

2

The Recognition of Lexical Units in Speech

Anne Cutler and James M. McQueen
MRC Applied Psychology Unit, Cambridge, UK

In order to understand language, a listener or reader must be able to recognise discrete lexical units. The range of possible utterances (or texts) that we may be confronted with is infinite, so it is impossible to store them in memory. Instead, we must store smaller units, usually words. Incoming spoken or written language must therefore be mapped onto these stored units, held in a mental lexicon.

The recognition of these lexical entries depends upon identification of where each corresponding unit in the input begins and ends—that is, on the recognition of word *boundaries*. The written forms of many languages facilitate the segmentation of visual language into discrete lexical units by providing an explicit boundary marker: the white spaces between words. There is, however, no equivalent to these white spaces in spoken language. The spectrogram shown in Fig. 2.1 illustrates this point. In the absence of a phonetic transcription, it is difficult to identify either how many words there are, or where they might begin. The continuous nature of speech prevents any objective marking of the onsets and offsets of each phoneme. Co-articulation operates both word-internally and across word boundaries. As the beginnings of phonemes cannot be reliably identified, it is impossible to mark word onsets. A trained spectrogram reader is likely to hypothesise phoneme candidates and then word candidates for the utterance in Fig. 2.1 without marking the precise location of any segment or word boundaries.

In the absence of reliable word boundary markers in speech, how might the *lexical segmentation problem* be solved? We suggest that listeners develop specific strategies to deal with this problem. These strategies tend to exploit the

Both authors are now at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

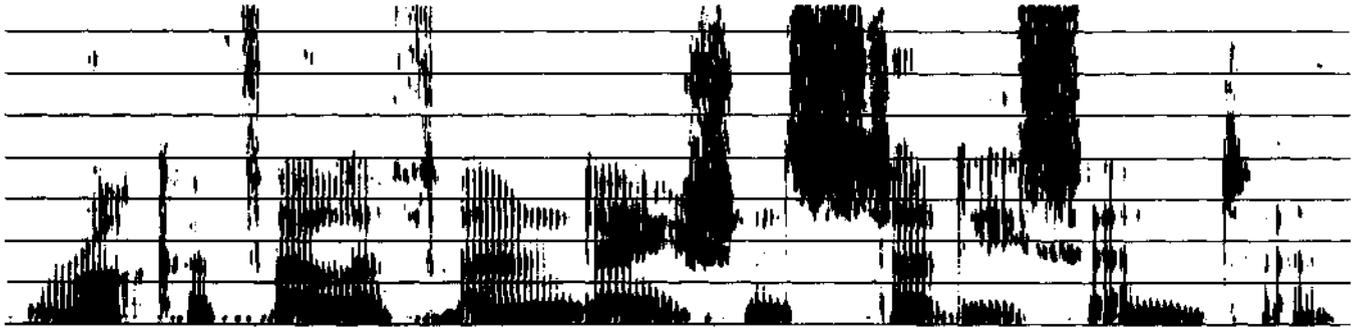


FIG.2.1 A spectrogram of a short phrase, spoken by a native speaker of British English. The phrase is "Rockefeller Foundation Study Centre."

structural regularities of the listener's language, and are therefore likely to be language-specific. In this chapter, we outline a segmentation strategy for English, based on the prosodic structure of the language. Note that prosodic information is not given to the reader of English. The failure of alphabetic scripts to code prosody may be one reason why these scripts have adopted the explicit boundary marker of white space between words.

THE METRICAL SEGMENTATION STRATEGY

The strategy that English listeners apparently use to solve the lexical segmentation problem in speech recognition exploits the rhythmic structure of the English language. English is a stress language, and its rhythm is stress-based. Languages with stress rhythm have, in effect, two distinct syllable types: strong and weak. Strong syllables contain full vowels: weak syllables contain reduced vowels (in English, often schwa). Words like "a-buse", "sa-loon", and "pro-ject" start with a weak syllable followed by a strong one. This distinction is insensitive to lexical stress: For a syllable to be strong it need only contain a full vowel; it does not matter whether this syllable carries the primary stress of a word, or only a secondary stress (e.g. "projectile" contains three syllables; the first is weak, the second is strong with primary stress, the final syllable is also strong with secondary stress).

The "metrical segmentation strategy" exploits stress rhythm in the following way: English listeners assume that every strong syllable is the onset of a new content word. Evidence in support of this claim has come from an analysis of misperceptions of speech (Cutler & Butterfield, 1992), from an examination of the prosodic structure of the English lexicon (Cutler & Carter, 1987), and from laboratory word-recognition tasks (Cutler & Norris, 1988). Here we will focus on the data from misperception, and on the statistical analyses of the vocabulary. First, we will discuss how speech appears to be misperceived exactly as the operation of the metrical segmentation strategy would predict. Secondly, we will show how a strategy of segmenting speech at the onsets of strong syllables is likely to be successful most of the time, because it accurately exploits the structure of the English vocabulary.

The misperceptions of speech that are of relevance to lexical segmentation are those which involve word boundaries. These *juncture misperceptions* are of two kinds. Word boundaries can be erroneously inserted or erroneously deleted. These errors can occur in two places: before strong syllables and before weak syllables. The metrical segmentation strategy predicts that two types of juncture misperception should be relatively common: erroneous insertions before strong syllables (postulating a word boundary before a strong syllable) and erroneous deletions before weak syllables (failing to detect a word boundary before a weak syllable). The strategy also predicts that the other two types of error should be rare. Listeners should tend not to insert word boundaries before weak syllables, nor should they tend to delete boundaries before strong syllables.

Cutler and Butterfield (1992) examined these predictions in two ways. First, they analysed as many spontaneous slips of the ear as they could find. An exploration of several collections of errors yielded 310 juncture misperceptions. These fell into the pattern predicted by the metrical segmentation strategy. Insertions before strong syllables (e.g. "analogy" -> "and allergy") were more common than insertions before weak syllables ("effective" -> "effect of). Deletions before weak syllables ("my gorge is" -> "my gorgeous") were more common than deletions before strong ("is he really" -> "Israeli").

Secondly, juncture misperceptions were induced in the laboratory by presenting listeners with speech that was very faint, i.e. presented at a level (preset for each subject) at which only 50% of words could be correctly identified. Listeners heard six-syllable unpredictable utterances, with alternating patterns of strong and weak syllables, and their task was to write down what they thought each utterance was. Cutler and Butterfield examined the boundary misplacements (in only those responses which contained the correct number of syllables with the correct rhythmic pattern); as in the slips of the ear, there were significantly more boundary insertions before strong than before weak syllables, and significantly more boundary deletions before weak than before strong syllables.

Thus the segmentation errors that listeners make are strongly in accord with the predictions of the metrical segmentation strategy. Moreover, this strategy is well adapted to the structure of English. In an analysis of the MRC Database (a phonetically transcribed word-list based on the *Shorter Oxford English Dictionary*), Cutler and Carter (1987) found that nearly three-quarters of the sample began with strong syllables. Cutler and Carter's count included all word classes. However, as might be expected, the strong-initial tendency is particularly marked for English nouns: Cutler, McQueen, and Robinson (1990) report that nearly 85% of nouns in the *Longman Dictionary of Contemporary English* (Procter, 1975) begin with strong syllables. As nouns are among the least predictable parts of speech, the metrical segmentation strategy will work most efficiently of all where, in effect, it is most needed in speech recognition.

Even more unpredictable than nouns as a whole, perhaps, are proper names for people. Thus it is interesting that the same pattern was found when Cutler et al. (1990) examined the prosodic structure of English first names. However, this additional analysis revealed one intriguing asymmetry. Although names have the same overall pattern as nouns, male names have even fewer weak-initial syllables than nouns, whereas female names have rather more weak-initial syllables than nouns. This results in a significant difference between the two groups of names: Female names are significantly more likely to begin with a weak syllable than male names. This asymmetry was even more marked when frequently used names were considered alone. Lists of the most popular names in Britain and the United States include virtually no male names with weak initial syllables—in other words, not only are there very few such names, but those that exist are very rarely used (e.g. *Sebastian, Demetrius*). On the other hand, about

16% of frequently used female names begin with a weak syllable—so not only are there more weak-initial names for females, but some of these are in very common use (e.g. *Elizabeth, Michelle*). Speculations about the reasons for this asymmetry between male and female names, while fascinating, are outside the scope of this chapter (the reader is referred to Cutler et al., 1990, for such speculations, and for further asymmetries in the phonological structure of the two name classes). Here, the point at issue is that (irrespective of variations *within* the set) the most unpredictable set of words in the English language consists largely of items the onsets of which will be correctly located by the metrical segmentation strategy. Moreover, the more common the item, the more likely it is to have the predicted prosodic form.

Exactly this frequency effect also holds true of the English language in general. Cutler and Carter (1987) found that the mean frequency of the strong-initial content words (nouns, verbs, and adjectives) in the MRC Database was significantly higher than the mean frequency of the weak-initial content words. This predicts that in real speech samples, strong-initial content words will outnumber weak-initial ones to a much greater extent than they do in the vocabulary. Again, this is true. Cutler and Carter analysed a 200,000-word corpus of spontaneous speech (*A Corpus of English Conversation*, Svartvik & Quirk, 1980), and found that only 9.8% of content words began with weak syllables. This statistic suggests that a metrical segmentation strategy that assumes that content words begin with strong syllables will operate successfully over 90% of the time.

Nevertheless, not *all* content words begin with strong syllables; for instance, the preceding sentence alone contains *statistic, suggest, assume, begin* and *successful*. How problematic is it for the metrical segmentation strategy that such words do occur? In a further exercise in lexical statistics, we attempted to address this issue. We examined all the weak-initial content words in the *Longman Dictionary of Contemporary English*. Consider words like *alert* and *assassin*. The metrical segmentation strategy will lead a listener to assume that the second (strong) syllable of a word of this type is the beginning of a new content word. If lexical access were to be initiated from the second syllable of these two words, however, no real words would be activated (that is, both *lert* and *sassin* are nonwords). If no real-word entry is mistakenly accessed, it can be assumed that the listener can readily recover from the misleading segmentation. However, some weak-initial words *do* contain strong-initial words embedded within them. For example, if we remove the initial syllable from prefixed weak-initial words such as *infrequent* and *displeasure*, we obviously obtain the words *frequent* and *pleasure*. But again, it is unlikely that the listener will experience much difficulty with this misleading segmentation either, because at least contact will have been made with a morphologically related lexical entry. Thus for both of these types of weak-initial word—the *alert* type and the *infrequent* type—no unrelated lexical competitors will be erroneously contacted via the operation of the metrical segmentation strategy.

The only words where the strategy would cause difficulty for the listener are those in which removing the initial, weak, syllable produces a totally unrelated strong-initial word. Presentation of such a word would, in effect, lead the listener up a lexical garden path, because a lexical entry would be activated that has nothing to do with the content of the actual utterance. Consider, for example, such words as *vermilion*, *informant* and *contributory*. It is unlikely that a listener's recognition of those words would be facilitated by contact with the lexical entries for *million*, *formant* and *tributary*. We computed, therefore (using the *Longman Dictionary of Contemporary English*), the proportion of the three types of weak-initial content word—the *alert* type, the *infrequent* type, and the *vermilion* type—in the English vocabulary.

Reassuringly, the results of this analysis suggest that lexical garden paths will be rare. As Fig. 2.2 shows, only 17.5% of weak-initial content words contain unrelated embedded words. We can combine this figure with those from Cutler and Carter's (1987) analyses to obtain an estimate of how often the metrical segmentation strategy will lead listeners up a garden path in average speech contexts. Cutler and Carter found that 41% of words in the *Corpus of English Conversation* were content words. Of these, as we mentioned earlier, only 9.8% began with weak initial syllables. From the present computation we can estimate that 17.5% of this weak-initial set were garden-path words. 17.5% of 9.8% of 41% is 0.7%—in other words, the garden-path words constitute less than 1% of words encountered in typical English speech. On this estimate the garden-path effect of words like *vermilion* under the metrical segmentation strategy would appear to be far less of a problem for spoken word recognition than that caused by lexical ambiguity (i.e. the multiple unrelated meanings of words such as *palm* and *bank*).

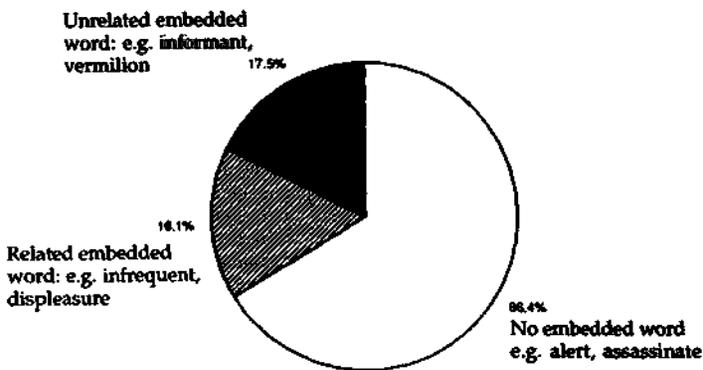


FIG.2.2 The proportion of weak-initial content words in the *Longman Dictionary of Contemporary English* with no embedded word beginning from the second syllable, with related words beginning from the second syllable, and with unrelated words beginning from the second syllable

This result highlights the fact that the metrical segmentation strategy will work very efficiently with the content words of the English language. As implied earlier, however, Cutler and Carter (1987) found that 59% of words in the spontaneous speech corpus were *not* content words. That is, a majority of words in typical English speech are function, or closed-class, words. How does the strategy deal with function words?

Because most function words are monosyllabic, lexical words counted for slightly more than half of the *syllables* in the corpus. Cutler and Carter computed the proportion of strong syllables that were initial or noninitial in both content and function words, and likewise for weak syllables. Nearly 75% of the strong syllables began (or were) content words; more than two-thirds of the weak syllables began (or were) function words. These statistics led them to suggest a refinement of the metrical segmentation strategy. The extended version of the strategy can be stated as follows: Assume that every strong syllable is the onset of a new content word, and also assume that weak syllables are the onsets of function words, but only when there are no other ongoing lexical hypotheses (that is, where the weak syllable must be an onset, assume the onset is of a function word). Cutler and Carter suggested a possible instantiation of this strategy in an algorithm that, given a strong syllable, attempts lexical look-up in a main lexicon of content words. In the absence of a strong syllable (i.e. with a weak syllable), the input is matched against a subsidiary lexicon, consisting of a list of function words.

Describing the segmentation strategy in this way leads to a further prediction for the data from the juncture misperception studies. This prediction concerns erroneous word boundary *insertions* only. Insertions before strong syllables should produce content words, but insertions before weak syllables should produce function words. Cutler and Butterfield (1992) confirmed this prediction, both for spontaneous and for laboratory-induced slips of the ear. When a word boundary was erroneously inserted before a strong syllable, the following word was significantly more likely to be a content word (e.g. *conduct ascents uphill* → *the doctor sends her bill*) than a function word (e.g. *within reviewed results* → *belief to who results*). When a word boundary was inserted before a weak syllable, however, subjects were more likely to misperceive the next word as a function word (e.g. *dusty senseless drilling* → *thus he sent his drill in*) than as a content word (e.g. *an eager rooster played* → *a new resolve again*).

To sum up so far, it appears that listeners use their accumulated knowledge of their own language in solving the lexical segmentation problem. The strategies they develop to make effective use of this knowledge are highly likely to be language-specific—the metrical segmentation strategy, for example, works well for English, but it would be of little use in a language with a markedly different prosodic distribution of content words, and of no use whatsoever in a language without rhythmic stress. In the next section, we consider a different question about how such strategies, exploiting listeners' knowledge of the distributional patterns within their vocabulary, might operate in speech recognition. Specifically, we

consider the issue of whether the mental lexicon is directly involved in the operation of the segmentation process.

THE LEXICON AND SEGMENTATION

There appear to be two distinct ways in which the English vocabulary might be involved in segmentation for lexical access. First, the vocabulary, as instantiated in the mental lexicon of the listener, could be involved on-line in the segmentation process. As we consider the segmentation process to be pre-lexical, acting to initiate lexical access attempts, this amounts to an interactive approach to lexical segmentation. It could be instantiated in an interactive activation framework. The activation strength of the majority of the content words in the lexicon (those with strong initial syllables) could directly modify the segmentation process, increasing the likelihood that a strong syllable will be taken as the onset of a content word.

Secondly, lexical segmentation could be an autonomous process, not subject to on-line control from the lexicon. It is quite conceivable that the segmentation process could take advantage of the structure of the English vocabulary without the lexicon being directly involved in its operation. Listeners, as we have argued, would develop an heuristic such as the metrical segmentation strategy through exposure to the language. As a result, the strategy would exploit the prosodic distribution of the language, and hence involve the structure of the lexicon in each segmentation decision that was made. Nevertheless, the strategy, once developed, could operate pre-lexically and entirely autonomously, on the basis of bottom-up information from the speech signal alone.

There is little evidence that deals directly with this issue. Nevertheless, we favour the second alternative, for two reasons. These will be discussed in turn. First, it is our view that the weight of the evidence from tasks requiring phonetic decisions favours an autonomous view of speech processing (Burton, Baum, & Blumstein, 1989; Cutler, Mehler, Norris, & Segui, 1987; Frauenfelder, Segui, & Dijkstra, 1990; McQueen, 1991). Cutler et al. (1987) have argued that the variability of lexical involvement in the phoneme monitoring task can be more parsimoniously explained by an autonomous model of spoken word recognition than an interactive model such as the TRACE model of McClelland and Elman (1986). Frauenfelder et al. (1990) failed to find support for the interactive predictions that a model like TRACE makes for phoneme monitoring in nonwords. The model predicts that the top-down facilitation from the lexicon that speeds detection of, for example, /p/ in the word *olympiade* relative to /p/ in the matched nonword *arimpiako* should also have inhibitory consequences. Detection of /t/, for example, in **the** nonword *vocabulaire* should be inhibited, because the lexical entry *vocabulaire* should boost the activation of the /l/ phoneme-node, which in turn should inhibit the /t/ node (in the TRACE architecture, the phoneme-nodes compete). But Frauenfelder et al. found no difference between detection times for targets such

as /t/ in *vocabulaire* and matched nonwords such as *socabulaire* (where there should be no top-down facilitation, and no resultant inhibition).

In addition to the phoneme-monitoring task, results from the phonetic categorisation task also favour an autonomous view of speech processing. Burton et al. (1989) found that the lexical shift in the phonetic categorisation of word-initial stops, originally reported by Ganong (1980), depends on the acoustic-phonetic quality of the materials. Listeners were asked to categorise a voice onset time (VOT) continuum ranging from /d/ to /t/, embedded in word-nonword (*duke-tuke*) and nonword-word (*doot-toot*) contexts. They were more likely to label ambiguous stops as /d/ when the voiced endpoint was a word (*duke*), and as /t/ when the unvoiced endpoint was a word (*toot*). But Burton et al. found that this lexical effect disappeared when the voicing continuum contained more acoustic-phonetic information. There was no lexical shift with a continuum where the burst and aspiration amplitude was manipulated in parallel with VOT. McQueen (1991) replicated this effect in word-final categorisation. Lexical effects should be larger and more robust word-finally because there is more time for lexical access before the final phoneme is heard. Listeners tended to label ambiguous fricatives in an /s/-/ʃ/ continuum according to the lexical status of the string: For example, ambiguous stimuli were more often categorised as /ʃ/ in the continuum *fish-fiss* than in the continuum *kish-kiss*. But this lexical shift only occurred when the fricatives were degraded by low-pass filtering.

As top-down processing in interactive models is considered to benefit speech perception, these demonstrations that lexical effects are not a mandatory feature of the perception of ambiguous phonemes (where one would predict top-down processing to be most in evidence) pose a problem for interactive models. McQueen (1991) also found that a specific reaction time prediction of the interactive TRACE model was not supported by the word-final categorisation data. Interactive models such as TRACE, which are based on activation, predict that top-down facilitation increases over time, as the evidence for a particular word accumulates. TRACE therefore predicts that the lexical shift in word-final categorisation should be larger for slower responses. Reaction times were split into three ranges, fast, medium, and slow, for each subject, following a procedure developed by Miller and Dexter (1988). Contrary to TRACE predictions, it was found that the lexical shift was largest in the fast RT range.

We have recently submitted these data to a more stringent test. Dividing each subject's overall data into three ranges ignores any variability in reaction time along the fricative continuum. Because responses to endpoint stimuli were faster than those to ambiguous region stimuli this range analysis will have tended, in the fast range, to increase the number of data points at the endpoints and decrease the number near the boundary. Conversely, in the slow range, there will have been relatively more data near the boundary and relatively less at the endpoints. Because the results suggest that there are lexical shifts in the fast range, and as these shifts are due to performance in the boundary region, it seemed important to attempt to

TABLE 2.1
Estimated 50% Crossover Points for Palatal Responses, as a Function of Reaction Time Range

<i>Continua</i>	<i>Fast Range</i>	<i>Medium Range</i>	<i>Slow Range</i>
Word-nonword	6.3	5.9	5.9
Nonword-word	5.9	5.7	5.7
Nonword-nonword	6.4	5.8	5.8

Note: The three ranges were based on each subject's data for each point along the continuum.

Source: Data from McQueen, 1991.

maximise the amount of data in the boundary region in the fast RT range. To this end, we performed the new RT range analysis. The data was again divided into three reaction time ranges for each subject, but now this analysis was performed separately for each of the nine fricatives along the continuum. This technique removes any variability in reaction time due to stimulus continuum position, and equates the amount of data contributing to each of the cells of the analysis.

Estimates of 50% crossover points were made by linear interpolation for each of the continuum types in each RT range. The means are given in Table 2.1. The percentages of palatal responses in the three RT ranges, with the data divided in this way, for each point along the continuum for each subject, are plotted in Fig. 2.3. The mean cut-off was 982msec (SD = 114msec) between fast and medium ranges and 1143msec (SD = 146msec) between medium and slow ranges.

The results replicated the original range analysis. There was a significant effect of continuum only in the fast range: $F(2,42) = 8.79$, $p < 0.001$. This was partially due to a lexical shift. In an analysis comparing the word-nonword and nonword-word continua only, there were reliably more /// responses when the palatal endpoint was a word than when this endpoint was a nonword: $F(1,21) = 7.83$, $p < 0.05$. Combining the old and new analyses, it appears that the overall lexical shift is due to an effect that is largest in the fastest RT range and smallest in the slowest range. The shift is only statistically reliable in the fast range, whether the ranges are defined from each subject's overall response times or from each subject's responses to each of the nine fricatives.

This analysis therefore confirms that, irrespective of variability between average RTs to endpoint vs. ambiguous stimuli, it is the fast responses that show the greatest lexical shifts. As we pointed out earlier, TRACE would have predicted the opposite pattern of results. Thus McQueen's study joins the body of recent research that has demonstrated weakness in the interactive account of speech processing. On the other hand, McQueen's data do not challenge an autonomous model. Frauenfelder et al.'s data, likewise, are problematic for an interactive model, but not for an autonomous model, because autonomous theory predicts that the perception of nonwords should be insulated from lexical processing. The autonomous race model (Cutler & Norris, 1979; Cutler et al., 1987), as an example,

predicts the pattern of RT range results that McQueen (1991) found for word-final categorisation. In this model, lexical and pre-lexical phoneme identification procedures are considered to race; the procedure which more rapidly produces an output on a particular trial will be responsible for the phonetic decision on that trial. Because in word-final categorisation the lexical route is faster on average

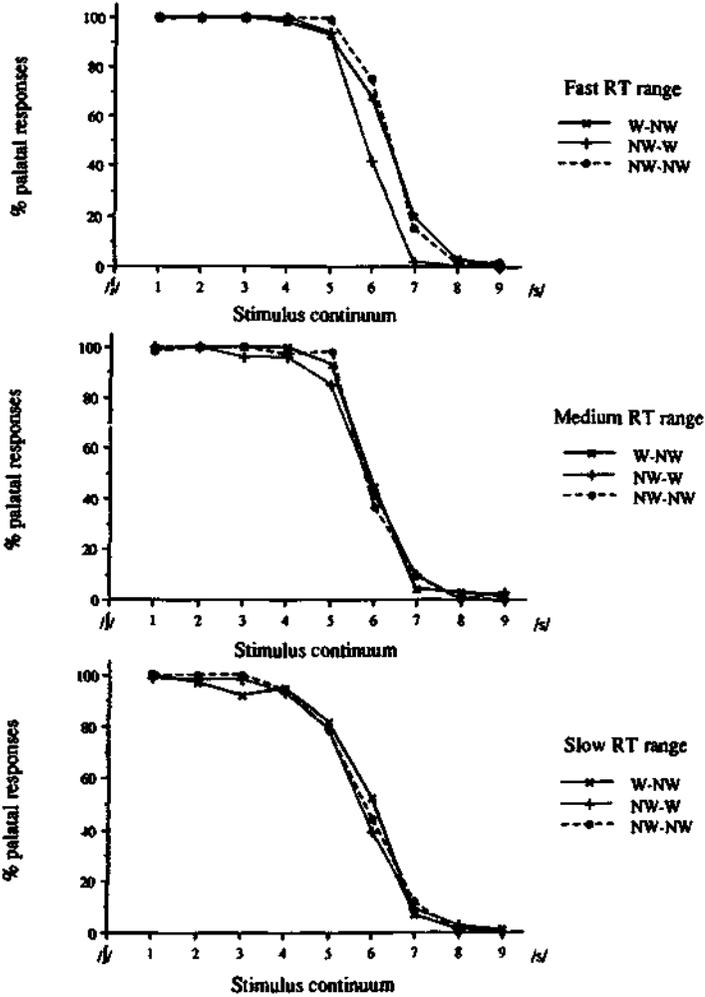


FIG.2.3 Percent palatal (/f/) responses, divided into three reaction-time ranges: fast, medium, and slow, defined for each point along the continuum for each subject. The data are plotted, collapsed across the three vowel blocks, as the proportion of the total number of responses to each stimulus within each range (data from McQueen, 1991).

than the pre-lexical route, the responses producing the lexical shift (i.e. those made via the lexical route) will tend to be the faster responses.

On the basis of evidence from tasks requiring explicit phonetic decisions, then, it would appear that autonomous theories of speech processing are more plausible than interactive theories. This might suggest in turn that a strictly autonomous view of lexical segmentation should be preferred. The inference from the phoneme-monitoring and phonetic categorisation data to the normal segmentation of continuous speech is, however, not a strong one. Nevertheless, the second line of evidence that favours an autonomous view of segmentation is based on the operation of the segmentation process itself. We return to the juncture misperceptions reported by Cutler and Butterfield (1992).

If the lexicon is directly involved in lexical segmentation, its influence should be reflected in the boundary error data: Asymmetries in the vocabulary should be correlated with asymmetries in the errors. This hypothesis could be relatively easily tested, because of a particular characteristic of the materials used in the faint speech experiment described earlier. In these materials, it will be recalled, half the vowels were weak and half were strong. The weak vowels were mostly schwa, which indeed is usually the case for weak vowels in English. However, although the English phoneme repertoire contains over a dozen strong monophthong vowels, plus several diphthongs, the experimental materials were constrained such that they contained only six strong vowels (/e, eI, A, u, I, i/). Cutler and Carter's (1987) statistics on the predominance of strong initial syllables in English indicate that for *any* full vowel, there are likely to be more words with this vowel in a strong initial syllable (for example, with the vowel /e/, words like *beg, chest, feather, residence, verisimilitude*) than with the same vowel in a strong second syllable, preceded by a weak syllable (e.g. *cadet, forgetful, suggestible, togetherness*). However, this ratio is likely to vary from vowel to vowel.

If the lexicon directly influences boundary placement, as an interactive account of speech recognition would suggest, then the larger the strong-initial to weak-initial ratio is for a particular vowel, the more likely a listener should be to misperceive a weak-initial word with that vowel as a strong-initial one. In other words, the frequency across vowels of boundary insertion errors in the faint speech experiment (i.e. errors in which a strong syllable is erroneously reported as word-initial) should directly reflect the size of the strong-initial to weak-initial ratio across vowels. A correlational analysis described in Cutler and Butterfield (1992) provided a test of this prediction. Using once again the *Longman Dictionary of Contemporary English*, we computed, for each of the six full vowels used in the faint speech experiment, the ratio of strong-initial words (*feather, etc.*) to weak-initial (*cadet, etc.*). As predicted, this ratio was quite variable across the six vowels. We then computed the ratio of boundary insertions to boundary deletions before strong syllables containing each of these vowels. As Fig. 2.4 shows, the two ratios appeared to be quite uncorrelated (indeed, their relation was, if anything, in the wrong direction: $r(5) = -0.31$, $p > 0.5$).

An interactive account of the role of the lexicon in speech segmentation would also predict another correlation between lexical distribution and missegmentation frequency—this time for *deletion* errors. Deletion errors, it will be recalled, occur far more often before weak syllables than before strong ones—that is, boundaries before weak syllables tend to be overlooked. The effect of such a deletion error is, of course, to attach the weak syllable to a preceding strong syllable. If the lexicon is involved in boundary errors, such deletion errors should also reflect lexical distributions; specifically, they should be more common for those vowels that most often occur with a following weak syllable. We assumed that the ratio of words with each vowel in a penultimate syllable followed by a weak final syllable (e.g. *feather, forgetful*) to words with the vowels in a final strong syllable (e.g. *chest, cadet, comprehend*) would also vary across the six vowels; and again, we computed the relevant statistics using the *Longman Dictionary*, and found that it did. The prediction from the interactive account is that the larger this ratio is for a particular vowel, the more likely it should be that listeners will erroneously overlook a word boundary after that vowel. We performed a correlational analysis and again found that the correlation was not significant: $r(5) = 0.71, p < 0.1$. Figure 2.5 shows the data.

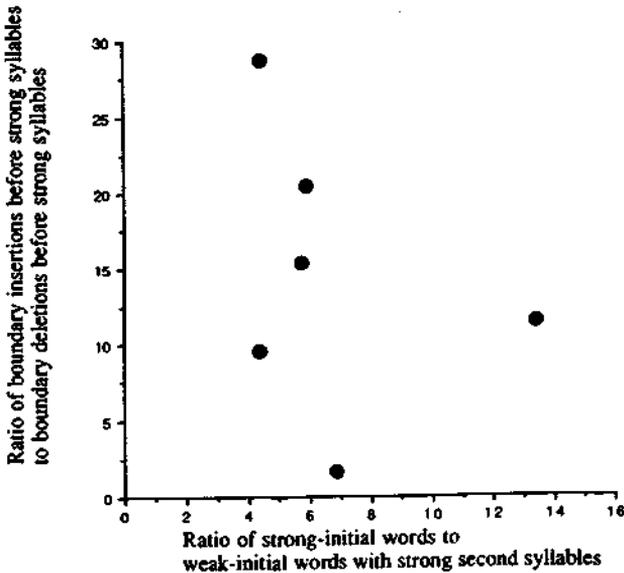


FIG.2.4 A scatter plot, for six vowels {/e, el, A, u, I, i/}. The ratios of words with each vowel in strong initial syllables to words with weak initial syllables followed by strong syllables containing the vowel are taken from the *Longman Dictionary of Contemporary English*. These are plotted against juncture misperception errors taken from Cutler and Butterfield (1992): The ratio of boundary insertions to boundary deletions before the strong syllables containing each of the six vowels.

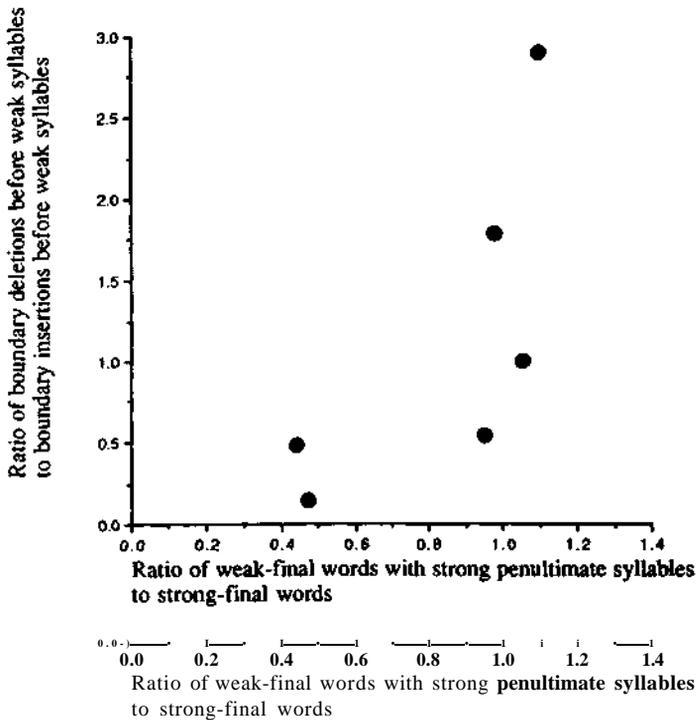


FIG. 2.5 A scatterplot, for six vowels (t, e, l, A, u, I, i). The ratios of words with each vowel in penultimate strong syllables to words with final strong syllables containing the vowel are taken from the *Longman Dictionary of Contemporary English*. These are plotted against juncture misperception errors taken from Cutler and Butterfield (1992): The ratio of boundary deletions to boundary insertions before weak syllables following strong syllables containing each of the six vowels.

The predictions of an interactive view of lexical segmentation have, therefore, not been confirmed. Of course, these correlational analyses can not be held to provide conclusive support for an autonomous position. Nevertheless, these results, in combination with the findings from the phonetic decision tasks, indicate that an autonomous account of lexical segmentation may be in several respects more plausible than an interactive account.

CONCLUSIONS

We have argued that in order to solve the lexical segmentation problem, listeners make use of the structural properties of their language. In English, they exploit the distributional properties of the vocabulary, as indicated by lexical statistics. We have claimed that listeners adopt a metrical segmentation strategy. They assume that strong syllables are the beginnings of content words. If listeners have to take weak syllables as onsets, they treat them as the onsets of function words.

Estimates suggest that in typical English speech, the metrical segmentation strategy would cause misleading lexical entries to be accessed for less than 1% of words.

The continuous temporal nature of speech is best suited to a bottom-up analysis. Evidence from paradigms such as phonetic categorisation indicates that the lexicon does not influence pre-lexical processing. This suggests that the lexicon is not involved on-line in segmentation. Converging evidence for this claim comes from a correlational analysis of juncture misperceptions. What appears to be most plausible is that listeners learn to use a metrical segmentation heuristic through experience with their language, and that this process then operates pre-lexically.

ACKNOWLEDGEMENTS

We acknowledge financial support from the Joint Research Councils Initiative in HCI and Cognitive Science: Grant No. E304/148. We would also like to thank the Longman Group UK Inc. for allowing us to use a machine-readable version of the *Longman Dictionary of Contemporary English*.

REFERENCES

- Burton, M.W., Baum, S.R., & Blumstein, S.E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 567-575.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.
- Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- Cutler, A., McQueen, J., & Robinson, K. (1990). Elizabeth and John: Sound patterns of men's and women's names. *Journal of Linguistics*, 26, 471-482.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141-177.
- Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- Frauenfelder, U.H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, 16, 77-91.
- Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.
- McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- McQueen, J.M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 433-443.
- Miller, J.L., & Dexter, E.R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 369-378.
- Procter, P. (Ed.) (1975). *Longman Dictionary of Contemporary English*. London: Longman.
- Svartvik, J. & Quirk, R. (1980). *A Corpus of English Conversation*. Lund: Gleerup.