Modelling the Acquisition of Lexical Segmentation

PAUL CAIRNS, RICHARD SHILLCOCK, NICK CHATER, & JOE LEVY*
University of Edinburgh

Introduction: The Segmentation Problem

Contrary to the impression that a listener might have, the words in a stream of conversational English are, in the main, not clearly delimited by acoustic discontinuities or pauses (Cole 1980). Given this, a neonate faced with the problem of learning a language must find some way of isolating the linguistically relevant chunks of the input in order to eventually compile a lexicon and access meaning.

In explaining how the infant learns to segment, there is a dichotomy between explanations that posit INTERACTIVE processing, and those that are BOTTOM-UP. These differing accounts are paralleled by a distinction in adult processing models that has been debated in the literature.

The interactive position (e.g., Cole and Jakimik 1980; Marslen-Wilson and Welsh 1978) is that the lexeme licenses segmentation. As the speech input arrives over time, the words that become incompatible with the input are incrementally eliminated until one winning candidate emerges. The stored lexical phonology of the winner specifies the boundary at which the next word begins. Thus, after the initial portion of the input has activated the lexical cohort, the information that facilitates segmentation crucially flows from the lexical level, and therefore is essentially top-down. The Cohort Model (Marslen-Wilson and Welsh 1978) formed the basis of the TRACE connectionist model (McClelland and Elman 1986) in which activation from lexical nodes serves to cut up the input — lexical segmentation is a by-product of word recognition.

Bottom-up accounts, on the other hand, propose that listeners segment on the basis of some marking in the speech signal. These theories preclude any top-down influence from the lexicon or other sources. The bottom-up cues that may be used to aid segmentation can be divided into three main groupings: (i) Acoustic/phonetic juncture markers or pauses (Lehiste 1971) (ii) Prosodic marking that specifies the initial portion of a word, given a pre-syllabified input (Cutler and Norris 1988; Cutler 1993; Cutler and Butterfield 1992). (iii) Distributional cues: for example differing probabilities of certain phonological sequences at various points in the speech stream (PHONOTACTICS — see Harrington, Watson and Cooper 1988).

The dominant bottom-up approach in current psycholinguistic theory is (ii), which has been thoroughly investigated by Cutler and colleagues. Her

*This work was supported by the U.K. Economic and Social Research Council (ESRC). Grant number: R000 23 3649

METRICAL SEGMENTATION STRATEGY (MSS) holds that segmentation depends on the exploitation of prosodic rhythm. Speakers of a particular language are sensitive to its rhythmic pattern, and chunk the input accordingly, initiating lexical access at the beginning of each chunk. In English, the relevant rhythm is the alternation between strong and weak syllables, where weak syllables have /ə/ or another lax vowel as nucleus — all others are strong. When a strong vowel is heard, a boundary is hypothesized at the beginning of the syllable of which the vowel is nucleus.

The interactive and bottom-up approaches represent the main division in models of adult segmentation behaviour. In acquisition, both approaches have been suggested. Therefore, there are two relevant questions which must be answered by any model of the acquisition of segmentation: (1) how likely is it that a learner could use a particular strategy given what we know about the language input?; and (2) how effective would that strategy be? Note that (2) can be partly answered by studying the effectiveness of the adult models.

The interactive strategy seems initially unlikely as a method for bootstrapping segmentation since it presupposes that the listener has a lexicon with which to segment. However, it has been suggested that the presence of words spoken in isolation in the input to the infant could allow a lexicon to be compiled without recourse to segmentation (e.g., Suomi 1993). When a word is heard in isolation it could be added to the lexicon and thus subsequently used in interactive segmentation. As time goes on, more words are added to the lexicon and thus more interactive segmentation is possible. This cumulative effect finally allows the infant to acquire a full lexicon. However, there are two problems with this account: First, only a small proportion of utterances are one-word, and; second, how is a child to know that any given utterance is composed of only one word, and not two or three?

Data from the CHILDES corpus (MacWhinney and Snow 1985) show that only around 14% of child-directed utterances are composed of just one word.[1] Moreover, these items are typically words such as *yeah*, *uhhuh*, and *ok* which will seldom occur embedded in phrases in any case, and are thus of little use in carrying out interactive segmentation. An analysis of the London-Lund Corpus (LLC — Svartvik and Quirk 1980) carried out by the authors, showed that this figure is somewhat higher for normal adult conversation: around 20%. So, it seems that we need further experimental or computational evidence before a decision can be reached as to the plausibility of the interactive strategy in infant development.

Of course, there are also doubts as to how effective an interactive strategy would be, even if it could be bootstrapped by the infant. These problems have been described in various works which criticise strictly left-to-right models such as the earlier versions of the Cohort Model. Critics have pointed out that the vocabulary structure of English is such that the majority of words do not become unique until after their offset (?), and that in experimental studies, listeners do not recognise some 20% of words until after their acoustic offsets (Bard, Shillcock and Altmann 1988). These problems have not gone unnoticed by interactive theorists. For example, Marslen-Wilson has produced a revised version of the Cohort model in which strictly left-to-right processing is abandoned (Marslen-Wilson 1987). However, it still seems to be the case that the interactive account, which involves some lexical matching, is prone to failure under conditions in which noise in the signal undermines the matching

[1] Thanks to Steve Finch (pers. comm.) for this statistic.

procedure.

The status of bottom-up theories in acquisition would seem to be less problematic. Although the MSS was originally a theory of adult behaviour, there has been recent discussion of how the MSS could be acquired (Cutler et al. 1992; Otake et al. 1993). The basis of the theory is that infants are endowed with a sensitivity to rhythm, and when presented with speech input will tend to extract portions of the stimulus that correspond to the prosodic sequences with the smallest periodicity. This approach seems much more likely as the basis of a theory of infant segmentation behaviour than does the interactive stance, principally because it seems to impose no unreasonable constraints on the nature of the input to the child, but also because it is robust and requires only minimal *a priori* knowledge. Furthermore, experimental work seems to demonstrate a sensitivity to metrical patterns in infants of 9 months (Jusczyk, Cutler and Redanz 1993). However, if we take the case of English, there remain questions that a fully developed theory of MSS acquisition would have to answer:

- Stress is a perceptual phenomenon — there is no simple relationship between perceived stress and acoustic variables such as amplitude duration and pitch. Furthermore, the mechanisms of stress realization vary from language to language. Therefore, the MSS will need to be augmented with an explanation of how the categories STRONG VOWEL and WEAK VOWEL come to be established as salient perceptual categories.

- If the categories STRONG VOWEL and WEAK VOWEL are perceptually salient primitives, then the input will be perceived as a string of strong and weak syllables: *SWSWWSWSSWSW*... But, even if the infant apprehends the rhythmic alternation in the input, there is still no reason why boundaries should be hypothesized before each strong syllable, as opposed to, say, *after* every strong syllable.

It has been shown that the MSS is an effective strategy for English due to the structure of the English lexicon, in which the overwhelming majority of content words have strong initial stress, and thus can be extracted by the MSS (Cutler and Carter 1987). However, it is important to note that the MSS does not cover exactly the same problem domain as the interactive accounts, since the complex issue of syllabification is presupposed by the model, yet not dealt with explicitly.

In our discussion of bottom-up accounts so far, we have concentrated on the MSS, which relies on prosodic marking, ignoring the role of the two other bottom-up information sources for segmentation that were outlined above. In the rest of this paper we will investigate the role of the third type of cue: PHONOTACTICS — constraints on the segmental phonological structure of words and syllables.

It is generally the case that sequences of segments are more constrained within words than across word boundaries. Thus, the sequence /ŋð/ is only licenced in English if there is a morpheme boundary between /ŋ/ and /ð/. However, phonotactics do not have to be absolute constraints, probabilistic structure is present too: thus the sequence /z ə/ is very common across word boundaries, but much less common word-internally.

The role of phonotactics has been studied in the speech recognition literature (see Harrington, Watson and Cooper 1988), but has received little attention in the domain of psycholinguistics. In this study, by attempting to quantify the role of phonotactics in bootstrapping segmentation, we will clarify the parameters of a more complete model which would integrate the various bottom-up information sources.

Empirical motivation for examining the role of phonotactics in infant segmentation behaviour comes from experiments which seem to demonstrate a sensitivity to phonotactic structure in infants as young as 9 months (Jusczyk et al. 1993). If infants are sensitive to such information then it seems likely that they would also come to appreciate it's power in predicting boundary location.

Bootstrap Modelling of Segmentation

We now outline a neural network model that is used to investigate the possible contribution of phonotactic information to bootstrapping segmentation. First we describe the phonological input corpus, before discussing the neural network model itself.

The input corpus

The London-Lund Corpus is a large body of English conversation transcribed orthographically and available on-line. Because of its size (around 460,000 words) an automatic method was developed for its phonetic transcription. First, the words are replaced by their phonemic citation forms using an on-line dictionary. Then, these forms are input to a set of re-write rules that introduce phonological alternations into the string (e.g., assimilation, vowel reduction). None of the rules uses word boundary information to specify its context of application. The output from the rule-set is a corpus of 1.5 million phonetic segments.

It is, of course, impossible to recreate the original speech data, but this method has two advantages: First, we need a very large corpus of conversational speech if its statistics are to be representative — at present there is no comparably large corpus with a genuine phonological transcription; Second, this method provides a higher-order approximation to genuine data, when compared with a corpus derived from a phonemic dictionary in combination with word frequency counts. Thus, our data will be representative of the distribution of strings of closed-class words such as *if I can*.

This segmentally transcribed version of the corpus is then retranscribed in terms of 9 binary features that represent the cognitive elements of Government Phonology (see Harris and Lindsey (in press); Kaye, Lowenstramm and Vergnaud 1985; Kaye, Lowenstramm and Vergnaud 1990). The nine elements are seen as cognitive categories which are universals and which are sufficient for the representation of speech. The sound patterns of the acoustic input are mapped onto these elements. Harris and Lindsey advance a Realisational Autonomy hypothesis, which allows each of the elements to be realised phonetically, independent of the other elements. Recent work (Williams and Brockhaus 1992) has shown how the government phonology elements can be automatically extracted from the speech stream, providing further motivation for the use of this representation in a psychologically responsible model.

The model

We have constructed unsupervised n-gram models in which no word boundary marking is employed in construction. These models, are admissible as models of infant development because no top-down information is employed during training. However, such models inherently employ phonemic categories, and hence cannot be used to address the pre-categorial phase of infant development. Therefore we developed a neural network model that is feature, rather than phoneme category based. We consider a feature based representation to be one step closer to the genuine speech signal.

The central idea of the network model is that of *prediction*: the network is presented with an uninterrupted stream of input speech and is asked to predict what the values of the nine Government Phonology features will be at the next time step. This is, of course, an extremely difficult task, with complete predictability being impossible. However, in striving to achieve its goal the network is forced to learn about the sequential featural statistics of English speech, thus lowering its prediction error. The rationale used in postulating boundary points follows from what we know about phonotactics: lack of constraint in phonotactic structure (high ENTROPY in information theoretic terms) will make the next segment difficult to predict. If prediction is hard, then error will be high. Thus, boundaries are proposed at prediction error peaks.

Network training

The network has a recurrent, self-supervised, architecture (see figure 1). The task is to echo the current slice of input, remember the previous, and, most importantly for this study, to predict the next. Noise is added to the input by flipping features from 0 to 1 (or vice versa) with a certain probability, in order to encourage the network to rely on sequential information (i.e., if the current segment is degraded, then the net will have an incentive to use the local phonetic context to recover its identity). The net is trained using Back-propagation Through Time (Rumelhart, Hinton and Williams 1986), a steepest descent procedure, and a cross-entropy error measure (see Hinton 1989 — cross entropy is a good measure to use when one wishes to interpret continuous valued outputs as probabilities of binary decisions). Training comprises two passes through a training stretch of the corpus one million phonemes in length (with different noise on each pass), thus two million phonemes in total. The learning rate is decayed as training progresses.

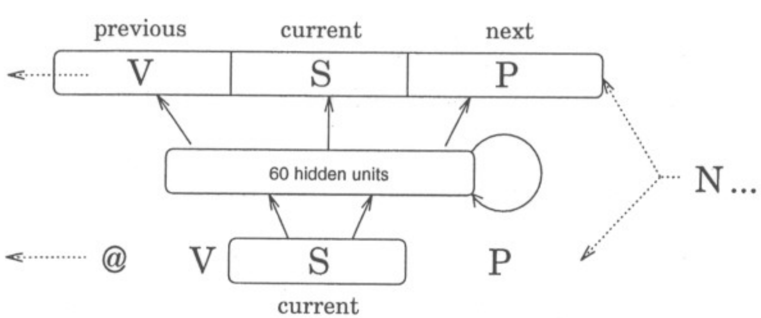

Figure 1: The Network – The solid arrows between layers indicate complete connectivity with modifiable uni-directional links. The dotted arrows show how the input corpus arrives over time to specify the input and output target.

Network segmentation

The model was tested by providing as input a noise-free 10,000 phoneme (about 2,700 words) stretch of corpus, and measuring the Cross Entropy error on the prediction sub-group of the output units. This yields a variable error signal in which we define a 'peak' by placing a cutoff point at varying numbers of standard deviations above the mean. As stated above, high error will tend to indicate increasing likelihood of a juncture point because lack of phonotactic constraints across boundaries makes prediction of word-initial segments difficult. At the cutoff that maximizes the network's performance, 21% of the boundaries are correctly identified with a hits:false-alarms ratio of 1.5:1.[2]

In the following sections we evaluate the significance of these results by comparison with a random segmentation algorithm which was averaged over five different runs. This algorithm was designed to yield a distribution of pseudo word lengths similar to that of the network. We consider this to be a more stringent test of the network's performance than comparison with a random segmentation algorithm that uses a uniform distribution.

Although network performance peaks with correct identification of about one in five boundaries in the test corpus, there is a sizable proportion of false alarms at this cutoff (i.e., cases in which the network predicts a boundary when in fact there is none). It may well be that although the false-alarms do not actually correspond to existing boundaries in the test stretch, they are actually plausible guesses based on the low-level data that is the only information source available to the model. We tested this hypothesis by examining the phonotactic acceptability of the boundaries that the model postulates, defined by the legality of the sequence of segments over the postulated boundary. Thus the sequence /tp#ra/ is a phonotactically malformed boundary postulate, while /pt#ra/ is well-formed. We found that false-alarm boundaries of the network are much more likely to be phonotactically well-formed than those of the random case (for the initial boundaries: $\chi^2_{(1)} = 221.8$, $p < 0.001$, while for the final boundaries $\chi^2_{(1)} = 119.1$, $p < 0.001$).

In summary, phonotactics provide a fairly weak source of information for the bootstrapping of segmentation, but the cumulative effect of such information may well be useful in the initial phases of compiling a lexicon.

Network segmentation and the MSS

In this section we provide a qualitative analysis of network segmentation and present the surprising result that there is a statistical basis for the emergence of the MSS in our purely bottom-up model.

We investigated the performance of the model by counting the instances in which a boundary is correctly postulated before a strong or weak syllable. As an operational definition of STRONG and WEAK we took the lax vowels /ə/, /ɪ/, and /ʌ/ to be weak, and all other monothongs and dipthongs to be strong. Given this criterion, in the 2,700 word test set 53% of the words are strong-initial.[3] The network performance is proportionally skewed towards

[2] To decide when network segmentation was optimal, we measured the MUTUAL INFORMATION for each cutoff, and maximised this value. This allows our estimation of network performance to be independent of any assumptions about the relative desirability/undesirability of hits/false-alarms.

[3] This is representative of the proportion for real speech, see Cairns et al. 1994.

successful detection of strong-initial words to a striking degree (see Figure 2a, $\chi^2_{(1)} = 77.2, p < 0.001$). A similar result was obtained when we changed the

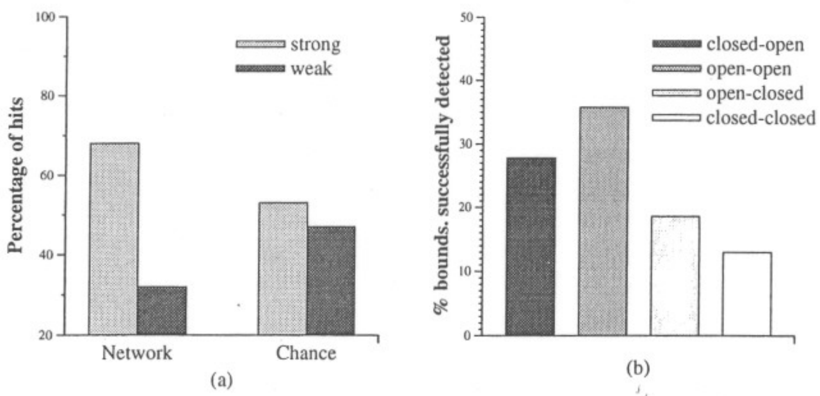

Figure 2: Network segmentation performance mimics the MSS: (a) shows the proportions of boundaries inserted before strong and weak syllables. (b) shows the breakdown of boundary insertions for boundaries between open- and closed-class words.

definition of WEAK to just /ə/ ($\chi^2_{(1)} = 70.4, p < 0.001$). A natural conclusion to draw is that the model is segmenting more before open-class words, and examination of the totals of hits before open- as opposed to closed-class shows that this is the case (recall that the majority of open-class content words in English have strong initial syllables). The initial portions of open-class words are much more likely to be detected than beginnings of closed-class items (see Figure 2b, $\chi^2_{(1)} = 14.0, p < 0.001$). Note also that the boundaries with which the model has most difficulty are the closed-closed boundaries, thus strings of closed-class words such as *up to the* are less likely to be segmented than strings of open-class items.

When we consider the contiguous pairings of these individual segmentations — the words that emerge from the network — the same pattern is evident. A word count of the LLC revealed that 65% of all items were closed-class, so one would expect that this ratio would hold in network output, all other factors being equal. While the network does not segment more whole words from the test stretch than it would by chance (showing that the model does not develop a lexicon), of the correctly extracted tokens 41% are closed-class. This is significantly less than the random segmentation performance: $\chi^2_{(1)} = 19.46$, $p < 0.001$.

So, our network produces segmentations which broadly mimic the pattern predicted by the MSS, yet the net is not retrodictive in the way that the MSS is: Crucially, the nuclear vowel of the initial syllable of a word is not visible to the network when it makes a segmentation decision (at least for CV...syllables — recall that boundaries are inserted on the basis of the ease of prediction of the first segment in the word). Thus, this model does not need to posit that

the Strong/Weak distinction has *a priori* perceptual salience for the infant. The reason why our network exhibits this pattern of results is simply that the initial segments of strong-initial, open-class words tend to be less predictable than those of closed-class words.

However, we would emphasize that phonotactic information is a relatively weak predictor, and that it is unlikely that a purely phonotactic approach could enable the infant to acquire a lexicon. Rather, we see phonotactic information as being a possible method for bootstrapping of the MSS, which is a more robust and reliable tool for lexical acquisition. This bootstrapping could be mediated by sensitivity to the correlation of the boundaries that phonotactics predict with metrical structure.

Adding categorial knowledge

The results from the previous section were obtained by segmenting with raw scores that were not normalized for phoneme type. This can be seen as simulating the phase of infant development in which phonemic categories, and information about their frequencies, are not available to the infant. However, we know that towards the end of the first year of life the child's phonological space is becoming structured in terms of phonemic categories (e.g., Kuhl 1983; Werker 1993). Therefore, we decided to mimic the effect of this phonemic restructuring in our model, to see if the qualitative pattern of segmentations would remain constant.

We carried out the same segmentation procedure as before, but this time normalizing the network error scores for phoneme type. We found an entirely different pattern of results with respect to strong-syllables and word class than before: in general the network no longer mimicked the MSS. Segmentation before strong as opposed to weak syllables was not significantly different from chance: $\chi^2_{(1)} = 0.387, n.s.$ Neither was segmentation before open as opposed to closed-class items: $\chi^2_{(1)} = 0.035, n.s.$ Furthermore, using phoneme-normalized scores, 78% of correctly extracted word tokens were closed-class, in contrast to the 41% with raw scores. This figure once again differs significantly from the expected distribution: $\chi^2_{(1)} = 8.07, p < 0.005$, except that now it is the closed-class items that are favoured, rather than the open-class.

The intuitive explanation of why segmentation behaviour should change in this way when scores are normalized is that closed-class words, because they are most frequent in the language, also contain the most frequent phonemes. Therefore, the network will predict these phonemes more easily than ones which do not occur often in closed-class words. Because predicting these segments is easier, errors are lower. Hence normalizing for phoneme type will augment the error scores for phonemes that most often occur in closed-class words, and effectively increase the probability of boundaries being proposed before such segments.

Conclusions

We have provided a computational underpinning to the claim that low-level phonotactics could be used by a neonate as a cue for initially breaking up the continuous stream of input speech.

Moreover, we have given an account of how the MSS could arise without recourse to positing sensitivity to metrical information as part of a genetic endowment. The network segmentation performance was significantly biased in favour of detecting open-class words that have strong initial syllables.

Furthermore, we have shown that our model's mirroring of the MSS disappears when we add knowledge about the frequencies of individual phoneme categories — detection of closed-class words becomes favoured.

We see the overall picture of the role of phonotactics that emerges from these results as follows: Initially, phonotactics could provide tentative segmentations from which the MSS could be induced in the pre-categorial infant. Once the MSS is in place, and the infant's phonological space comes to be structured with the phonemic categories of English, then the MSS would pick out the open-class words, while phonotactics could help in isolating the closed-class items. Also, we would not wish to rule out some form of lexical competition in segmentation once a lexicon is in place.

The disappearance of MSS-like behaviour raises the possibility of a critical period for realization of the MSS: If we assume that categorial knowledge is pervasive after a certain stage of development, then the utility of the phonotactic strategy for bootstrapping the MSS is only visible in the pre-categorial phase. However, while standard accounts of critical period phenomena suppose that the driving force behind the changes is the infant's maturation (e.g., Lenneberg 1967), our work suggests that there may be alternative explanations arising from informational constraints and interactions *external* to the child, but intricately linked to the learning process.

References

Bard, E. G., R. C. Shillcock and G. T. M. Altmann (1988) The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics* **44**(5), 395–408.

Cairns, P., R. Shillcock, N. Chater and J. Levy (1994) Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation. Submitted to *Cognitive Psychology*, currently being revised.

Cole, R. A. and J. Jakimik (1980) A model of speech perception. In R. A. Cole, ed., *Perception and Production of Fluent Speech*, pp. 133–163. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cole, R. A., ed. (1980) *Perception and Production of Fluent Speech*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cutler, A. and D. Norris (1988) The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology (Human Perception and Performance* **14**, 113–121.

Cutler, A. (1993) Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research* **22**(2), 109–131.

Cutler, A. and S. Butterfield (1992) Rhythmic cues to speech segmentation - evidence from juncture misperception. *Journal of Memory and Language* **31**(2), 218–236.

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

Cutler, A., J. Mehler, D. Norris and J. Segui (1992) The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* **24**, 381–410.

Harrington, J., G. Watson and M. Cooper (1988) Word boundary identification from phoneme sequence constraints in automatic continuous speech recognition. In *Proceedings of the 12th International Conference on Computational Linguistics*, pp. 225–230.

Harris, J. and G. Lindsey ((in press)) The elements of phonological representation. In J. Durand and F. Katamba, eds., *New Frontiers in Phonology*. Harlow: Longman.

Hinton, G. E. (1989) Connectionist learning procedures. *Artificial Intelligence* **40**, 185–234.

Jusczyk, P. W., A. Cutler and N. J. Redanz (1993) Infants' preference for the predominant stress patterns of English words. *Child Development* **64**, 657–687.

Jusczyk, P. W., A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud and A. M. Jusczyk (1993) Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* **32**, 402–420.

Kaye, J. D., J. Lowenstramm and J. R. Vergnaud (1985) The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* **2**, 305–328.

Kaye, J. D., J. Lowenstramm and J. R. Vergnaud (1990) Constituent structure and government in phonology. *Phonology* **7**(2), 193–231.

Kuhl, P. K. (1983) Perception of auditory equivalence classes for speech in early infancy. *Infant Behaviour and Development* **6**, 263–285.

Lehiste, I. (1971) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* **51**(6 (2)), 2018–2024.

Lenneberg, E. (1967) *Biological foundations of language*. New York: John Wiley and Sons.

Luce, P. A., D. B. Pisoni and S. D. Goldinger (1990) Similarity neighborhoods of spoken words. In G. T. M. Altman, ed., *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, Mass.: MIT Press.

MacWhinney, B. and C. Snow (1985) The child language data exchange system. *Journal of Child Language* **12**, 271–296.

Marslen-Wilson, W. (1987) Parallel processing in spoken word recognition. *Cognition* **25**, 71–102.

Marslen-Wilson, W. and A. Welsh (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* **10**, 29–63.

McClelland, J. L. and J. L. Elman (1986) Interactive processes in speech perception: The TRACE model. In J. L. McClelland and D. E. Rumelhart, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2: *Psychological and Biological Models*, pp. 58–121. Cambridge, Mass.: MIT Press.

Otake, T., G. Hatano, A. Cutler and J. Mehler (1993) Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* **32**, 258–278.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986) Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: *Foundations*, pp. 318–362. Cambridge, Mass.: MIT Press.

Suomi, K. (1993) An outline of a developmental model of adult phonological organization and behaviour. *Journal of Phonetics* **21**, 29–60.

Svartvik, J. and R. Quirk (1980) *A Corpus of English Conversation*. Lund: Liber-Laromedel Lund.

Werker, J. F. (1993) Developmental changes in cross-language speech perception: Implications for cognitive models of speech processing. In G. T. M. Altmann and R. C. Shillcock, eds., *Cognitive Models of Speech Processing: The Sperlonga Meeting II*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Williams, G. and W. Brockhaus (1992) Automatic speech recognition: a principle-based approach. In A. Göksel and E. Parker, eds., *Working papers in linguistics and phonetics 2*, pp. 371–401. London: School of Oriental and African Studies.