Chapter 5

Exploiting Prosodic Probabilities in Speech Segmentation

Anne Cutler

Recognizing Continuous Speech

Shillcock (this volume) has shown that listeners hearing the word *trombone* in a sentence momentarily entertain the hypothesis that they might be hearing the word *bone.* Why does this happen? Surely speech recognition would be more efficient if listeners accessed only the lexical representations of words that really occurred and not also words that might be embedded within occurring words?

It is the argument of this chapter that Shillcock's finding arises naturally from a strategy used by listeners to deal with the problems of speech segmentation. The essence of the segmentation problem is that word boundaries in continuous speech are not reliably marked. Recent studies of human speech processing have suggested that listeners may use heuristic strategies for overcoming the absence of word-boundary information. Such strategies may allow listeners to guide their attempts at lexical access by postulating word onsets at what linguistic experience suggests are the most likely locations for word onsets to occur.

Cutler and Norris (1988) have proposed such a strategy based on metrical structure. In a stress language like English, syllables can be either strong or weak. Strong syllables contain full vowels, while weak syllables contain reduced vowels (usually a schwa). Cutler and Norris found that hsteners were slower to detect the embedded real word in *mintayf (in* which the second vowel is strong) than in *mintef* (in which the second vowel is schwa). They suggested that listeners were segmenting *mintayf* prior to the second syllable, so that detection of *mint* therefore required combining speech material from parts of the signal that had been segmented from one another. No such difficulty arose for the detection of *mint* in *mintef,* since the weak second syllable was not segmented from the preceding material.

Cutler and Norris proposed that in continuous speech recognition in English, listeners generally approach the problem of segmentation for lexical access by applying a metrical segmentation strategy (MSS): strong syllables are taken as likely lexical (or content) word onsets, and the continuous speech stream is segmented at strong syllables so that lexical-access attempts can be initiated. This explains why *bone,* even when it is embedded in *trombone,* should be momentarily considered to be a possible new word: *bone* is a strong syllable.

The success rate of such a strategy depends, of course, on how realistically it reflects the structure of the language. Hypothesizing that strong syllables are likely to be lexical word onsets and that weak syllables are not will only prove to be an efficient strategy for detecting actual word onsets if most lexical words actually begin with strong syllables and not with weak syllables. As the next section shows, the MSS is indeed well adapted to the characteristics of the English vocabulary.

Assessing **Prosodic Probabilities for English**

To estimate the success rate of the MSS, Cutler and Carter (1987) examined the metrical structure of word-initial syllables in English. First they looked at the metrical structure of words in the English vocabulary. The MRC Psycholinguistic Database (Coltheart 1981, Wilson 1988) is a lexicon of over 98,000 words and is based on the *Shorter Oxford English Dictionary.* Over 33,000 entries have phonetic transcriptions. Figure 1 shows the metrical characteristics of the initial syllables of the transcribed words in this lexicon divided into four categories: monosyllables (such as *bone* and *splint),* polysyllables with primary stress on the first syllable (such as *lettuce* and *splendour),* polysyllables with secondary stress on the first syllable (such as *trombone* and *psychological),* and polysyllables with weak initial syllables (in which the vowel in the first syllable is usually schwa, as in *averse* and *trapeze,* but may also be a reduced form of another vowel, as in *invest* and *external).* Words in any of the first three categories satisfy the MSS. It can be seen that these categories together account for 73 percent of the words analyzed.

In English the most common word *type* (as opposed to token) is clearly a polysyllable with initial stress. However, individual word types differ in the frequency with which they occur. Frequency-of-occurrence statistics (Kucera and Francis 1967) are listed in the MRC Database, and Cutler and Carter found that the mean frequency for the four metrical word categories did indeed differ. First, monosyllables occur on average far more

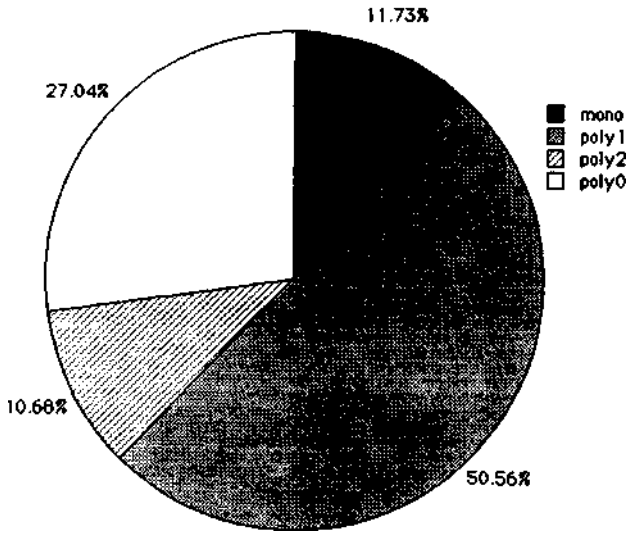

Figure 1
Metrical structure of the initial syllable of words in the MRC Psycholinguistic Database (mono = monosyllabic words; poly 1 = polysyllabic words with initial primary stress; poly 2 = polysyllabic words with initial secondary stress; poly 0 = polysyllabic words with weak initial syllable).

frequently than any type of polysyllable. (Note that to analyze frequency of occurrence, Cutler and Carter considered only the lexical, or content, words in the database and excluded the grammatical, or function, words which accounted for less than 1 percent of the phonetically transcribed words. These were overwhelmingly monosyllabic and of high frequency; their inclusion would have inflated the mean frequency of monosyllables still further.) Second, within the set of polysyllables, words with strong initial syllables occur somewhat more frequently than words with weak initial syllables. If the type counts in figure 1 are multiplied by their mean frequencies, one can estimate that although there are more than seven times as many polysyllables in the language as there are monosyllables, average speech contexts are likely to contain almost as many monosyllables as polysyllables. Moreover, only about 17 percent of lexical tokens in most speech contexts will begin with weak syllables.

Cutler and Carter tested this estimate against a natural speech sample, the *Corpus of English Conversation* (Svartvik and Quirk 1980), using the frequency count of this corpus prepared by Brown (1984). The London-Lund corpus consists of approximately 190,000 words of spontaneous

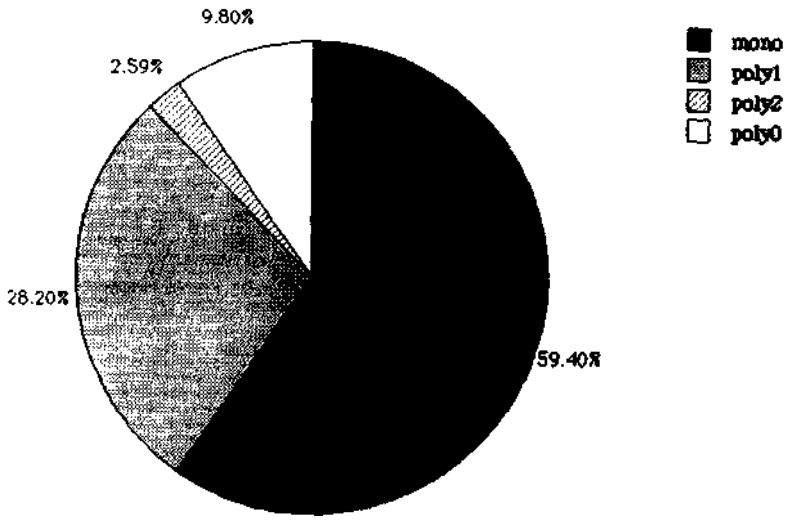

Figure 2
Metrical structure of the initial syllable of lexical words in the *Corpus of English Conversation* (mono = monosyllabic words; poly 1 = polysyllabic words with initial primary stress; poly 2 = polysyllabic words with initial secondary stress; poly 0 = polysyllabic words with weak initial syllable).

British English conversation. Figure 2 shows the distribution of metrical categories for lexical words in this corpus. The three categories with strong initial syllables account for 90 percent of the tokens; only 10 percent of the lexical words have weak initial syllables.

Although figure 2 covers all the lexical words in the London-Lund corpus, it actually accounts for only 41 percent of all words in the sample; the majority of words in the corpus are grammatical words. But because hardly any grammatical words have more than one syllable, figure 2 nevertheless accounts for 51 percent of all *syllables.* In fact, with some reasonable assumptions it was possible to compute the probable distribution of syllables in this speech sample. Cutler and Carter assumed that grammatical words such as *the* and of were in general realized as weak syllables. If so, the most likely distribution of syllables is that given in table 1. It can be seen that about three-quarters of all strong syllables in the sample were the sole or initial syllables of lexical words. Of weak syllables, however, more than two-thirds were the sole or initial syllables of grammatical words.

Thus a listener encountering a strong syllable in spontaneous English conversation seems to have about a three to one chance of finding that

Table 1
Strong (full) syllables versus weak (reduced) syllables in the *Corpus of English Conversation*

	Strong	Weak
Sole or initial syllable of lexical word	74%	5%
Noninitial syllable of lexical word	12%	23%
Sole or initial syllable of grammatical word	11 %	69%
Noninitial syllable of grammatical word	3%	3%
Total number of syllables	93,989	145,888
Percentage of syllables in corpus	39%	61 %

strong syllable to be the onset of a new lexical word. A weak syllable, on the other hand, is most likely to be a grammatical word. It seems, therefore, that English speech indeed provides an adequate basis for the implementation of a segmentation strategy such as Cutler and Norris's MSS, whereby strong syllables are assumed to be the onsets of lexical words.

Testing the Performance of the Metrical Segmentation Strategy

Cutler and Carter tentatively proposed some characteristics of an algorithm for implementing the MSS. In their proposal they suggested that listeners might not only assume strong syllables to be the onsets of lexical words but also take into account the likely distribution of weak syllables. The algorithm in outline has six steps:

1. Assume separate lexical (L) and grammatical (G) lists.
2. If the initial syllable of the input is weak, go to the G list. If it is strong, go to the L list.
3. The lookup process in each list returns the longest candidate consistent with the input up to a strong syllable.
4. Occurrence of a strong syllable terminates the current lookup process and initiates a new L lookup.
5. If either lookup fails, the input is submitted to the other list.
6. If both lookups fail, backtracking is necessary; that is, a previous decision must be canceled (e.g., by accepting a shorter candidate word, by undoing the word assignment of the previous syllable and attaching it to the current input, by continuing the current lookup process into a following strong syllable, etc.).

The performance of this algorithm on the London-Lund corpus can only be assessed by considering all words in context, which, in view of the size

of the corpus, was impracticable. Cutler and Carter therefore created a minicorpus specifically to test the algorithm. A native speaker read onto tape a 97-word passage, and phonetically trained listeners noted which syllables were strong and which were weak. The algorithm given above performed extremely well, assigning 82 percent of all words (including 92 percent of lexical words) to the correct list on the first pass.

Cutler and Carter did not compare the performance of the MSS with that of other strategies. However, a 1989 study by Briscoe did undertake such a comparison. In this study the performance of three other routines for generating lexical hypotheses in continuous-speech recognition was compared with the performance of the MSS. The basis of Briscoe's comparison was the number of partial lexical hypotheses that the four strategies generated.

The three comparison strategies were loosely based on other existing proposals in the psycholinguistic literature.

1. Lexical-access attempts were initiated at each new phoneme.
2. Lexical access was tried at each syllable boundary.
3. Lexical access was initiated at sentence onset and subsequently at the conclusion of each successful lexical access. This amounts to a word-by-word segmentation, which is assumed in many automatic systems for continuous-speech recognition.
4. Lexical-access attempts were initiated at the onset of each strong syllable. This strategy was based on the specific proposals for the implementation of the MSS in Cutler and Carter 1987 and listed above.

Each strategy was implemented, and their respective performances were compared on a string of phonetic segments transcribing one sentence from the 97-word passage produced for Cutler and Carter's test of the MSS. Three transcription levels were used:

a. A fine-class transcription, in which each phoneme was explicity identified

b. A fine-class transcription of strong syllables with a broad-class transcription of weak syllables into such broad categories as vowel, stop consonant, nasal, etc.

c. A midclass transcription of strong syllables into more constrained categories such as voiced stop, back vowel, etc., again with a broad-class transcription of weak syllables

At transcription level (a), strategies 3 and 4 produced noticeably fewer lexical hypotheses than strategies 1 and 2, but this is only to be expected,

since if all segmental information is available, the number of lexical hypotheses is simply a function of the number of segmentation points. The more frequent the segmentation points (phonemes or syllables versus strong syllables or words), the more hypotheses will be generated. In particular, the performance of strategies 3 and 4 at level (a) was very similar, even in their errors (both, for instance, treated *rainbow* initially as two words). Performance of the four strategies at transcription levels (b) and (c), however, is a more interesting comparison, since these levels of accuracy arguably offer a more realistic approximation of the likely information available to any recognizer. And it is at these levels that the greatest difference between the strategies appeared: at levels (b) and (c), strategy 4 performed much better than all three of the other strategies, producing significantly fewer partial lexical hypotheses. Interestingly, strategy 3, the word-by-word segmentation routine, which seems superficially to be the most common sense approach, produced an enormously increased number of hypotheses at level (c). Note that Harrington and Johnstone (1987) have computed that most English sentences of reasonable length allow millions of possible parses with broad- or middle-class phonetic transcription. At level (c) strategy 3 in fact produced very many more potential parses than did strategy 2, which includes the constraint that new words can only begin at syllable boundaries. This suggests that some constraint on which segment boundaries potentially begin words is virtually indispensible. A further aspect of Briscoe's study is that at level (c) the MSS-based strategy 4 was tested in two versions: one in which there was a single lexicon and segmentation was attempted only at the onsets of strong syllables, and one in which the lexicon was split into separate lexical and grammatical word lists and a weak syllable was initially looked up in the latter list. This second version performed best of all.

Thus on Briscoe's metric of counting lexical hypotheses (which amounts to an assessment of wasted effort in speech segmentation), the MSS is particularly well adapted to dealing with continuous spoken English and is more robust than alternative strategies in coping with the effects of reduction of fine-grain information. Since such reduced information might be argued to be all that the recognizer has to work with in many speech situations, it appears, at least from this limited study, that the MSS is the most realistic of the strategies Briscoe contrasted.

These tests strongly indicate that the MSS is a realistic strategy and should perform well on continuous spoken English. Some relevant evidence from human speech recognition is described in the next section.

Predicting Listeners' Segmentation Performances

Cutler and Norris's 1988 experiment, which motivated their proposal of the strategy, presented listeners only with nonsense bisyllables. Some evidence that listeners may indeed use the strategy in the segmentation of continuous speech was subsequently produced. This evidence comes from segmentation errors, the way in which word boundaries tend to be misperceived.

The absence of reliable correlates of a word boundary makes misperception of the location of a word boundary in speech easy in principle. Butterfield and Cutler (1988) examined listeners' misperceptions of continuous speech in the light of Cutler and Norris's proposed MSS. If listeners are indeed assuming strong syllables to be word-initial and weak syllables to be non-word-initial, word boundary misperceptions should be very unequally distributed across the four possible types of errors. Specifically, erroneous insertion of a boundary before a strong syllable and erroneous deletion of a boundary before a weak syllable should prove to be relatively common, whereas erroneous insertion of a boundary before a weak syllable and erroneous deletion of a boundary before a strong syllable should be relatively rare. Butterfield and Cutler examined both spontaneous and experimentally elicited misperceptions.

Psycholinguists have for many years collected and analysed the slips of the ear that occur in conversation, and in fact, many of these contain word-boundary misplacements. Butterfield and Cutler examined all the errors listed in published studies of slips of the ear (Bond and Garnes 1980; Browman 1978, 1980; Celce-Murcia, 1980; Garnes and Bond 1975, 1980) plus all the slips of the ear included in a speech error collection that I had assembled over several years. Among these slips, over one hundred involved misplacement of a word boundary across at least one syllabic nucleus. (We excluded errors in which a boundary was misplaced across only one or two consonants such as *up with Anne -> up a fan* because they are irrelevant to the hypothesis about metrical syllable structure.) Some slips in fact involved more than one misplaced boundary (such as *for an occasion -> fornication).*

Some examples of errors are shown in table 2. Butterfield and Cutler found in this set of naturally occurring errors precisely the pattern predicted by the MSS: insertions of a word boundary before a strong syllable *(disguise -> he skies)* and deletions of a word boundary before a weak syllable *(ten to two -> twenty to)* outnumbered by more than two to one

Table 2
Slips of the ear

Coke and a Danish	->	Coconut Danish
it was illegal	->	it was an eagle
ten to two	->	twenty to
disguise	->	the skies
reverse	->	your purse
my gorge is rising	->	my gorgeous …
by tonight	->	butter knife
she'll officially	->	Sheila Fishley
she's a must to avoid	->	she's a muscular boy
variability	->	very ability
in closing	-------->	enclosing
effective	---------->	effect of
paint your ruler	->	paint remover

insertions of a boundary before a weak syllable *{variability -> very ability)* or deletions of a boundary before a strong syllable (in *closing -> enclosing).*

However, Butterfield and Cutler found that the contextual information available for these errors was insufficient to determine what opportunities the listeners had had for word-boundary misplacement. Thus the statistical significance of the asymmetric distribution of the natural slips was impossible to ascertain. It was possible, though, to carry out a statistical comparison of the relative frequency of the words that were actually spoken versus the words that were erroneously perceived. After all, it may simply be the case that when listeners are presented with an utterance that for some reason is difficult to perceive, they reconstruct a plausible version. In this case the distribution of word-boundary misperceptions across strong and weak syllables may simply fall out of the fact that, as Cutler and Carter (1987) showed, words with strong initial syllables tend to have a higher frequency of occurrence than words with weak initial syllables. Of course, this frequency analysis was not simple to perform. First, many of the slips of the ear involved proper names, the frequency of which is impossible to assess. Second, grammatical words such as *the* and of have such a high frequency of occurrence that any error that includes a grammatical word not in the target utterance will necessarily have a higher mean frequency of occurrence than the target, whereas any error omitting a grammatical word present in the target will necessarily have a lower mean frequency of occurrence than the target. However, it seems reasonable to suppose that

if frequency effects are operative, they should show up in the lexical words analyzed separately from the grammatical words. For instance, it would not seem particularly surprising were *She wants a cot and blanket* to be heard as *She wants a cotton blanket,* and although the mean frequency of *cot* and *and* is higher than the frequency *of cotton,* it is surely more relevant that the frequency *of cot* alone is lower than the frequency *of cotton.* Thus Butterfield and Cutler simply compared the frequency of lexical words in targets and errors.

The results of the frequency analysis showed, unsurprisingly, that there was a general tendency for word-boundary insertions to result in errors containing higher-frequency words than the target and for word-boundary deletions to result in errors containing lower-frequency words than the target. This is unsurprising because boundary insertions are likely to produce a percept containing shorter words, while boundary deletions are likely to produce a percept containing longer words, and as is well known, shorter words tend to be more frequent than longer words. This predictable effect is less important than the fact that less than half of the errors overall contained higher-frequency words than their targets. Overall there was *no* significant tendency for errors to contain higher-frequency words than targets. Moreover, there was no significant difference in the nature of the frequency effect between the two types of errors predicted by the MSS and the two types of errors not predicted.

Thus the evidence from spontaneous slips of the ear suggests that listeners do indeed rely on a strategy of assuming that strong syllables begin words. However, slips of the ear occur infrequently and are difficult to collect. As noted above, they are also difficult to analyze in many ways. Therefore, Butterfield and Cutler followed up their analysis of spontaneous misperceptions with an experiment involving deliberately induced misperceptions. In this study, unpredictable utterances (e.g., "achieve her ways instead") were presented to listeners at a level minimally above their threshold for speech reception (which was determined separately for each listener in an extensive pretest). The subjects' task was to write down what they thought was said.

Some sample responses are listed in table 3. Excluding responses that were entirely correct, consisted of no responses, or consisted of only a few syllables, those responses that preserved the number of syllables (six) in the target utterance comprised nearly half of the responses. Of these 40 percent contained word-boundary misplacements. Some responses contained more than one boundary misplacement, so the total number of errors

Table 3
Example responses to faint speech

Stimulus	Responses
achieve her ways instead	a cheaper way to stay
	the chief awaits his men
soon police were waiting	soon the beast will waken
	soon to be awakened
conduct ascents uphill	the doctor sends her bill
	conduct a sense of ill
sons expect enlistment	some expect a blizzard
	sons expected missing
dusty senseless drilling	dust is senseless ruin
	thus he sent his drill in

available for analysis was 257. The distribution of these errors across the four possible error classes is shown in table 4. It can be seen that exactly the pattern predicted by the proposed strategy emerges: erroneous insertions of word boundaries before strong syllables and deletions of word boundaries before weak syllables greatly outnumber insertions of boundaries before weak syllables or deletions of boundaries before strong syllables.

Because the opportunities for each type of error could be determined exactly in this case, the difference could be evaluated statistically. Butterfield and Cutler found that it was indeed significant. Moreover, analysis of only the first missegmentation in each response (on the grounds that later word choices to a certain extent follow from earlier choices) revealed the same pattern—far more insertions before strong syllables than before weak and far more deletions before weak syllables than before strong—with the same level of statististical significance. And once again a comparison of the frequency of lexical words in the targets and in the errors showed no overall preference for higher-frequency responses and no significant difference in frequency effects across the responses that were predicted by the strategy and those that were not.

Note that this lack of a frequency effect is here, as with the spontaneous slips of the ear, strong evidence against any interpretation of the pattern of results in terms of simple plausibility of responses. If subjects had simply been choosing likely responses, their responses would have tended to be of higher frequency than the (improbable) stimuli; they were not. Moreover, it is also evidence against simple random choices of words as responses, since the skew in the frequency distribution of the English vocabulary is

Table 4
Frequencies of boundary misplacements in response to faint speech

Boundary misplacement	No. of occurrences
Insertion before a strong syllable *(sons expect enlistment -> some expect a blizzard)*	144
Deletion before a weak syllable *(achieve her ways instead -> a cheaper way to stay)*	52
Total no. of misplacements predicted by the MSS	196
Deletion before a strong syllable *(soon police were waiting -> soon to be awakened)*	13
Insertion before a weak syllable *(dusty senseless drilling -> thus he sent his drill in)*	48
Total no. of misplacements not predicted by the MSS	61
Total no. of misplacements	257

such that random choices predict that responses should have tended to be of *lower* frequency than the stimuli; again, they were not.

One further characteristic of the error pattern in this experiment is worthy of note. Although word-boundary insertions before weak syllables, which are predicted to be relatively uncommon, are indeed the second rarest type of error, they nevertheless occur four times as often as the rarest type of error, boundary deletions before strong syllables. From Cutler and Carter's examination of natural speech, one can predict the prosodic probabilities of weak syllables and hence the way they are most likely to be misperceived. In the spontaneous speech corpus that Cutler and Carter examined, more than two-thirds of all weak syllables were monosyllabic grammatical words. Thus one might predict that a weak syllable in faintly perceived speech is most likely to be perceived as a monosyllabic function word. A subsidiary prediction about the misperception data might then be that erroneous insertions of word boundaries before weak syllables should tend to involve erroneous reports of monosyllabic function words.

This is indeed the case. Exactly two-thirds of the boundary insertions before weak syllables (32 out of 48 cases) involved monosyllabic function words (such as *dusty senseless drilling -> thus he sent his drill in).* Examination of the natural slips of the ear showed that a large number of the erroneous insertions of word boundaries before weak syllables in that corpus also involved monosyllabic function words (e.g., *descriptive -> the script of).* Word-boundary misplacements by human listeners therefore seem to reflect the prosodic probabilities of English remarkably accurately. The initial statement of the MSS, which referred only to lexical word

boundaries, may underestimate the degree to which the segmentation of continuous speech is driven by prosodic probability.

A complete implementation of the MSS would certainly have to take into account the distribution of strong and weak syllables with respect to different types of word boundaries. Cutler and Carter's tentative algorithm was deliberately oversimplified in an attempt to see how well the crudest implementation would perform. In that their algorithm distinguishes lexical from grammatical word hypotheses, it does in fact predict the predominance of grammatical words among weak-initial-syllable responses. However, its assumption that the longest word consistent with the input is accepted obviously has to be modified to take into account contextual acceptability. Cutler and Carter suggest several other ways in which their outline proposal can be substantially refined. Some further considerations involved in applying the MSS are discussed in the next and final section.

Conclusion: Applying a Metrical Segmentation Strategy

This chapter has argued that the absence of reliable word-boundary information in continuous speech can in part be overcome by exploiting the prosodic probabilities of the language. In English, where there is a strong likelihood that lexical words will begin with strong syllables, a strategy of assuming that a strong syllable is likely to be the onset of a new lexical word and that a weak syllable is not will successfully locate most lexical word boundaries. Evidence from human perceptual performance suggests that listeners do make use of such a segmentation strategy.

What exactly is a segmentation strategy? Let us first consider the term *segmentation.* It is important to be clear that this notion is logically distinct from a process *of classifying* the speech signal. A traditional preoccupation of Psycholinguistics has been the search for units of perception, that is, the postulated prelexical units of representation into which incoming speech signals are translated in order that lexical entries (presumably coded in terms of the same units) may be accessed. Among such postulated units are phonemes (Foss and Gernsbacher 1983) and syllables (Mehler 1981, Segui 1984). Clearly, the process of turning a continuous speech signal into a sequence of labeled discrete units involves dividing up the signal or segmenting it; that is, classification logically entails segmentation. But, as Norris and Cutler (1985) have argued in more detail, the reverse is not true. Simply making a division at a particular point in the signal does not necessarily imply that what is on either side of the division point is assigned a label, that is, classified.

Thus Cutler and Norris (1988) were able to point out that the MSS is compatible with a model of speech perception involving classification and also with a model involving no classification. They suggested, for instance, that in a model involving a phonemic level of representation, the occurrence in the input of one of a specified set of phonemes (the set of full vowels) could instigate a lexical-access attempt that starts either from that vowel or from its preceding syllabic onset. On the other hand, in a model involving no prelexical classification, a segmentation device could monitor the incoming signal for a high-energy quasi-steady-state portion of a specified minimum relative duration (full vowels are, after all, among the most readily identifiable portions of speech signals). Whenever this specification was met, the segmentation device could divide the speech at a point suitably prior to the onset of the steady state and again instigate a lexical-access attempt from that point, with the input to the lexicon being a relatively untransformed portion of the speech signal, of which only the onset need be defined.

Thus, although the metrical segmentation strategy is based on the distinction between strong and weak syllables, syllables per se are not part of its operation. It is really the strong and weak *vowels* that matter. On any implementation of the strategy, the occurrence of a full vowel must trigger segmentation. But segmentation probably does not then occur precisely at the vowel itself, if only because it is more efficient to locate the actual onset of the word. In principle, lexical access *can* be based on strong vowels; one could, for instance, imagine a lexicon in which *hat, bedazzle, straggler,* etc., were all in some sense stored together. But there is no doubt that accurate location of the word onset is more useful, and for the MSS this means locating the left boundary of the syllable in which the detected vowel occurs. The right boundary is quite unimportant, especially in an implementation of the MSS such as that proposed by Cutler and Carter (1987), in which the lookup process starts at each strong syllable and *continues,* if necessary, over subsequent weak syllables, returning in each case the longest candidate consistent with the input.

Location of a syllable's left boundary means correctly attaching the syllabic onset to the vowel. In English, onsets can be null, or they can contain up to three phonemes (e.g., *oak, soak, stoke,* and *stroke* are all English words). There is evidence that consonant cluster onsets in English are perceived as integral units (Cutler, Butterfield, and Williams 1987); this could facilitate the process of locating the left boundary of a syllable if a strong vowel is detected.

The importance in the MSS proposal of the vowel plus its preceding onset means that the proposal resembles other models in the literature that share this feature, for example, the notion of demisyllables as representational units (Fujimura and Lovins 1978) and the consonant-vowel units that figure in Dogil's (1986) "pivot parser." But it is nevertheless not a proposal about representational units, i.e., about classification. It is only a proposal about segmentation for lexical access, about locating those points in a continuous speech signal that are the most efficient points from which to initiate attempts at lexical access.

The second component in the proposal for a segmentation strategy is the notion of *strategy*. It is not intended that this should be considered as a conscious operation on the listener's part. The prelexical level is presumably not a processing level open to conscious inspection and control. Metrical segmentation is best thought of as the operation of an autonomous and automatic device, the purpose of which is to initiate lexical-access attempts with maximum efficiency, i.e., with as little waste as possible. Its operation should be guided by experience, probably by very early experience with one's native language.

Thus native speakers of different languages might use a number of different variants of the same basic type of segmenting device. The MSS is a specific proposal about how such a device operates for a free-stress language like English. But even in languages with other prosodic structures there might still be quite similar possibilities for segmentation routines. In a fixed-stress language like Polish, for instance, the relationship between stress placement and lexical-word boundaries might well be exploited by a segmentation device. Segmentation of nonstress languages like French does not have such an obvious prosodic basis, since in such languages there is no opposition between strong and weak syllables; all syllables are effectively equal in their contribution to linguistic rhythm. But much the same sort of device may still operate. For instance, with no prosodic basis for distinguishing likely word-initial syllables from likely noninitial syllables, a segmentation device of the general type embodied by the MSS might treat all syllables as equally likely to begin a word and simply segment speech signals at the onset of every syllable. There is evidence that simple syllable-based segmentation does indeed occur in the perception of French (Cutler, Mehler, Norris, and Segui 1986).

The metrical segmentation strategy may be only one of a number of operations that participate in the recognition of continuous speech. Its particular contribution is to increase the efficiency of the initial process of lexical access. Evidence from comparative implementations suggests that

its contribution to efficiency is high, even though the exigencies of experimental design mean that its operation in human recognition can best be appreciated from its occasional failures, such as finding *bone* in *trombone.*

References

Bond, Z. S., and Garnes, S. 1980. Misperception of fluent speech. In R. Cole (ed.), *Perception and Production of Fluent Speech.* Hillsdale, N.J.: Erlbaum.

Briscoe, E. J. 1989. Lexical access in connected speech recognition. *Proceedings of the Twenty-Seventh Congress, Association for Computational Linguistics,* Vancouver.

Browman, C. P. 1978. Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics* 42.

Browman, C. P. 1980. Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand.* New York: Academic Press.

Brown, G. D. A. 1984. A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavior Research Methods, Instrumentation, and Computers* 16:502-532.

Butterfield, S., and Cutler, A. 1988. Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. *Proceedings of SPEECH 88* (Seventh symposium of the Federation of Acoustic Societies of Europe), pp. 827-833. Edinburgh.

Celce-Murcia, M. 1980. On Meringer's corpus of "slips of the ear." In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand.* New York: Academic Press.

Coltheart, M. 1981. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology* 33A: 497-505.

Cutler, A., Butterfield, S., and Williams, J. N. 1987. The perceptual integrity of syllabic onsets. *Journal of Memory and Language* 26:406-418.

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

Cutler, A., Mehler, J., Norris, D., and Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25:385-400.

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

Dogil, G. 1986. Phonological pivot parsing. *Proceedings of COLING 86—Twelfth International Conference on Computational Linguistics,* Bonn.

Foss, D. J., and Gernsbacher, M. A. 1983. Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 22:609-632.

Fujimura, O., and Lovins, J. B. 1978. Syllables as concatenative phonetic units. In A. Bell and J. B. Hooper (eds.), *Syllables and Segments.* Amsterdam: North-Holland.

Games, S., and Bond, Z. S. 1975. Slips of the ear: Errors in perception of casual speech. *Proceedings of the Eleventh Regional Meeting, Chicago Linguistic Society,* pp. 214-225.

Games, S., and Bond, Z. S. 1980. A slip of the ear? A snip of the ear? A slip of the year? In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand.* New York: Academic Press.

Harrington, J., and Johnstone, A. 1987. The effects of word boundary ambiguity in continuous speech recognition. *Proceedings of the Eleventh International Congress of Phonetic Sciences,* vol. 3, pp. 89-92. Tallinn, Estonia.

Kucera, H., and Francis, W. N. 1967. *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Mehler, J. 1981. The role of syllables in speech processing. *Philosophical Transactions of the Royal Society* B295:333-352.

Norris, D., and Cutler, A. 1985. Juncture detection. *Linguistics* 23:689-705.

Segui, J. 1984. The syllable: A basic perceptual unit in speech processing. In H. Bouma and D. G. Bouwhuis (eds.), *Attention and Performance,* vol. 10. Hillsdale, N.J.: Erlbaum.

Svartvik, J., and Quirk, R. 1980. *A Corpus of English Conversation.* Lund: Gleerup.

Wilson, M. D. 1988. MRC Psycholinguistic Database: Machine-usable dictionary, version 2.0. *Behavior Research Methods, Instrumentation, and Computers,* 20:6-10.