PROSODY IN SITUATIONS OF COMMUNICATION: SALIENCE AND SEGMENTATION

Anne Cutler

MRC Applied Psychology Unit, Cambridge, UK.

ABSTRACT

Speakers and listeners have a shared goal: to communicate. The processes of speech perception and of speech production interact in many ways under the constraints of this communicative goal; such interaction is as characteristic of prosodic processing as of the processing of other aspects of linguistic structure. Two of the major uses of prosodic information in situations of communication are to encode salience and segmentation, and these themes unite the contributions to the symposium introduced by the present review.

1. INTRODUCTION

Communication is what speech is for. Everything about speech is somehow involved in the relationship between speaker and listener. Is there anything special to say about the role of prosodic structure in this relationship?

One rather negative claim that has shown up in a number of forms is that prosody is in some sense not central to the message being communicated. Among the reasons cited are that prosody encodes affect, which, while it may be communicated, is not part of linguistic structure; or that the dimensions of prosody are duration, intensity and fundamental frequency, and since every speech sound must have some duration, intensity and fundamental frequency, prosody simply falls out of the fact that speech is realised acoustically. The fact that most orthographies do not encode prosody is sometimes seen as supporting evidence for the claim that prosody is inessential.

These days it is presumably unnecessary to argue against this point of view. However, the contributions to the present symposium certainly provide counter-evidence to it. In this introductory review paper, I shall present evidence from studies of speech processing showing that the processing of prosody is subject to the same interacting constraints of the perception and production systems as affect the processing of other aspects of linguistic structure.

2. PROSODY

There is, of course, no one-to-one mapping between form and function in prosody, although for administrative convenience many researchers often act as if there were. Strong correlations certainly exist, for instance between certain kinds of pitch movement and the presence or absence of syntactic closure, but if we know one thing about prosodic function, it is that its relationship to prosodic form is highly complex and to a considerable degree context-dependent.

This symposium is not a theoretical treatment of prosody from either single

perspective, however; it is a discussion about prosody in situations of communication. The complexity of the relationship between form and function implies matching complexity in the prosodic processing which speakers and listeners perform in the course of communicating. In the following section I review some of the considerable recent literature on the interaction of perceptual and production processes, with emphasis on the perception and production of prosody.

The complexity of prosodic structure, and the necessity for hierarchical structural descriptions, is a recurring theme also in the five other contributions to the present symposium. In this introductory paper, I have chosen to follow two further themes which run through the symposium: the way prosodic structure can express relative *salience*, and the way it can communicate information about *segmentation*, at various levels of linguistic structure.

3. SITUATIONS OF COMMUNICATION

Let us define a situation of communication for our present purposes as a speaker speaking and a listener listening. (This is not to deny that there are many other kinds of communication, and some of them - sign language, for instance - certainly involve prosody.) The speaker's production processes and the listener's perceptual processes are obviously not independent, if only in the trivial sense in that one operates on the output of the other. However, there are some interesting further aspects of non-independence. Speech production processes can be actively constrained by characteristics of the perceptual process; and such effects can certainly be observed in the processing of prosody.

3.1 Perceptual constraints on production

As I have argued elsewhere [10], speakers' choices in production are often quite obviously constrained by the needs of listeners. This happens even at what one might consider quite low levels. For instance, why are the utterances of a speaker with a pipe clenched between the teeth not incomprehensible? If the processes of production were to run their normal course, the output might be considerably distorted; instead, adjustments occur (see e.g. [26]), with the effect that the processes of perception are enabled to run their normal course. Similarly, consider the Lombard reflex [27]: when ambient noise increases, speakers involuntarily speak more loudly. Interestingly, speakers in this situation adjust the individual formant frequences of their speech to compensate for the spectral characteristics of the noise [31]. The result, once again, is that the output sounds as close to the speaker's normal output as is possible.

At a slightly higher level, we see the same constraints operating on phonological processes of elision and assimilation. The process of palatalisation, whereby an alveolar stop and a following palatal glide become affricated, can apply across a word boundary - thus *did you* becomes [dɪdʒu] - and the effect is obviously to obscure the onset of the post-boundary word. But as Cooper and Paccia-Cooper [6] have shown, palatalisation across a word boundary is significantly less likely if the post-boundary word is unpredictable - for instance, low frequency, or contrastively stressed. The effect of this is that the words which the listener most needs to hear are less likely to be obscured. Likewise, speakers making up nonce words prefer to choose affixes which leave the base word intact over affixes

which require stress shifts or vowel changes (so *dowagerish* is preferred to *dowagerial*; [8]); again, the effect is that listeners can make sense of the new word because it contains a known word unaltered within it.

It is unsurprising that effects of this kind are apparent also in the realm of prosody. The mis-stressing of words impairs word recognition most severely if the stress shift causes vowel quality changes [2, 14]; and when speakers make a slip of the tongue involving mis-stressing, they are most likely to correct it if a vowel was changed [9]. Furthermore, they are more likely to add contrastive stress to the correction if there is high contrast between the error and the intended word [25]. Thus both the frequency and urgency of error repair are directly correlated with the likelihood that the error will disrupt comprehension.

Likewise, the work of two contributors to this symposium has shown how well-attuned are the processes of accent placement to listeners' needs. Fowler and Housum [20] showed that deaccented productions of words in a story could function as better retrieval cues for listeners than the same words accented on first mention. We know that listeners hearing a story construct an overall representation of the story situation [3, 22]; Fowler and Housum speculated that deaccenting could function as a signal to listeners that the concept in question is *already in* the story representation. Thus on hearing a word which in the phonetic context is obviously deaccented, listeners automatically access the already-constructed representation; for this reason such words function particularly effectively as retrieval cues. Similarly, Terken and Nooteboom [32] found that true-false decisions could be made more rapidly if new sentence subjects were accented but previously mentioned subjects were deaccented.

3.2 The Role of Speaker Awareness

Speakers' choices when they are deliberately trying to make themselves clear are also well attuned to listeners' needs. When marking word boundaries, for instance with a pause, speakers pay most attention to marking exactly the boundaries which listeners most often overlook, i.e. boundaries before weak syllables [13]. When trying to make syntax explicit, they add syntactic markers such as relative pronouns and complementisers [33], the presence of which, as perceptual research (e.g. [21]) has shown, makes syntactic processing significantly easier.

Prosody can be consciously used by speakers who are trying to be clear. Thus speakers who realise a listener is having difficulty understanding tend to speak more slowly, louder, and with raised pitch [5]. One communication situation in which this very noticeably happens is when an adult is talking to a child. A recent study by Fernald [19] has shown how effectively prosody can be used in this way. Fernald recorded the same mothers talking either to their infant child or to their husband, in specific types of interaction: expressing approval, attracting attention, giving solace etc. She then filtered all the recorded utterances and asked listeners to identify the type of interaction involved in each. The listeners' choices corresponded with the original context to a significantly greater degree for the infant-directed utterances than for the adult-directed utterances. Since the filtering process had left nothing in the speech signal intact except for the prosody, it would seem that, as Fernald concluded, speech to infants is more heavily loaded than speech to adults on prosodic signals of interactive intent.

In most speech situations, however, speakers are not making deliberate efforts to speak clearly. And as

Lehiste showed in a classic study [23], the availability of prosodic cues which will be of use to listeners may depend crucially on speaker awareness of potential problems for the listener. Prosody can in many cases very effectively signal which of two alternative syntactic parses is intended, for instance for syntactic ambiguities such as *The German teachers attended a meeting*, or *She hit the man with the stick* (see, e.g., [30]). In Lehiste's experiment, speakers read out a number of sentences, some of which were syntactically ambiguous; Lehiste then ascertained whether or not the speakers had been aware of the ambiguity, and which interpretation they had intended in their reading. The speakers then produced the sentences twice more, consciously intending each of the two different interpretations. All the versions were then played to listeners, who, Lehiste found, could much more accurately judge which interpretation had been intended in the versions produced with awareness of the ambiguity. Where the speaker had been unaware of the ambiguity, in fact, the listener judgements were often at chance.

3.3 The Speaker-Listener Contract

We can use the term *speaker-listener contract* to signify the proposal that participants in spoken communication have a shared goal: maximising the probability of successful message transmission. As the above review suggests, prosody is as much involved as any other aspect of linguistic structure in speakers' efforts to do their part in achieving this goal. The evidence reviewed included contrastive stress on error corrections; deaccenting of previously mentioned referents; and explicit cues to speech segmentation at the word and the phrase level. Thus both salience and segmentation figure in prosodic contributions to realisation of the speaker-listener contract.

4. SALIENCE

In a language which has sentence accent, listeners accord a high priority to the task of detecting where accent falls in a speaker's utterance. Prosodic cues are exploited to enable listeners to direct attention to the location of sentence accent [7]. If part of the normally available prosodic information is absent, listeners will exploit what remains [15]; but it seems that no one prosodic dimension is paramount in signalling accent location, because conflict between different sources of prosodic information (e.g. rhythm and pitch) leaves listeners unable to predict where accent will occur [11]. The importance of seeking accent location is explained as a search for focussed, or semantically central, aspects of the speaker's message [16].

The processing advantage enjoyed by accented words does not of course imply that if every word in an utterance were to be accented, the listener could process the entire utterance at a faster rate. Salience is necessarily a relative concept. As the work of Fowler and Terken, cited above, has conclusively shown, appropriate deaccenting is just as informative, and just as important, as accent.

In this symposium the contributions of Fowler, Ladd and Terken all make a further contribution to our understanding of the phonology and processing of sentence accent. As Ladd argues, relative salience expresses a syntagmatic relationship (between nodes in a metrical tree, in the metrical notation which Ladd uses), which co-exists with paradigmatic category distinctions between levels of accent (or, in Ladd's terms, levels of sentence stress). Ladd's intention in making this proposal is to reconcile apparently conflicting views of stress: on the one hand, the consensus of

contemporary phonologists that stress is an abstract relational construct, and on the other, the paradigmatic approach whereby stress is a property of syllables, which has proven persistently useful to non-phonologists (such as syntacticians and of course psycholinguists).

The role of relational structure in the expression of salience is also central to Terken's contribution, which focusses on the way in which the processes of speech production translate such relational structure into relative acoustic salience (in the fundamental frequency contour, in this instance), and the way in which the processes of speech perception interpret fundamental frequency variation as information about relative salience.

Fowler and Levy extend our understanding of how relative salience in a context finds expression in linguistic output by drawing parallels between lengthening and shortening effects in both prosodic and lexical forms. Unpredictable topics are referred to by longer expressions, and/or the words expressing them are realised with greater duration. The effect is to provide listeners with more speech evidence for less predictable concepts. This is powerful evidence for the operation of the speaker-listener contract at multiple levels of linguistic structure.

5. SEGMENTATION

Segmentation is one of the listener's major tasks; boundaries must be identified between units at several linguistic levels. Firstly, the continuity of the speech signal results in very few reliable cues to word boundaries being realised; listeners therefore have to exploit whatever sources of information they can to work out how speech signal divide up into individual words. Secondly, listeners must group words into phrases, that is, they must detect syntactic boundaries. Thirdly, they must identify larger units of semantic structure, sometimes referred to as topic structure [4], or paragraph structure [24]. And fourthly, they must perceive structure at the interactional level, i.e. speaker turns.

Prosody contributes to the listener's performance of all these segmentation tasks. At the lexical segmentation level, listeners can exploit their linguistic experience to develop heuristic segmentation procedures based on where word boundaries are most likely to occur in their native language; in English, I have argued, such procedures are based on the predominance of strong initial syllables in the vocabulary [12]. At the syntactic level, as was discussed above, prosodic cues to boundaries are readily exploited by listeners [23, 29, 30].

In comparison with the quite large amount of research on lexical juncture, and yet larger body of work on syntactic boundaries, segmentation of discourse into topic or paragraph units has received relatively little attention. (Three studies in the early 1980s should be mentioned: Brown, Currie and Kenworthy [4] reported that speakers tended to raise the pitch of their speech when introducing a new topic; Menn and Boyce [28] reported the same finding in parents' conversations with children. Lehiste [24] analysed the average duration of phonetic segments and words in non-final, phrase-final and paragraph-final position; she found both phrase-final and (somewhat greater) paragraph-final lengthening.) It is therefore timely that the contribution to this symposium by Bruce describes an ongoing project which has as one of its principal aims the investigation of prosodic cues to segmentation at this level of linguistic structure.

Segmentation of conversation into participant turns is, finally, addressed in this symposium by Couper-Kuhlen. The literature on prosodic cues to turn-taking has been bedevilled by confusion between the speaker and listener perspectives; Duncan [18], for instance, isolates several prosodic characteristics of speakers' turn-final uttterances and terms them "cues" without, however, any evidence that listeners actually use them as such (see Cutler and Pearson [17] for a critique). Couper-Kuhlen reports evidence that co-operative rhythmic synchronisation of speech occurs in smooth turn-taking; in this study the listeners' reception of speakers' signals is attested by the synchronisation of the initial rhythmic intervals of the new turn (produced by the listener-turned-speaker) with the final rhythmic intervals of the old turn produced by the previous speaker. Like the other contributors, Couper-Kuhlen also highlights the importance of hierarchical structure in prosody, such structure being fundamental to the turn-taking metric which she proposes.

6. CONCLUSION

It is no surprise to find that salience and segmentation form unifying themes for contributions to a symposium on prosody. According to Bolinger [1], these (or, in his words, obtrusions for prominence and the expression of closure) are the two major language-universal uses of prosody. In situations of communication, much of speakers' and listeners' prosodic processing is devoted to these goals.

One thing to note about the importance of prosodic segmentation cues is that it mirrors the importance of segmentation in orthographic representations - lexical segmentation is explicitly coded in nearly all orthographies, and syntactic segmentation in most; higher-level segmentation is likewise signalled by textual devices. As this review has tried to show, and as the symposium will further stress, both salience and segmentation are central to successful communication, and prosody is thus central to linguistic structure.

7. REFERENCES

[1] Bolinger, D.L. (1978), "Intonation across languages", in J.H. Greenberg (Ed.) *Universals of Human Language*, Stanford: Stanford University Press.

[2] Bond, Z.S. & Small, L.H. (1983), "Voicing, vowel and stress mispronunciations in continuous speech", *Perception & Psychophysics, 34*, 470-474.

[3] Bransford, J. & Franks, J. (1971), "The abstraction of linguistic ideas", *Cognitive Psychology, 3*, 331-350.

[4] Brown, G., Currie, K.L. & Kenworthy, J. (1980), *Questions of Intonation*, London: Croom Helm.

[5] Clark, J.E., Lubker, J. & Hunnicutt, S. (1988), "Some preliminary evidence for phonetic adjustment strategies in communication difficulty", in R. Steele & T. Threadgold (Eds.) *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam: J. Benjamins.

[6] Cooper, W.E. & Paccia-Cooper, J. (1980), *Syntax and Speech*, Cambridge, MA: Harvard University Press.

[7] Cutler, A. (1976), "Phoneme-monitoring reaction time as a function of preceding intonation contour", *Perception & Psychophysics, 20*, 55-60.

[8] Cutler, A. (1980), "Productivity in word formation", *Papers from the Sixteenth Regional Meeting, Chicago Linguistic Society*, 45-51.

[9] Cutler, A. (1983), "Speakers' conceptions of the functions of prosody", in A. Cutler & D.R. Ladd (Eds.) *Prosody: Models and Measurements*, Heidelberg: Springer.

[10] Cutler, A. (1987), "Speaking for listening", in A. Allport, D.G. MacKay, W. Prinz & E. Scheerer (Eds.) *Language Perception and Production: Relationships between Listening, Speaking, Reading and Writing*, London: Academic Press.

[11] Cutler, A. (1987), "Components of prosodic effects in speech recognition",

Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia; Vol. 1, 84-87.
[12] Cutler, A. (1990), "Exploiting prosodic probabilities in speech segmentation", in G. Altmann (Ed.) *Cognitive Models of Speech Processing,* Cambridge, MA: MIT Press.
[13] Cutler, A. & Butterfield, S. (1990), "Durational cues to word boundaries in clear speech", *Speech Communication, 9,* 485-495.
[14] Cutler, A. & Clifton, C.E. (1984), "The use of prosodic information in word recognition", in H. Bouma & D.G. Bouwhuis (Eds.) *Attention and Performance X: Control of Language Processes,* Hillsdale, NJ: Erlbaum.
[15] Cutler, A. & Darwin, C.J. (1981), "Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency", *Perception & Psychophysics, 29,* 217-224.
[16] Cutler, A., and Fodor, J.A. (1979), "Semantic focus and sentence comprehension", *Cognition, 7,* 49-59.
[17] Cutler, A. & Pearson, M. (1985), "On the analysis of prosodic turn-taking cues", in C. Johns-Lewis (Ed.) *Intonation in Discourse,* London: Croom Helm.
[18] Duncan, S. (1973), "Toward a grammar for dyadic conversation", *Semiotica, 9,* 29-47.
[19] Fernald, A. (1989), "Intonation and communicative intent in mothers' speech to infants: Is the melody the message?", *Child Development, 60,* 1497-1510.
[20] Fowler, C.A. & Housum, J. (1987), "Talkers' signalling of 'new' and 'old' words in speech, and listeners' perception and use of the distinction", *Journal of Memory & Language, 26,* 489-504.
[21] Hakes, D.T. (1972), "Effects of reducing complement constructions on sentence comprehension", *Journal of Verbal Learning & Verbal Behavior, 11,* 278-286.
[22] Haviland, S.E. & Clark, H.H. (1974), "What's new? Acquiring new information as a process in comprehension", *Journal of Verbal Learning & Verbal Behavior, 13,* 512-521.
[23] Lehiste, I. (1973), "Phonetic disambiguation of syntactic ambiguity", *Glossa, 7,* 107-122.
[24] Lehiste, I. (1980), "Phonetic characteristics of discourse", *Transactions of the Committee on Speech, Acoustical Society of Japan, 4,* 25-38.
[25] Levelt, W. & Cutler, A. (1983), "Prosodic marking in speech repair", *Journal of Semantics, 2,* 205-217.
[26] Lindblom, B., Lubker, J. & Gay, T. (1979), "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", *Journal of Phonetics, 7,* 147-162.
[27] Lombard, E. (1911), "Le signe de l'élévation de la voix", *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx, 37,* 101-119.
[28] Menn, L. & Boyce, S. (1982), "Fundamental frequency and discourse structure", *Language & Speech, 25,* 341-383.
[29] Scott, D.R. (1982), "Duration as a cue to the perception of a phrase boundary", *Journal of the Acoustical Society of America, 71,* 996-1007.
[30] Streeter, L.A. (1978), "Acoustic determinants of phrase boundary perception", *Journal of the Acoustical Society of America, 64,* 1582-1592.
[31] Summers, W.V., Pisoni, D.B., Bernacki, R., Pedlow, R. & Stokes, M. (1988), "Effects of noise on speech production: Acoustic and perceptual analyses", *Journal of the Acoustical Society of America, 84,* 917-928.
[32] Terken, J. & Nooteboom, S.G. (1987), "Opposite effects of accentuation and deaccentuation on verification latencies for given and new information", *Language & Cognitive Processes, 2,* 145-163.
[33] Valian, V.V. & Wales, R.J. (1976), "What's what: talkers help listeners hear and understand by clarifying syntactic relations", *Cognition, 4,* 115-176.