Laboratory Phonology 7

edited by
Carlos Gussenhoven
Natasha Warner

Phonological Processing:
Comments on Pierrehumbert, Moates et al., Kubozono, Peperkamp & Dupoux, and Bradlow

Anne Cutler

1. Introduction

Processing is a very general term. Preceded by *phonological* its domain of reference can be located by presumption in the realm of language (if only because the phonological processing of food, or of passport applications, seems rather improbable). But it remains an unsatisfyingly ambiguous expression, allowing either an interpretation in which information of a phonological nature is processed (cf. mathematical instruction), or one in which unspecified information is processed via the application of phonology (cf. mathematical reasoning). Perhaps for this reason psycholinguists, though they are excessively fond of the general term processing, do not standardly use phonological processing to refer to any aspect of their models.

Processing in psycholinguistics covers all mental operations involved in the use of language (and most particularly in listening and speaking). The five papers commented on here all refer in some way to evidence from language performance, and hence imply such operations. But they involve processing of several different kinds: production of words (Pierrehumbert), production of phonemes (Bradlow), perception of words (Moates, Bond and Stockmal), perception of vowels (Kubozono) and perception of stress (Peperkamp and Dupoux).

An introductory overview of the consensus model of word and Phoneme production in psycholinguistics is contained in the chapter by Levelt (this volume; for further detail see Levelt, Roelofs & Meyer, 1999). It would be redundant to recapitulate the summary

here. The following section offers a matching overview of the consensus model of listening to spoken language (with special reference to the processing levels just mentioned: words, phonemes, stress).

The evidence on which the modelling efforts are based — in production as well as in comprehension research — has been collected in most part from behavioural experiments in the psycholinguistic laboratory. In more than three decades of spoken-language comprehension research, an extensive arsenal of empirical techniques has been developed (see Grosjean & Frauenfelder, 1996, for a review). In some of these techniques task performance is assessed qualitatively, but the majority involve measurement of response time (RT), i.e. the latency with which listeners can perform some simple operation such as repeating a heard input, making a binary decision about it, or signalling detection of a target. RT measures represent the best attempts to investigate processing "on-line", i.e. in the course of its operation. In psycholinguistics, on-line measures are often preferred over less direct measures of performance.

2.   Spoken-word recognition

Listening to spoken language involves recognising, in the incoming speech signal, discrete portions which correspond to stored representations in the listener's lexicon. Several facts about spoken-word recognition make it a challenging research area. First, the process takes place in time - words are not heard all at once, but from beginning to end. Second, words are rarely heard in isolation, but rather within longer utterances, and there is no reliable equivalent in speech of the spaces which demarcate individual words in a printed text. Thus listening to speech necessarily involves segmentation, i.e. division of the continuous input into the portions corresponding to individual words. Third, spoken tokens of individual words are highly variable, because speakers' voices differ greatly and background noise and other aspects of the listening situation can affect intelligibility. And fourth, spoken words are not highly distinctive; language vocabularies of tens of thousands

of words are constructed from a repertoire of on average only 30 to 40 phonemes (Maddieson, 1984). As a consequence words tend to resemble other words, and may have other words embedded within them (thus *great* contains possible pronunciations of *grey* and *rate* and *eight,* it resembles *grape* and *crate* and *greet,* it occurs embedded within possible pronunciations of *migrate* or *grating* or even *league rating,* and so on). How do listeners know when to recognise *great* and when not?

All current models of spoken-word recognition assume that whatever words are supported by the speech signal, irrespective of whether such support is intended by the speaker, may become active in the listener's recognition system. There is now abundant experimental evidence indicating that words may become activated when they are embedded within other words (such as *grey* in *grating;* Cluff & Luce, 1990; Shillcock, 1990; Gow & Gordon, 1995), or when they are spuriously present across two other words (such as *great* in *league rating;* Tabossi, Burani & Scott, 1995), and that partially overlapping words *{grey, great, grating)* may become simultaneously active (Zwitserlood, 1989; Marslen-Wilson, 1990; Connine, Blasko & Wang, 1994; Vitevich & Luce, 1998; Soto-Faraco, Sebastian-Galles & Cutler, 2001). Recognition then ensues after a process of competition between the activated candidate words. Again, there is empirical evidence which solidly supports the contribution of competition in word recognition (Goldinger, Luce & Pisoni, 1989; McQueen, Norris & Cutler, 1994; Norris, McQueen & Cutler, 1995; Vroomen & De Gelder, 1995; Soto-Faraco et al., 2001).

Since TRACE (McClelland & Elman, 1986a), models of spoken-word recognition have been computationally implemented, allowing explicit simulation of experimental findings. Of all such models, only Shortlist (Norris, 1994) currently allows simulations with a realistically sized dictionary of tens of thousands of words. In Shortlist, competition involves lateral inhibition between simultaneously active candidates for any part of the speech signal. The more active a candidate word is, the more it may inhibit activation of its competitors. Activated and competing words need not be aligned with one another, and the competition process in conse-

quence offers a potential solution to the segmentation problem. Thus although the recognition of *league rating* may involve competition from *grey, great* and *eight,* this will eventually be overcome by joint inhibition from *league* and *rating.*

Models of spoken-word recognition differ on a number of dimensions, most notably on whether they allow bidirectional or only unidirectional flow of information between processing levels. However, there is widespread agreement on the above outline architecture. In the following section the models are compared with regard to the role they allow to phonological constructs.

3.   Phonology in spoken-word recognition

Do phonological constructs define entities which play a necessary role in models of perceptual processing? The answer to this question must be no. Years of psycholinguistic research have been invested in examination of whether various phonological constructs function as "perceptual units", but, as I have argued in a previous contribution to the Laboratory Phonology series and elsewhere (Cutler, 1992a, 1992b), the questions asked by spoken-word recognition researchers have not allowed for answers which might be useful to phonology. In the modelling framework sketched above, the phoneme can certainly be said to have a vital role, insofar as the phoneme is by definition the minimal sequential unit of distinction between two words. Each such minimal distinction between two simultaneously active lexical candidates will influence the process of competition between them. But such influence does not depend on an explicit representation of the phoneme as the means by which the distinction is achieved.

Some current computational models (e.g. TRACE, Shortlist) in fact operate with explicit representations of phonemes; but this is always described as a computational convenience rather than an inherent component of the model. Evidence that listeners effectively exploit coarticulatory cues to upcoming phonemes (e.g. Whalen, 1984, 1991; Streeter & Nigro, 1979; Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 1999) has been interpreted as arguing against an explicit role for phonemes in the lexical

access process (e.g. Marslen-Wilson & Warren, 1994), although in reality it is, like the fact that a minimal difference between two words is by definition a difference of one phoneme, neutral with respect to the representational issue.

Similarly, episodic models of word recognition (e.g. Goldinger, 1998) are often held to be incompatible with the notion of obligatory intermediate representations in terms of phonemes or other units; but they too are in fact neutral on this issue. It is true that they offer a framework which can be realised without such representations; but so do non-episodic models. In short, the search for evidence which will settle the issue of "perceptual units" may be a hopeless quest; many psycholinguists have indeed abandoned it as such.

The performance evidence discussed in the papers to which these comments refer correspondingly does not provide evidence concerning the role of phonological constructs in the perception model (or, for that matter, in the production model discussed by Levelt). However, the papers do to a considerable extent address questions which have more to do with processing (and how it should be modelled) than with phonological structure (and its role in the grammar). In particular conjunctions of subsets of the five papers raise a number of interesting questions of importance for the processing model. Five of these questions, of varying degrees of granularity, will be considered in part 4 below; they concern respectively the relation between perception and production, the flexibility of the language processing system, the relative contribution of vowels and consonants in spoken-word recognition, the processing implications of phoneme coarticulation, and the position of language-specificity in the processing system.

4.   Processing questions from a phonological perspective

4.1 How close is the relation between perception and production?

Listeners are also speakers; barring impairment, any language user's processing of spoken language includes both perception and

production. For this reason alone perception and production must to a substantial extent be considered together. Production is ultimately very dependent on perception, in that we speak the words and structures we have heard. On the other hand, the translation of sound to meaning and of meaning to sound require different types of processing, which, as brain imaging and neurological impairment evidence attests, are to a considerable degree subserved by different mechanisms in the brain (Price, Indefrey & Van Turennout, 1999). Thus the closeness (or otherwise) of the relation between the language input and output systems is a regular object of study, and two traditional types of evidence concerning this issue are addressed in the present set of papers.

Evidence that speakers often cater to the needs of listeners can readily be found (see e.g. Cutler, 1987). The study of clear speech has provided a rich source of relevant data: deliberately or not, speakers adjust their clarity of articulation and other aspects of speaking style when listeners are in difficulty (Picheny, Durlach & Braida, 1985, 1986). Many adjustments they make would be difficult to bring under conscious control — for instance, adaptation of vowel formant structure to compensate for the formant masking in background noise (Van Summers, Pisoni, Bernacki, Pedlow & Stokes, 1988), or selective emphasis on the word boundaries which perceptual processing is most likely to overlook (Cutler & Butterfield, 1990). The vowel production measurements described by Bradlow surely belong in this class, although in her data clear articulation is applied to all segments irrespective of intrinsic inter-segment confusability within the vowel inventory of the language. The vowel space expansion in clear speech conditioned by listener difficulty parallels that recently reported for infant-directed speech (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg & Lacerda, 1997). These results suggest close attunement of production and perception systems.

Pierrehumbert also posits a close perception-production relation, though of a different kind. In her proposal, word production is based very directly on aggregated experience of word perception. In fact the evidence which she reviews does not actually include lexeme-specific effects of the kind that the proposal in principle

predicts — i.e. idiosyncratic properties of production associated with individual words; rather, she points to talker-repetition advantages in perception, systematic word frequency effects on articulation, and sociolinguistic effects (which are in a way also frequency effects, of a group-specific nature). For further discussion of this issue see Levelt (this volume); it will be interesting to see whether future studies will produce evidence for the truly lexeme-specific production effects which would support Pierrehumbert's position.

Another type of evidence often called upon in discussions of perception-production relations concerns neutralisation effects. If underlying distinctions are masked, i.e. neutralised, are speakers ignoring perceptual exigencies? If phonologically conditioned neutralisation in fact turns out to be incomplete, is this because perceptual needs have prevailed? So far, a simple conclusion for this literature has proved elusive; the picture is complex, with many sources of information contributing to the realisation or nonrealisation of a contrast, systematic differences being produced but below the level at which listeners can profit from them, and differences which listeners are well able to use being unreliably produced (see e.g. Warner, Jongman, Sereno & Kemps, submitted).

Kubozono describes a case of neutralisation which is remarkable in that information available to nonnative listeners is ignored by native listeners. That is, the neutralisation is itself part of the native user's system. This pattern in fact suggests that his finding belongs in the realm of effects which are captured by some other aspect of processing. A parameter of variation which provides information of one kind may thereby become unavailable for use as information of another kind; systematic durational cues to stress may be overlooked by listeners for whom duration is a cue to Phonemic quantity distinctions (Berinstein, 1979); systematic palatalisation variation as a function of syntactic structure may be overlooked by listeners for whom palatalisation is sociolinguistically informative (Scott & Cutler, 1984). It is thus not unreasonable to propose that the nonnative listeners in Kubozono's study may have been exploiting a type of information which for the native listeners was already captured by another function.

4.2 How flexible is the language processing system?

To what extent can the operation of the processing system be varied by the language user? This too has been a question which has long aroused research interest (indeed, some of the relevant research is mentioned under the topic of clear speech in section 4.1 above, or under coarticulatory effects in section 4.4 below).

The modulating effects of attention are at issue in contributions in this section. Attention is (like the larger issue of consciousness to which it is very closely connected) a notoriously elusive psychological concept; nonetheless, processing effects reasonably ascribed to attentional variation are widespread. Particularly the type of psycholinguistic experiment in which explicit phonemic decisions are required provides an appropriate environment for such effects to manifest themselves. Thus presence or absence of a secondary task encouraging attention to meaning (Dell & Newman, 1980), or simply varying the relative monotony of stimulus materials (Cutler, Mehler, Norris & Segui, 1987) both affect whether or not lexical characteristics influence listeners' response latencies in a phoneme detection experiment; this is explained as indicating reduced attention to lexical processing when it is not explicitly required for task performance, or even when words and nonwords hardly vary. Implicit direction of attention to target position is also possible in the same task (Pitt & Samuel, 1990; Pallier, Sebastian-Galles, Felguera, Christophe & Mehler, 1993). Direction of attention to target phoneme location also improves listener performance in distinguishing noise-masked versus noise-replaced phonemes (Samuel & Ressler, 1986).

Pierrehumbert (citing Johnson, 1997) sees a role for attention in explaining speaker normalisation effects in listening. The mental representation of a given word is the aggregate of perceptual episodes involving that word; however, if attention can play the role assigned to it, it is clear that individual episodes must be tagged for origin. Pierrehumbert suggests that listening can selectively refer incoming input from a given speaker to previous exemplars of the same speaker's production, or at least weight those exemplars more highly (it might seem that non-matching exemplars should

in fact be ruled out completely; the question of whether or not this is possible is not addressed). This would allow more accurate phonemic classifications, as occurs with speaker normalisation. Pierrehumbert also proposes a more nebulous effect in speech production, whereby selective attention to exemplars of a given speaker can lead to productions which more closely imitate that speaker. Most speakers are of course notoriously bad at imitating other individuals successfully, so the extent of this component of Pierrehumbert's proposal, and its precise role, need to be spelt out in greater detail. One potential function of a mechanism of this kind might be switches in speaking style and register.

A more straightforward role for attention in listening is proposed by Peperkamp and Dupoux. Although the long-term goal of their project is specification of how infants learn the phonology of a native language, their model also delivers predictions about adult listening. In a series of experiments following the earlier work of Dupoux, Pallier, Sebastian-Galles and Mehler (1997), they have examined the phenomenon that listeners can sometimes correctly perceive a nonnative contrast in a forced-choice discrimination task, but are unable to match tokens varying in the same contrast in an ABX categorisation task. They propose that whether attention can be paid to a stress contrast is determined, analogously to the case of segmental contrasts, by whether it functions to distinguish between words; stress contrasts however obviously differ from segmental contrasts in listeners' relative discrimination success. Peperkamp and Dupoux's account offers a potential explanation for this in a relatively late setting of the Stress Parameter (contrastive vs. non-contrastive) in acquisition: at the time that the parameter is set, language learners already have acquired at least a content/function word opposition.

4.3 Do vowels and consonants differ in their contribution to the recognition of spoken words?

Whether vowels and consonants constitute dichotomous classes is a contentious issue for processing models, as for linguistics. Within

the continuous-activation framework sketched in section 2 above, there would seem to be no basis for a categorical distinction; any information which allows a distinction to be made between two words, be it vocalic or consonantal information, should be equally useful to the processor. Consistent with this, Soto-Faraco et al. (2001) demonstrated exactly equivalent effects of single-vowel and single-consonant mismatch on lexical activation; the effect of the vowel signalling a difference between *sardina* and *sardana* (Soto-Faraco et al.'s experiments were conducted in Spanish) was in no way different from the effect of the consonant signalling a difference between *papilla* and *patilla* or (with more features mismatching) between *cinico* and *civico.*

Nevertheless there are a number of robust vowel-consonant differences which have appeared in perceptual experiments. One of the most striking is the consistent finding, from experiments using the word reconstruction task (van Ooijen, 1996), that it is easier to locate a real-word candidate by altering a vowel in the input than by altering a consonant. Listeners presented with *eltimate* or *weddow* find it easier to reconstruct these nonword inputs into *ultimate* and *widow* than *estimate* and *meadow.* As Moates et al. describe, this vowel/consonant asymmetry has been demonstrated in English (van Ooijen, 1996), in Dutch and in Spanish (Cutler, Sebastian-Galles, Soler Vilageliu & van Ooijen, 2000), and most recently, in a modified variant of the task, in Japanese (Cutler & Otake, 2002).

A difference between vowel and consonant processing has also been observed in the phoneme detection task. Response times to vowels are inversely correlated with the target duration — the longer the vowel token, the faster listeners detect it (van Ooijen, 1994; Cutler, van Ooijen, Norris & Sanchez-Casas, 1996). Again, this effect appears in both English and Spanish (Cutler et a1, 1996), and it is not observed with consonants (van Ooijen, 1994).

Both these effects have been explained in terms of learned responses on the part of listeners to contextually induced variability of phonetic segments in speech. Vowels are more variably realised; because of this listeners have built up a history of initially erroneous vowel identifications which have had to be corrected, and they

have learned to be cautious when required to make a definite identification (as in the phoneme detection task). However, longer vowel tokens are more likely to approach a canonical realisation, and the longer the token, the more likely listeners are to achieve a confident detection response without additional evidence from post-vocalic context. The experience with correction of an initially inaccurate vowel hypothesis has in turn rendered, in the word reconstruction task, the adjustment of a vowel a much more readily available operation than the adjustment of a consonant. This explanation is fully compatible with the word activation framework described above; in principle all types of phonetic information are equal, but in practice some are more variable than others.

The present papers add usefully to this discussion. Moates et al. present a further demonstration of the vowel advantage in word reconstruction; in addition, they show that responses involving reconstruction of more frequent phonemes are easier to produce than responses requiring production of less frequent phonemes. This finding strengthens the interpretation that patterns of performance in this task reflect listeners' past experience with hypotheses about phonetic identity in speech input. Kubozono's results show that indirect (visual) cues to a vowel length contrast may be ignored, as listeners have never learned (or needed) to rely on them; Peperkamp and Dupoux argue that the type of vowel contrasts which a language makes can modulate listeners' sensitivity to stress distinctions. Again, both results show listeners' processing shaped by past phonetic experience.

Bradlow's claim that coarticulated segments do not suffer from reduced distinctiveness may seem incompatible with the explanation of vowel/consonant processing differences in terms of experience of variability. However, distinctiveness in the utterance context is not necessarily the same as distinctiveness for the fraction of a second which corresponds to a segment's central realisation. Listeners have learned that speakers will give them all the information they need, though this information may be considerably distributed (and perhaps more so for vowels than for consonants). In word recognition all that is required is attention to the output of the lexical processor; incoming phonetic information will

act to constrain word-candidate activation as soon as it is available, whether or not its availability can be tied to some particular stretch of the speech signal uniquely corresponding to a given segment. Tasks such as phoneme detection and word reconstruction, on the other hand, require attention to individual segments, and in such tasks, differences in listeners' readiness to make confident identifications of one versus another segment type may more easily be observed.

4.4 Is coarticulation a severe problem for the language processor?

The fact that phonetic evidence which distinguishes one word from another (and thus by definition represents phonemes in the speech signal) is continuous rather than discrete has long been seen as a problem for listeners. Phonemes do not correspond to clearly separable sequential portions of the signal; if they can indeed be said to be present in speech, then at the very least they overlap. Listeners have the task of decoding an encoded representation (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967) if they are to extract from speech a percept in terms of a string of phonemes.

There is an enormous literature on the nature of coarticulatory effects (see e.g. Farnetani, 1997; Hardcastle & Hewlett, 1999) and their reflection in listeners' judgements (see e.g. Nygaard & Pisoni, 1995). But the very notion of coarticulation implies some reality to separable segments which in principle might be articulated without reference to one another (see Beckman, 1999, and Kuhnert & Nolan, 1999, for discussion of this issue). Only if (a) non-coarticulated segments are easier for listeners to process than coarticulated segments, and (b) segment perception is required of the listener, could one make a strong case that coarticulation complicates the listening process.

Certainly coarticulation leads to variability in the portions of speech which cue identity of any given phoneme, and where phonemic decisions are indeed required, variability may slow decision-

making. Thus the greater readiness of listeners to replace vowels than consonants in word reconstruction, as discussed above and as replicated once again by Moates et al., has been attributed to a side effect of coarticulation. Cutler et al. (2000) argued that vowels are more likely to be initially misidentified than consonants, so that listeners are usually more likely to have to revise decisions about vowels than about consonants in lexical processing. This experience translates into greater readiness to try another vowel than another consonant in the reconstruction task. The robustness of the effect across languages argues in favour of this account, given that it has been observed not only in vowel-rich languages like English and Dutch (van Ooijen, 1996; Cutler et al., 2000; Moates et al.) but also in two languages with five-vowel inventories, Spanish (Cutler et al., 2000) and Japanese (Cutler & Otake, 2002).

Explicit phonemic decision-making is however not required in everyday listening, and, despite the undoubted underlying contribution played by the phoneme as the minimal distinction between words, implicit phonemic decision-making may not be part of spoken-word recognition either (see Norris, McQueen & Cutler, 2000 for further discussion). Listeners need to identify words, not the component parts of words; coarticulation may speed this process rather than retarding it.

The type of word processing model which Pierrehumbert argues for, in which stored representations of words directly reflect a history of individual word processing episodes, seems in principle to require no phonemic decision-making. Interestingly, coarticulatory phenomena offer relevant evidence for evaluating this type of model. Studies of coarticulation reveal regularities which are determined by phonemic environment - the gestures which correspond to /k/ are different if the following vowel is high front /i/ rather than low back /c/, for instance. Such studies have not revealed a role for the word itself as a determiner of regularity - high frequency words such as *key* and *cause* and low frequency words such as *kiwi* and *caucus* show the same patterns of variation. Without some expansion of the episodic modelling framework beyond word-specific phonetics, such regularities must presumably be ascribed to chance.

The role of coarticulation in listening is central to Bradlow's contribution; she argues that coarticulation is not at all harmful to the listener's interests, and bases her argument on the fact that speakers do not suppress coarticulatory influences when they are deliberately trying to speak clearly. Word recognition theorists have argued for more than a decade (see e.g. McClelland & Elman, 1986b) that since coarticulatory effects are predictable, they should rather be helpful than harmful for word identification. Indeed, there is abundant evidence that listeners are very adept at extracting information from cross-phonemic coarticulatory influences. Typically, in the kind of experiment which has provided such evidence, listeners are presented with cross-spliced speech. For instance, a token of *slee-* from *sleep* and another token of *slee-* from *sleek* may be judged to have the same phonemic structure, but if the former is spliced to the /k/ of *sleek* and the latter to the /p/ of *sleep,* recognition of those final consonants, and of the whole words, is impaired (Martin & Bunnell, 1981, 1982; Whalen, 1984, 1991; Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 1999). Gating experiments (in which listeners are presented with word fragments, of increasing size, and are asked to guess word identity) also show that listeners make effective use of coarticulatory cues to identify upcoming phonemes (Ellis, Derbyshire & Joseph, 1971; Lahiri & Marslen-Wilson, 1991).

4.5   Where does language-specificity occur in the processing system?

The processing model sketched in part 2 above has no features which limit it to a specific language or group of languages. All listeners will have discrete memory representations of sound-meaning pairings, which will be activated by incoming speech signals; multiple simultaneous activation and inter-word competition presumably form the basis of a universal model of spoken-word recognition.

This is not to claim that all languages provide the same kind of information for lexical activation. Phoneme repertoires differ in

their size and in the features which distinguish phonemes; thus in a language which distinguishes dental from retroflex articulation, this difference will affect word-candidate activation, while the same difference might occur in speech in another language but be irrelevant for defining word identity. Similarly, suprasegmental information will be useless for distinguishing words in some languages, but is vital to distinguish words of a tone language. Nevertheless, the basic architecture of activation and competition is assumed to be constant.

The universal architecture includes effects which in principle might have been language-specific. Norris, McQueen, Cutler & Butterfield (1997) showed that the competition process is effectively modulated by a viability filter on the ongoing parse of the signal which would result from putative candidate words. If accepting a potential word would mean that a residue of the speech signal would be left over and could not be parsed as a word, that potential word is reduced in activation. Norris et al. called this effect the Possible-Word Constraint (PWC): it rules out residues which would make it impossible to parse the input as a continuous sequence of words. It is particularly useful as a way of ruling out activated words which are spuriously present via embedding; *ring* in *bring* can be rejected because the leftover *b* is not a viable word candidate.

Interestingly, what counts as an unparseable residue does not seem to differ across languages, although languages differ in what they allow as minimal words. Single consonants without vowels, such as *b,* are always unacceptable residues; but monomoraic syllables do not violate the PWC even in English, a language with a bimoraic minimal word (Norris, McQueen, Cutler, Butterfield & Kearns, 2001), nor does a single syllable violate the PWC in Sesotho, a language in which surface words must be bisyllabic (Cutler, Demuth & McQueen, 2002). The PWC might have tested residues against the vocabulary requirements of the language in question (bimoraic, bisyllabic, etc.), but it appears instead to be universal.

However, the PWC co-operates with language-specific effects in the competition process. Thus segmentation of continuous speech into its component words is sensitive to the boundaries of rhythmic

units, but languages differ in rhythmic structure; segmentation in English, a language with stress rhythm, is sensitive to foot boundaries (Cutler & Norris, 1988), but segmentation in Japanese, a language with moraic rhythm, is sensitive to mora boundaries (Otake, Hatano, Cutler & Mehler, 1993). These boundaries also define the domain of operation of the PWC (Norris et al., 1997; McQueen, Otake & Cutler, 2001).

Thus the processing model of word recognition is universal but realised with respect to the features specific to a particular language. The evidence provided in the present papers does not challenge this framework. Bradlow's findings imply that coarticulated information about segmental structure is equivalently available in languages with large or small vowel repertoires; Moates et al.'s results extend the cross-linguistically consistent pattern whereby listeners consider vowel information to be more mutable than consonant information. Peperkamp and Dupoux show that stress contrasts, though realised similarly in a pair of languages, may be salient for word recognition only in one, and Kubozono shows how a temporal distinction, though realised similarly word-medially and word-finally, may be exploited for discriminating between words in only one of these positions. A welcome next step would be experiments in which these effects were investigated in laboratory paradigms specifically designed to study activation and competition.

5.   Conclusion

As the above comments suggest, current psycholinguistics is characterised by multiple lines of research which concern, one way or another, the role of phonological constructs in processing. The evidence provided by the five papers discussed fits into several of these themes, confirms and amplifies some conclusions from existing work, and is in turn illuminated by some previous findings.

Those papers which raise perceptual issues do not challenge the generally agreed architecture of the spoken-word recognition system, with its central role for automatic lexical activation and

inter-word competition. Nor do the papers which address issues of language production provide direct evidence which would require incorporation in the model described by Levelt (this volume) and variously addressed also by papers on which Levelt comments. (In fact it is tempting to speculate that the methodological mismatch - psycholinguistics prefers on-line procedures, while laboratory phonology rarely uses such methods - might make a direct challenge difficult to mount in any case.)

Finally, consider the observation, pointed out in part 3 above, that psycholinguists have abandoned the issue of whether phonological constructs might serve as obligatory perceptual entities. If the present set of papers is a representative sample of (laboratory) phonological studies of processing topics, we may conclude that this question is a non-issue for laboratory phonologists as well.

References

Beckman, M. E.
   1999   Implications for phonological theory. In W. J. Hardcastle & N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques,* (pp. 199-225). Cambridge: Cambridge University Press.

Berinstein, A. E.
   1979   A cross-linguistic study on the perception and production of stress. UCLA *Working Papers in Phonetics,* 47.

Cluff, M. S. & Luce, P. A.
   1990   Similarity neighborhoods of spoken two-syllable words, Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance,* 16, 551—563.

Connine, C. M., Blasko, D. G. & Wang, J.
   1994   Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics,* 56, 624-636.

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,* (pp. 23-40). London: Academic Press.

Cutler, A.
   1992a   Psychology and the segment. In G. J. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* (pp. 290-295). Cambridge: Cambridge University Press.

Cutler, A.
   1992b   Why not abolish psycholinguistics? In W. U. Dressier, H. C.
            Luschiitzky, O. E. Pfeiffer & J. R. Rennison (eds.), *Phonologica
            1988,* (pp. 77-87). Cambridge: Cambridge University Press.
Cutler, A. & Butterfield, S.
   1990   Durational cues to word boundaries in clear speech. *Speech Com-
            munication, 9,* 485-495.
Cutler, A., Demuth, K. & McQueen, J. M.
   2002   Universality versus language-specificity in listening to running
            speech. *Psychological Science,* 13, 258—262.
Cutler, A., Mehler, J., Norris, D. G. & Segui, J.
   1987   Phoneme identification and the lexicon. *Cognitive Psychology,* 19,
            141-177.
Cutler, A. & Norris, D. G.
   1988   The role of strong syllables in segmentation for lexical access. *Jour-
            nal of Experimental Psychology: Human Perception and Perfor-
            mance,* 14, 113—121.
Cutler, A., Ooijen, B. van, Norris, D. & Sanchez-Casas, R.
   1996   Speeded detection of vowels: A cross-linguistic study. *Perception &
            Psychophysics,* 58, 807-822.
Cutler, A. & Otake, T.
   2002   Rhythmic categories in spoken-word recognition. *Journal of Mem-
            ory and Language,* 46, 296—322.
Cutler, A., Sebastian-Galles, N., Soler Vilageliu, O. & Ooijen, B. van
   2000   Constraints of vowels and consonants on lexical selection: Cross-
            linguistic comparisons. *Memory & Cognition,* 28, 746—755.
Dell, G. S. & Newman, J. E.
   1980   Detecting phonemes in fluent speech. *Journal of Verbal Learning &
            Verbal Behavior,* 19, 609-623.
Dupoux, E., Pallier, C, Sebastian-Galles, N. & Mehler, J.
   1997   A destressing 'deafness' in French. *Journal of Memory and Lan-
            guage,* 36, 406-421.
Ellis, L., Derbyshire, A. J. & Joseph, M. E.
   1971   Perception of electronically gated speech. *Language and Speech,* 14,
            229-240.
Farnetani, E.
   1997   Coarticulation and connected speech processes. In W.J. Hard-
            castle & J. D. M. H. Laver (eds.), *A Handbook of Phonetic Science,*
            (pp. 371-404). Oxford: Blackwell.
Goldinger, S. D.
   1998   Echoes of echoes? An episodic theory of lexical access. *Psychologi-
            cal Review,* 105,251-279.
Goldinger, S. D., Luce, P. A. & Pisoni, D. B.
   1989   Priming lexical neighbours of spoken words: Effects of competition
            and inhibition. *Journal of Memory & Language,* 28, 501-518.

Gow, D. W. & Gordon, P. C.
   1995   Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance,* 21, 344—359.
Grosjean, F. & Frauenfelder, U. H. (eds.)
   1996   Spoken Word Recognition Paradigms. *Special issue of Language and Cognitive Processes, 11(6).*
Hardcastle, W. J. & Hewlett, N. (eds.)
   1999   *Coarticulation: Theory, Data and Techniques.* Cambridge: Cambridge University Press.
Johnson, K.
   1997   Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (eds.), *Talker Variability in Speech Processing,* (pp. 145—166). San Diego: Academic Press.
Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U. & Lacerda, F.
   1997   Cross-language analysis of phonetic units in language addressed to infants. *Science,* 277, 684-686.
Kuhnert, B. & Nolan, F. J.
   1999   The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques,* (pp. 7—30). Cambridge: Cambridge University Press.
Lahiri, A. & Marslen-Wilson, W. D.
   1991   The mental representation of lexical form, A phonological approach to the recognition lexicon. *Cognition,* 38, 245-294.
Levelt, W. J. M., Roelofs, A. & Meyer, A. S.
   1999   A theory of lexical access in speech production. *Behavioral and Brain Sciences,* 22, 1-38.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M.
   1967   Perception of the speech code. *Psychological Review,* 74, 431—461.
Maddieson, I.
   1984 *Patterns of Sounds.* Cambridge: Cambridge University Press.
Marslen-Wilson, W. D.
   1990   Activation, competition and frequency in lexical access. In G. T. M. Altmann (ed.), *Cognitive Models of Speech Processing,* (pp. 148-172). Cambridge MA: MIT Press.
Marslen-Wilson, W. & Warren, P.
   1994   Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review,* 101, 653 — 675.
Martin, J. G. & Bunnell, H. T.
   1981   Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America,* 69, 559—567.

Martin, J. G. & Bunnell, H. T.
    1982   Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance,* 8, 473—488.

McClelland, J. L. & Elman, J. L.
   1986a The TRACE model of speech perception. *Cognitive Psychology,* 18, 1-86.

McClelland, J. L. & Elman, J. L.
    1986b   Exploiting lawful variability in the speech wave. In J. Perkell & D. H. Klatt (eds.), *Invariance and Variability in Speech Processes,* (pp. 360-385). Hillsdale, NJ: Erlbaum.

McQueen, J. M., Norris, D. G. & Cutler, A.
    1994   Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 20, 621-638.

McQueen, J. M., Norris, D. G. & Cutler, A.
    1999   Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance,* 25, 1363—1389.

McQueen, J. M., Otake, T. & Cutler, A.
    2001   Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language,* 45, 103—132.

Norris, D. G.
    1994   Shortlist: A connectionist model of continuous speech recognition. *Cognition,* 52, 189-234.

Norris, D. G., McQueen, J. M. & Cutler, A.
    1995   Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 21, 1209-1228.

Norris, D. G., McQueen, J. M. & Cutler, A.
    2000   Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences,* 23, 299—325.

Norris, D. G., McQueen, J. M., Cutler, A. & Butterfield, S.
    1997   The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology,* 34, 191—243.

Norris, D. G., Cutler, A., McQueen, J. M., Butterfield, S. & Kearas, R.
    2001   Language-universal constraints on speech segmentation. *Language and Cognitive Processes,* 16, 637—660.

Nygaard, L. C. & Pisoni, D. B.
    1995   Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (eds.), *Speech, Language and Communication.* Volume 11 of E. C. Carterette & M. P. Friedman (eds.), *Handbook of Perception and Cognition,* (pp. 63-96). NY: Academic Press.

Ooijen, B. van
    1994   *Processing of vowels and consonants.* PhD Dissertation, University
              of Leiden.
Ooijen, B. van
    1996   Vowel mutability and lexical selection in English: Evidence from a
              word reconstruction task. *Memory & Cognition, 24,* 573—583.
Otake, T., Hatano, G., Cutler, A. & Mehler, J.
    1993   Mora or syllable? Speech segmentation in Japanese. *Journal of
              Memory and Language,* 32, 358—378.
Pallier, C, Sebastian-Galles, N., Felguera, T., Christophe, A. & Mehler, J.
    1993   Attentional allocation within the syllabic structure of spoken words.
              Journal of Memory and Language, 32, 373—389.
Picheny, M. A., Durlach, N. I. and Braida, L. D.
    1985   Speaking clearly for the hard of hearing I: Intelligibility differences
              between clear and conversational speech. *Journal of Speech and
              Hearing Research,* 28, 96-103.
Picheny, M. A., Durlach, N. I. and Braida, L. D.
    1986   Speaking clearly for the hard of hearing II: Acoustic characteristics
              of clear and conversational speech. *Journal of Speech and Hearing
              Research,* 29, 434-446.
Pitt, M. A. & Samuel, A. G.
    1990   Attentional allocation during speech perception, How fine is the
              focus? *Journal of Memory and Language,* 29, 611 —632.
Price, C, Indefrey, P. & Turennout, M. van
    1999   The neural architecture underlying the processing of written and
              spoken word forms. In P. Hagoort & C. Brown (eds.), *Neurocogni-
              tion of Language,* (pp. 211-240). Oxford: Oxford University Press.
Samuel, A. G. & Ressler, W. H.
    1986   Attention within auditory word perception, Insights from the pho-
              neme restoration illusion. *Journal of Experimental Psychology: Hu-
              man Perception & Performance,* 12, 70-79.
Scott, D. R. & Cutler, A.
    1984   Segmental phonology and the perception of syntactic structure.
              Journal of Verbal Learning and Verbal Behavior, 23, 450-466.
Shillcock, R. C.
    1990   Lexical hypotheses in continuous speech. In G. T. M. Altmann
              (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and
              Computational Perspectives,* (pp. 24-49). Cambridge, MA: MIT
              Press.
Soto-Faraco, S., Sebastian-Galles, N. & Cutler, A.
    2001   Segmental and suprasegmental mismatch in lexical access. *Journal
              of Memory and Language,* 45, 412-432.
Streeter, L. A. & Nigro, G. N.
    1979   The role of medial consonant transitions in word perception. *Jour-
              nal of the Acoustical Society of America,* 65, 1533-1541.

Tabossi, P., Burani, C. & Scott, D.
    1995   Word identification in fluent speech. *Journal of Memory and Language, 34,* 440—467.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. & Stokes, M. A.
    1988   Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84,* 917—928.

Vitevitch, M. S. & Luce, P. A.
    1998 When words compete, Levels of processing in spoken word recognition.- *Psychological Science, 9,* 325-329.

Vroomen, J. & de Gelder, B.
    1995   Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 98—108.

Warner, N., Jongman, A., Sereno, J. & Kemps, R.
    submitted   Subphonemic durational differences in production and perception. *Journal of Phonetics.*

Whalen, D. H.
    1984   Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49—64.

Whalen, D. H.
    1991   Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50,* 351-360.

Zwitserlood, P.
    1989   The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25-64.