12

From performance to phonology: comments on Beckman and Edwards's paper

ANNE CUTLER

Beckman and Edwards have presented experimental evidence for the association of syllabic lengthening in speech with two phonological effects. One is the presence of an intonational phrase boundary. Lengthening associated with intonational phrase boundaries occurs consistently across speakers and speech rates. The other effect is the presence of a word boundary. In contrast to phrase-final lengthening, word-final lengthening occurs inconsistently. It is more evident at a slow rate of speech; even then, not all speakers show it with all sentences.

Beckman and Edwards's aim in examining the occurrence of lengthening effects is to make claims about the proper model of English phonological structure. They relate their experiments to a long tradition of phonological interpretation of timing effects in speech, with two separate (and as Beckman and Edwards point out, at least in part incompatible) strands: lengthening which is claimed to accompany the terminal boundary of a phonological unit, and shortening which is claimed to adjust the interval between stressed syllables in the direction of greater regularity. The effects which Beckman and Edwards found in their experiments are of the first sort.

This commentary concerns the assumptions on which undertakings such as Beckman and Edwards's are based. Precisely how does a speaker's performance shed light on the phonology of a language?

Two useful distinctions may be drawn. The first is between what one may call a strong and a weak version of the claim to "psychological reality." In the strong version, the claim that a particular level of linguistic analysis X or postulated process Y is psychologically real implies that the ultimately correct model of human language processing will include a level of representation corresponding to X or a mental operation corresponding to Y. In the weak version, "psychological reality" implies only that language users can draw on knowledge of their language which is accurately captured by the linguistic generalization in question. (Note that for certain linguistic constructs this latter claim is equivalent to no more than the construct's descriptive adequacy; for example, the intuitions which language

users are predicted to entertain by the weak reading of "psychological reality of the phoneme" are the same distributional data which led to the postulation of the construct phoneme in the first place.)

The second useful distinction is between the processes of language production and language perception; or rather, between the relevance of phonological structure to production and to perception respectively. Crudely speaking, one might think of phonological structure exercising a particular effect on language production because the structure is relevant in some way (strongly or weakly) to the process of language production. Alternatively, one might imagine the same effect appearing in production simply because speakers want to make sure that the structure is correctly interpreted by hearers. In this case the presence of the effect in question can be said to be constrained by the nature of the process of perception rather than the process of production, in the sense that the speaker's behavior is governed by a model of the hearer. Some examples should make this second distinction clearer.

Production as a motivation

Cooper and Paccia-Cooper (1980) carried out one of the most extensive investigations of temporal (and other phonological) effects associated with syntactic structure. Their stated aim was "to make use of temporal phenomena [to] arrive at the form of a speaker's grammatical code" (p.9). Among other effects, they found much evidence of segmental lengthening at syntactic boundaries; their preferred explanation of this phenomenon is that "lengthening effects are produced in order to allow the speaker an extra fraction of time for planning upcoming material in the next phrase. This planning may involve not only semantic and syntactic computations but also the resetting of ... articulatory postures" (p. 199). Since this explanation cannot account for why lengthening also occurs in utterance-final position, Cooper and Paccia-Cooper offer as a subsidiary explanation the possibility that "The lengthening effect could be attributable to execution ... [and] represents a general relaxation response ... The internal clock, in effect, runs more slowly at the ends of major constituents, presumably due to processing fatigue" (p. 199). Cooper and Paccia-Cooper point to some independent supporting evidence for each of these claims, and discuss and dismiss alternative explanations.

Each of these mechanisms whereby segmental lengthening effects might arise is purely internal to the process of speech production. The syntactic structure around which Cooper and Paccia-Cooper built their investigations is, according to their explanations, the framework within which speech is produced. Speakers speak clause by clause and phrase by phrase; when they need extra time for planning the next unit of their utterance, or when they reach the end of a planning unit and are able to relax, the unit in question is a syntactic one.

Although Cooper and Paccia-Cooper were concerned in their investigations

exclusively with syntactic structure, and couched their production-based account in syntactic terms, there is of course no reason why the argument could not equally well be applied, *mutatis mutandis*, to phonological structure. Speakers speak phonological phrase by phonological phrase, require planning time prior to the next phonological phrase, relax at the end of a phonological phrase.

Perception as a motivation

Lehiste (1977) made a powerful argument for the perceptual origins of lengthening effects in speech. She reviewed evidence showing that listeners judge speech to be more rhythmically regular than it actually is, and, in effect, "expect isochrony" (p. 253). Speakers can trade on this expectation by using manipulation of speech rhythm to communicate information to listeners. For instance, they can signal the presence of a syntactic boundary by lengthening the interval between two stress beats on either side of the boundary. Scott (1982) provided confirming evidence that listeners' use of lengthening as a cue to the presence of a phrase boundary is not, in Beckman and Edwards's words, a direct decoding of syntax but is mediated by phonological structure in the form of inter-stress interval ratios. It is suggested, then, that lengthening associated with syntactic structure is produced by the speaker not for any production-internal reason such as the need for planning time or the relaxation of articulatory gestures, but for the benefit of the listener's perception of structure within the utterance.

Note that invoking perceptual motivation for lengthening effects in speech production is not a particularly radical claim. Production involves a range of phenomena in which the choices by speakers are constrained by the needs of listeners (see Cutler 1987 for a review of some of these). The construction of a neologism, the way in which a slip of the tongue is corrected (and whether it is corrected at all), the choice of whether or not to obscure a word boundary by allowing the application of elision or assimilation – all of these production processes are constrained ultimately by the requirements of the perceiver. It is surely no more radical to claim that speakers should attempt to make phonological structure clear to their listeners.

Comparing production and perception arguments

Naturally, independent evidence can be invoked to motivate either a production-based or a perception-based account of speech performance phenomena. This was the approach taken by Lehiste (1977), who first reviewed the considerable independent evidence for listeners' over-estimation of rhythmic regularity before constructing her argument about the perceptual motivation for lengthening in production. But it should be noted that there is also independent evidence for the importance of rhythmic regularity in speech production, beyond the very existence

of lengthening/shortening effects in normal utterances. For instance, it is a startling fact that slips of the tongue which alter the number of syllables in an utterance (such as omission or addition of syllables) significantly more often result in an utterance more rhythmic than the target utterance would otherwise have been (Cutler 1980). This suggests that there may be a production advantage to regularity of speech rhythm. Here, something which the speaker did *not* want to say can be viewed as evidence of a speech production effect which is independent of the process of constructing an utterance, to the extent that pressure towards rhythmicity can on occasion interfere with the process of construction.

However, one can also invoke evidence which is internal to the phenomenon at issue in order to distinguish between the relative merits of production-based and perception-based accounts. For instance, it may seem obvious that performance effects which have their origin in some characteristic of the production process should necessarily show up in all utterances where they are applicable. Performance effects which result from the speaker's intention to produce perceptual effects in the listener, on the other hand, might well be subject to the speaker's option – for instance, their presence might be argued to be contingent upon the speaker's recognition of the listener's needs.

One of the most relevant experiments here is a study by Lehiste (1973), which dealt with syntactically ambiguous sentences. At issue was whether speakers would produce disambiguating cues (temporal cues, among others) which would enable listeners to distinguish between the various readings of an ambiguous utterance. Crucially, Lehiste told her set of speakers that all the sentences were ambiguous only *after* they had spoken the sentences. Then she (a) recorded which reading the speakers had intended in this first production, and (b) had the speakers produce each sentence again in each of its readings. All productions were then played to listeners who were asked to judge which reading had been intended. The listeners were often unable to tell which reading had been intended when the speakers were unaware of ambiguity. When the speakers were aware of the ambiguity and were attempting to signal one reading rather than another, however, their signals were generally distinguishable. Lehiste's physical measurements of the various productions confirmed that the consciously disambiguated utterances contained marked temporal and other cues to syntactic structure which were absent in the first, unaware, productions.

Lehiste's study suggests that disambiguating prosody (of which in her results a very large component was temporal manipulation) is indeed to a considerable extent under the control of speaker awareness, which in turn suggests that it is not produced by factors internal to the production process.

Strictly speaking, the mere absence of a particular effect on a particular occasion is not by itself evidence against a production-based account of that effect. Floor effects could for instance result in a temporal effect becoming so small as to be insignificant at fast rates of speech. Alternatively, an independent component of

the speech production process could interact in an unknown manner with the component producing the temporal effect in such a way that the effect came and went. However, if no such competing process can be independently motivated, and if rate of speech is explicitly controlled, then consistency versus inconsistency of a particular temporal effect is a useful guideline to its genesis.

In terms of the strong versus weak distinction of psychological reality, a perception-based account implies only weak psychological reality, in the sense that speakers could not produce phonological distinctions for listeners unless they had a mental representation of the phonological structure available to them. A production-based account generally implies strong psychological reality; for example, Cooper and Paccia-Cooper's account is based on the assumption that speech is planned in chunks which are syntactically defined, so that the syntactic structure is present as a level of representation in the process of speech production. It is conceivable, however, that a production-based account could entail only weak psychological reality – for instance, if it only required reference to a mental representation of phonological structure itself independent of the production process, rather than requiring that phonological structure be a necessary component of one or more stages of processing.

Assumptions underlying Beckman and Edwards's arguments

We come now to the present argument from performance to phonology. Beckman and Edwards are interested in the proper model of English prosodic phonology, and the way in which temporal effects in speech performance might shed light on this. Precisely what assumptions underlie their conception of the relationship between the phonology and the performance?

Beckman and Edwards do not explicitly state the reasons why they consider that performance sheds light on phonology. However, there are certain clues to their assumptions in their text. Let us concentrate on the case of word-final lengthening, which is, after all, both the major new contribution of Beckman and Edwards' paper and the primary puzzle which is still left at its end. On the one hand, Beckman and Edwards speak of word-final lengthening as being a "boundary mark" (p. 164), which "marks accentual phrases" (p. 170). All of these wordings suggest that Beckman and Edwards consider word-final lengthening, at least, to be under the control of speaker option, and to be used for marking structural boundaries for the sake of the listener. On the other hand, Beckman and Edwards also speak of a constituent "triggering" word-final lengthening (p. 170), which suggests a strong production-based motivation.

In fact, the results of the various experiments support different conclusions about word-final lengthening. In experiments 1 and 2 the word-final effect is robust and reliable, which would at least be consistent with a production-based explanation. In experiment 3, however, only two subjects show simple word-final

lengthening. This is consistent with an explanation in which presence of the effect depends upon subjects' awareness of the structure with which the effect is associated.

As Beckman and Edwards are aware, however, this state of affairs is highly unsatisfactory. Words *per se* are not necessarily phonological units. Moreover, it is beyond question that literate speakers are aware of words as units of their utterances. Thus there is a distinct possibility that the word-final lengthening effect is simply a direct analogue encoding of a non-phonological effect. This kind of explanation is what Beckman and Edwards are confessedly biased against, and it is understandable that they expend considerable effort on resisting it. Furthermore, since natural speech tends to contain a majority of monosyllabic words (Cutler and Carter, 1987), such an explanation would suggest that perhaps the majority of temporal variation above the segmental level relates only to lexical and not at all to phonological units.

Because the primary problem for an interpretation of the word-final effect is the discrepancy between the results of experiments 1 and 2, on the one hand, and experiment 3, on the other, Beckman and Edwards attempt to reach a solution by looking for a crucial difference between the experiments. They themselves point out that their chosen solution, a possible difference in focus structure between the sets of materials in the earlier and later experiments, is both highly speculative and unable to account fully for all aspects of their results. However, there are further characteristics of the particular experimental materials Beckman and Edwards used which may also prove to be relevant.

Consider first of all the question of subjects' awareness of structure. The Lehiste (1973) experiment showed that bringing a contrast to speakers' attention heightens the likelihood that the contrast will be marked in performance. The sentences used in all of Beckman and Edwards's experiments left speakers in no doubt of the contrasts under investigation. One cannot therefore rule out the possibility that the temporal distinctions they produced arose from their desire to emphasise the perceived contrast. The opposition was of course clearest in the pairs of sentences used in experiments 1 and 2, in which the same phonetic sequence received different lexical/syntactic parses.

Consider further the phonetic structure of these sentences. Speakers were required to articulate three /p/ phonemes, each followed by a different vowel, in the crucial region to be measured. It is not particularly far-fetched to suggest that the difficulty of articulating such a sequence with clarity might have been a contributory factor in rendering speakers more prone to produce lengthening effects. Indeed, research on tongue twisters suggests that lengthening (that is, slowing of articulation) is a characteristic response of speakers faced with an articulatorily awkward output (Kupin 1976). Significantly, in experiment 3, in which the crucial regions of the sentences contained no sequences of repeated phonemes, far less word-final lengthening was observed.

Suppose that instead of "Pop" and "Poppa" preceding the "opposed"– "posed" contrast, Beckman and Edwards had used, for example, "Dad" and "Dadda" (or perhaps, to avoid all repetitions, "Beck" and "Becca"). Would speakers have produced the word-final effects quite so consistently?

These speculations are clearly testable. Firstly, one could use materials involving less phonetic repetition. Independently, one could investigate the effect in a situation in which subjects were as far as possible unaware of the contrast to be tested.

If the word-final lengthening effect were to disappear as a result of these manipulations, one need look no more for a phonological explanation of it. If it should persist, then it is appropriate to continue the attempt to sort out whether or not it is a true phonological effect, and, orthogonally, whether it reflects characteristics of the process of speech production.

My personal bet is that the word-final lengthening effect is both non-phonological and at least partly artifactual. But I hope I am wrong. The most interesting kind of claim about speech performance effects is the strong implementation claim, since it has implications both for linguistic and psychological theory. There is precious little precedent for claims about the role of phonological structure in speech production; I wish Beckman and Edwards had been able to advance unassailable evidence in support of such a claim.

References

Beckman M. E. and J. Edwards (this volume) Lengthenings and shortenings and the nature of prosodic constituency.

Cooper W. E. and J. Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Cutler, A. 1980. Syllable omission errors and isochrony. In H. W. Dechert and M. Raupach (eds.) *Temporal Variables in Speech*. The Hague: Mouton, 183–190.

1987. Speaking for listening. In A. Allport, D. G. MacKay, W. Prinz and E. Scheerer (eds.) *Language Perception and Production: Relationships between Listening, Speaking, Reading and Writing*. London: Academic Press, 23–40.

Cutler, A. and D. M. Carter, 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2: 133–142.

Kupin, J. 1976. Tongue twisters. Paper presented to the Linguistic Society of America, Philadelphia, December 28–30.

Lehiste, I. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa* 7: 107–122.

1977. Isochrony reconsidered. *Journal of Phonetics* 5: 253–263.

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.