where sentence accent falls in a speaker's utterance; preceding prosodic cues enable listeners to direct attention to accents [Cut76]. If part of the normally available prosodic information is absent, listeners will make use of what remains [CD81]; but it seems that no one prosodic dimension is paramount in signalling accent location, because conflict between different sources of prosodic information (e.g., FO and rhythm) leaves listeners unable to predict where accent will fall [Cut87].

The importance of seeking accent location is explained as a search for focussed, or semantically central, aspects of the speaker's message [CF79]. In fact listeners are extremely efficient at processing the mapping

of discourse structure onto accentuation patterns, and extremely sensitive to mismatch in the mapping [BM83, TN87, FH87, TH94, vDL94].

Since the target of most listening is spontaneously uttered speech, it is reasonable to assume that the processing abilities of listeners as evidenced in these laboratory studies have been developed in the spontaneous situation. Yet the experimental studies cited above have been carried out almost without exception using speech materials which have been carefully constructed for the purpose and read from text in a laboratory situation. It is thus reasonable to ask whether we are as yet in possession of the full story regarding the processing which listeners apply to the speech they hear in most everyday situations.

Spontaneous speech and read speech differ with regard to prosodic structure: the former has, in particular, shorter prosodic units and more frequent pauses and hesitations (see, e.g., Crystal & Davy [CD69]). Thus listening procedures which involve the tracking of prosodic contours of some extended duration may be less well served by the average spontaneously spoken utterance. Indeed, Mehta and Cutler [MC88] found that the pattern of listeners<sup>1</sup> responses in a phoneme detection task performed on spontaneously uttered materials differed from the response pattern obtained with exactly the same materials produced as read speech. In particular, an effect of response facilitation for phoneme targets occurring later in the sentence, which appears consistently with read-speech materials including those of Mehta and Cutler's own study, disappeared with the spontaneous utterances. This effect has been variously interpreted as reflecting prediction of target location from syntactic, semantic, or prosodic structure; since the materials in Mehta and Cutler's read versus spontaneous conditions did not differ syntactically or semantically, the failure to find the position effect in one of two conditions which differed prosodically certainly supports a prosodic interpretation, and suggests that prosodic prediction may be of limited applicability with spontaneous input.

On the other hand, Mehta and Cutler did find response facilitation for targets on accented as opposed to unaccented words in spontaneous utterances. Note that intonational prominence—obtrusion of an intonational peak from median FO across an utterance—tends in fact to be greater in spontaneous than in read speech [vB90, vB91]. Mehta and Cutler argued that the prosodic characteristics of spontaneous speech (such as shorter prosodic units and hence more frequent occurrence of relative accent) allow rich opportunities for the exercise of some processing strategies in the listener's repertoire, but poorer opportunities for the exercise of others. Moreover, there may of course exist processing strategies of particular and exclusive usefulness for the processing of spontaneous speech which have as yet not been revealed by experimental investigation. Certainly the modelling of the prosodic structure of spontaneous speech is, given listener sensitivity to effects at this level, an enterprise likely to pay off in the construction of user-friendly synthesis and recognition systems.

## References

- [BCK80] G. Brown, K. L. Currie, and J. Kenworthy. *Questions of Intonation*. London: Croom Helm, 1980.
- [BM83] J. K. Bock and J. R. Mazzella. Intonational marking of "given" and "new" information: Some consequences for comprehension. *Memory & Cognition*, 11:64-76, 1983.
- [Bol78] D. L. Bolinger. Intonation across languages. In J. H. Greenberg, editor, *Universals of Human Language*. Stanford: Stanford University Press, 1978.
- [CD69] D. Crystal and D. Davy. *Investigating English Style*. London: Longman, 1969.
- [CD81] A. Cutler and C. J Darwin. Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, 29:217-224, 1981.
- [CF79] A. Cutler and J. A Fodor. Semantic focus and sentence comprehension. *Cognition*, 7:49-59, 1979.
- [Cut76] A. Cutler. Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 20:55-60, 1976.
- [Cut87] A. Cutler. Components of prosodic effects in speech recognition. In *Proceedings of the 11th International Congress of Phonetic Sciences, Tallin, Estonia*, Vol. 1, pp. 84-87, 1987.
- [FH87] C. A. Fowler and J. Housum. Talkers' signalling of 'new' and 'old' words in speech, and listeners' perception and use of the distinction. *Journal of Memory & Language*, 26:489-504, 1987.
- [MB82] L. Menn and S. Boyce. Fundamental frequency and discourse structure. *Language & Speech*, 25:381-383, 1982.
- [MC88] G. Mehta and A. Cutler. Detection of target phonemes in spontaneous and read speech. *Language and Speech*, 31:135-156, 1988.
- [TH94] J. Terken and J. Hirschberg. Deaccentuation and persistence of grammatical function and surface position. *Language & Speech*, 37:125-145, 1994.
- [TN87] J. Terken and S. G. Nooteboom. Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language & Cognitive Processes*, 2:145-163, 1987.

- [vB90] F. J. Koopmans van Beinum. Spectro-temporal reduction and expansion in spontaneous speech and read text: The role of focus words. In *Proceedings of the International Conference on Spoken Language Processing, Kobe, Japan, Vol. 1*, pp. 21-24, 1990.
- [vB91] F. J. Koopmans van Beinum. Spectro-temporal reduction and expansion in spontaneous speech and read text: Focus words versus non-focus words. *Proceedings of the ESCA Workshop on Phonetics and Phonology of Speaking Styles, Vol. 5*, pp. 1-36, 1991.
- [vDL94] W. van Donselaar and J. Lentz. The function of sentence accents and given/new information in speech processing: Different strategies for normal-hearing and hearing-impaired listeners? *Language and Speech*, 37:375-391, 1994.