JOURNAL OF MEMORY AND LANGUAGE 31, 218–236 (1992)

# Rhythmic Cues to Speech Segmentation: Evidence from Juncture Misperception

ANNE CUTLER AND SALLY BUTTERFIELD

*MRC Applied Psychology Unit, Cambridge, United Kingdom*

Segmentation of continuous speech into its component words is a nontrivial task for listeners. Previous work has suggested that listeners develop heuristic segmentation procedures based on experience with the structure of their language; for English, the heuristic is that strong syllables (containing full vowels) are most likely to be the initial syllables of lexical words, whereas weak syllables (containing central, or reduced, vowels) are nonword-initial, or, if word-initial, are grammatical words. This hypothesis is here tested against natural and laboratory-induced missegmentations of continuous speech. Precisely the expected pattern is found: listeners erroneously insert boundaries before strong syllables but delete them before weak syllables; boundaries inserted before strong syllables produce lexical words, while boundaries inserted before weak syllables produce grammatical words. 

## INTRODUCTION

BE ALERT! YOUR COUNTRY NEEDS LERTS!

is an old joke, but it works. It turns up on bumper stickers, lapel badges, lavatory walls, even keyrings.[1] The following joke, however, does not work:

BE A PAL! APAL FOLKS HAVE MORE FUN!

Why does the first joke work while the second one does not? That is, why is splitting the adjective *alert* into an article plus a pseudo-noun funny, while joining the article and noun *a pal* into a pseudo-adjective is not?

We suggest that the answer has nothing to do with nouns versus adjectives, or with splitting versus joining words. It has to do with the expectations which speakers of English have about where word boundaries ought to occur in English utterances.

Finding where new words begin in continuous speech is a problem, since word boundaries are rarely reliably marked. Cutler and Norris (1988) proposed that speakers of English use the rhythmic patterns of utterances to guide hypotheses about where new words begin. In English, which is a stress language, speech rhythm has a characteristic pattern which is expressed in the opposition of strong versus weak syllables. Strong syllables bear primary or secondary stress and contain full vowels, whereas weak syllables are unstressed and contain short, central vowels such as schwa. (Although there are levels of stress

This research was supported by a grant from the Alvey Directorate, UK, to Cambridge University, the Medical Research Council and STC Technology Ltd. The study of time-compressed speech referred to in the discussion is supported by a grant from the Human Frontier Science Program. We thank Bill Barry and Mary R. Smith for assistance, Ian Nimmo-Smith for statistical advice, and Dennis Norris for valuable discussions. Further thanks go to all who sent examples of unpublished slips of the ear, in particular, Zinny Bond and Cathe Browman. For computing the lexical statistics from *Longmans Dictionary* we thank James McQueen, and he acknowledges the Longman Group, Inc. for allowing access to the machine-readable version of the dictionary. Preliminary reports of parts of this research were presented to the "Seventh Symposium of the Federation of Acoustic Societies of Europe," Edinburgh, August 1988, to the "24th International Congress of Psychology," Sydney, September 1988, and to the "116th Meeting of the Acoustical Society of America," Honolulu, November 1988. Correspondence and reprint requests should be addressed to the first author at MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge CB2 2EF, UK.

[1] The first author thanks the second author for the gift of the keyring.



0749-596X/92 $3.00

within strong syllables, the only difference which matters for metrical rhythm is the binary opposition of strong versus weak.) Cutler and Norris' proposal accounts for the results of an experiment in which they found that listeners were slower to detect the embedded real word in *mintayf* (in which the second vowel is [e], i.e., strong) than in *mintef* (in which the second vowel is schwa, i.e., weak). That is, these two bisyllables differ in their metrical structure: one has two strong syllables, the other a strong and a weak syllable. Cutler and Norris suggested that listeners treat strong syllables as likely to be the initial syllables of new (lexical) words. In effect, listeners would employ a strategy of segmenting speech signals at the onset of each strong syllable. In the experiment, therefore, *mint* would be relatively difficult to detect in *mintayf* because listeners were segmenting *mintayf* prior to the second syllable, so that detection of *mint* in this case required combining speech material from parts of the signal which had been separated from one another by segmentation. No such difficulty would arise for the detection of *mint* in *mintef*, since the weak second syllable would not be segmented from the preceding material.

The statistics of the English vocabulary show that assuming strong syllables to be word-initial will be a pretty good bet. Cutler and Carter (1987) found that in a computer-readable English dictionary containing over 33,000 entries about 12% of the words were monosyllables (such as *camp* or *lodge*), just over 50% were polysyllables with primary stress on the first syllable (such as *camphor* or *cycle*), a further 11% were polysyllables with secondary stress on the first syllable (such as *campaign* or *psychological*), while the remaining 27% were polysyllables with weak initial syllables (in which the vowel in the first syllable is usually schwa, as in *camellia*, but may also be a reduced form of another vowel, as in *illogical*). All of the first three categories have strong initial syllables, and these categories together account for 73% of the words in the list.

Moreover, frequency of occurrence statistics show that monosyllables occur on average far more frequently than any type of polysyllable; and within the set of polysyllables, words with strong initial syllables occur more frequently than words with weak initial syllables. Although there are more than seven times as many polysyllables in the English language as there are monosyllables, average speech contexts are likely to contain almost as many monosyllables as polysyllables (among the lexical, or content, words; grammatical, or function, words are nearly all monosyllabic).

Cutler and Carter examined a natural speech sample consisting of approximately 190,000 words of spontaneous British English conversation. Almost 60% of the lexical words in this corpus were monosyllables. 28% were polysyllables with initial primary stress, and a further 3% were polysyllables with initial secondary stress. Most noticeably, perhaps (especially when one considers that a relatively high proportion of the speech in this corpus came from conversation among academics!), less than 10% of the lexical words were polysyllables with weak initial syllables. In other words, the three categories with strong initial syllables accounted, together, for over 90% of the lexical word tokens.

However, this lexical word count disguises one important fact: the majority of words in the corpus were, in fact, grammatical words. But because hardly any grammatical words had more than one syllable, the lexical word total nevertheless accounts for 51% 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. In that case, 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 syl-

lable in spontaneous English conversation would seem to have about a three to one chance of finding that strong syllable to be the onset of a new lexical word. A weak syllable, on the other hand, would be most likely to be a grammatical word. It would appear, therefore, that English speech indeed provides a good basis for the implementation of a segmentation strategy which assumes strong syllables to be the onsets of lexical words.

Puns such as "Be a lert," therefore, work because they conform to the natural strategy for segmenting continuous English speech: insert a boundary before a strong syllable and assume what follows is a lexical word. A word beginning with a weak syllable—a "minority" sequence—is treated as if it were an initial strong syllable preceded by a grammatical word. Because this sequence seems the more natural, the pun is easily apprehended.

Analogously, "Be apal" does not work because it attempts the opposite; it treats a natural sequence of a grammatical word plus strong initial syllable as if it were a single word beginning with a weak syllable. This requires deletion of a boundary before a strong syllable, exactly the reverse of the natural strategy.

We believe that the structure of natural strategies for segmenting continuous speech in large part accounts for the patterns of acceptability in puns involving word boundary shifts and in many similar word usage phenomena. It is, of course, difficult to substantiate this claim with systematic data; for example, definitive collections of puns do not exist. However, it is noteworthy that in four books of collected graffiti (Rees, 1979, 1980, 1981, 1982) we found numerous examples in support of our hypothesis—involving a word boundary being inserted before a strong syllable ("transcend dental medication") or deleted before a weak syllable ("Laura Norda")—but no unambiguous counterexamples.

Folk etymologies offer another source of such evidence; a dictionary of British pub names (Dunkling & Wright, 1987) lists numerous pub name etymologies (most of them, in Dunkling and Wright's opinion, completely spurious). Thus "Goat and Compasses" is held to derive from "God encompass us," and "Barley Mow" from "Bel Amour." Each of these postulates both insertion of a boundary before a strong syllable and deletion of a boundary before a weak syllable. In other examples insertion of a word boundary before a *weak* syllable is proposed (e.g., derivation of "Cat in the Well" from "St. Catherine's Well"), but every such example involves interpreting the weak syllables as function words; no example postulates deletion of a boundary before a strong syllable.

We further observe that the same pattern occurs in children's jokes; puns which insert a word boundary before a strong syllable, or delete a word boundary before a weak syllable, are common:

Q: When is a door not a door? A: When it's ajar.

Q: Why is a ship called "she"? A: Because it's often abroad.

Q: Why won't you starve in the desert?
A: Because of the sand which is there.

Again, in published collections of such jokes (e.g., Ahlberg & Ahlberg, 1982) we found that many more of them produced more natural patterns from less natural patterns than vice versa; it is impossible to provide systematic figures but—as with puns—readers can of course check our impressions against their own memories.

Finally, we note that jokes about misperceptions conform to the same generalization—in a well-worn British example, "Send reinforcements, we're going to advance" is heard over a field telephone as "Send three-and-fourpence, we're going to a dance." A similar American example reports schoolchildren patriotically reciting "I led the pigeons to the flag." In each of these a boundary is inserted before every strong syllable which in the original utterance was not word-initial. Of course there are also counterexamples ("Shirley, good

Mrs. Murphy," from the 23rd psalm, involves insertion of a boundary before a weak syllable, for instance); but our impression is that again the majority of cases involve boundary insertion before strong syllables and deletion before weak syllables.

Of course informal evidence of this kind, impressive though it may be in aggregate, does not offer a route by which our hypothesis can be put to an explicit test. Misperceptions, however, as exemplified in the jokes above, do admit of more systematic investigation. Any misperception of an utterance more than one word in length offers the opportunity for misperception of where in the utterance word boundaries, or junctures, occur. In the present study we examine juncture errors in misperceptions of continuous speech.

We can make two general classes of predictions, which arise directly from the hypothesis about lexical segmentation proposed by Cutler and Carter (1987) and Cutler and Norris (1988). Cutler and Carter cast the rhythmic segmentation hypothesis in the form of an algorithm, the principal steps of which are:

1.1. The main lexicon contains only lexical words; grammatical words constitute a separate list.

1.2. An initial segmentation process scans the input and places markers at the onset of each strong syllable.

1.3.1. If the initial string of the current input is not preceded by a marker, it is submitted to the grammatical list; if it is preceded by a marker, it is submitted to the main lexicon.

1.3.2. The lookup process in both the main lexicon and the grammatical word list returns the longest candidate consistent with the input, *except that* the occurrence of a marker indicating the beginning of a strong syllable will terminate the current lookup process and initiate a new lookup process in the main lexicon.

The first type of prediction concerns what kind of juncture errors occur before what kind of syllable. There are only two possible types of juncture errors: insertion of a boundary, where there is none in the input, and deletion of a boundary that is in the input. Likewise, there are two kinds of syllables from the point of view of metrical rhythm: strong and weak. If human listeners are indeed undertaking a first-pass segmentation of speech signals along the lines proposed in 1.1–1.3.2, then clearly they will be more likely to make some kinds of juncture misperceptions than others. Specifically, the obligatory initiation of a new lookup process occasioned by every strong syllable (1.3.2) will tend to induce errors in which strong syllables are erroneously taken to be word-initial; likewise, the obligatory attachment of weak syllables to preceding syllables wherever possible (again (1.3.2)) will tend to induce errors in which boundaries preceding weak syllables are overlooked. In brief, therefore, errors involving insertion of a word boundary before a strong syllable or deletion of a boundary before a weak syllable should prove to be relatively common, whereas errors involving insertion of a boundary before a weak syllable or deletion of a boundary before a strong syllable should be relatively rare.

The second type of prediction arises from the word class correlates of the strong/weak distinction, as expressed in the algorithm in (1.3.1). All strong syllables initiate lookup processes in the main lexicon, whereas when a weak syllable initiates a lookup process, it is in the grammatical list. This should lead to strong syllables being interpreted as new lexical words, while weak syllables are interpreted as grammatical words, which in turn leads to a prediction which specifically concerns boundary insertions: when boundaries are inserted prior to strong syllables the word following the boundary should be taken to be a lexical word, whereas when boundaries are inserted prior to weak syllables the word following the boundary should be taken to be a grammatical word.

The present study tests each of these pre-

dictions against evidence from the misperception of continuous speech. We use two sources of evidence: spontaneous misperceptions and laboratory-induced misperceptions.

## 1. Spontaneous Misperception

### *Procedure*

The psycholinguistic literature contains a number of studies of spontaneous misperceptions, or "slips of the ear" (e.g., Bond, 1973; Bond & Garnes, 1980; Browman, 1978, 1980; Celce-Murcia, 1980; Garnes, 1977; Garnes & Bond, 1975, 1980). Many of these include a large number of examples of errors. We examined all the published error examples we could find, plus all the slips of the ear included in a speech error collection assembled over several years by the first author. Finally, we asked other researchers in the field to send us slips of the ear; in response to this request, two leading researchers sent us large samples of unpublished slips.

Bond and Garnes (1980) report that misperceptions of juncture are relatively common and accounted for about 18% of their corpus of spontaneous slips of the ear. Among the slips that we analysed, we found in all 246 which involved misplacement of a word boundary across at least one syllabic nucleus. (Errors in which a boundary was misplaced across only one or two consonants—such as "up with Anne" → "up a fan"—were excluded, because they are irrelevant to the hypothesis about rhythmic structure.) Some slips in fact involved more than one misplaced boundary (such as "for an occasion" → "fornication," in which boundaries before two weak syllables have been deleted); the 246 misperceived utterances contained a total of 310 juncture misplacements.

Some example errors are shown in Table 1. We found that in this set of naturally occurring errors all possible types of word boundary misplacement appeared: insertions of a word boundary before a strong syllable (e.g., "analogy" → "and allergy"); insertions of a boundary before a weak syllable (e.g., "effective" → "effect of"); deletions of a boundary before a strong syllable (e.g., "is he really" → "Israeli"); deletions of a boundary before a weak syllable (e.g., "my gorge is" → "my gorgeous").

TABLE 1
Examples of Spontaneous Slips of the Ear

| Input | | Error |
|---|---|---|
| She'll officially | → | Sheila Fishley |
| She's a must to avoid | → | She's a muscular boy |
| How big is it? | → | How bigoted? |
| By loose analogy | → | By Luce and Allergy |
| The parade was illegal | → | The parade was an eagle |
| Into opposing camps | → | Into a posing camp |
| My gorge is still rising | → | My gorgeous . . . . |
| I'm not sure about this yet but | → | I'm not sure about this shepherd |
| Is he really? | → | Israeli? |
| I can't fit any more on | → | I can't fit any, moron |
| In closing | → | Enclosing |
| The effective firing rate | → | The effect of . . . . |

The rhythmic segmentation hypothesis predicts first that insertion errors will occur more often before strong syllables than before weak, while deletion errors will occur more often before weak syllables than before strong, and second that insertions before strong syllables will tend to postulate lexical words while insertions before weak syllables will tend to postulate grammatical words. Of course the context in which an individual utterance occurs will to some extent constrain the range of possible misperceptions. But note that Cutler and Carter's (1987) corpus analysis suggests that both types of prediction are counterintuitive. Cutler and Carter estimated on the basis of their analysis that among nonword-initial syllables, weak syllables on average outnumber strong syllables by more than three

to one. This makes the opportunity for erroneous word boundary insertions much greater before weak syllables than before strong. Likewise, Cutler and Carter found that grammatical words outnumbered lexical words in this corpus (the ratio was 59:41); this would suggest that, *ceteris paribus*, erroneous word boundary insertions ought to produce three grammatical words to every two lexical words.

### *Results and Discussion*

Table 2 shows the distribution of the 310 boundary misplacements across the four possible categories of insertions versus deletions before strong versus weak syllables. It can be seen that, as predicted, erroneous boundary insertions occur more often before strong than before weak syllables, while erroneous boundary deletions occur more often before weak than before strong syllables. The interaction is highly significant (with correction for continuity, $\chi^2$ [1] = 22.48, $p < .001$). Binomial tests on boundary insertions versus deletions show that each difference is separately significant: $z = 3.79, p < .001$ for insertions, $z = 2.87, p < .005$ for deletions.

Table 3 shows the distribution of word types following erroneously inserted boundaries. As predicted, when boundaries are inserted before strong syllables, the strong syllable is nearly always taken to be the beginning of a lexical word; but when boundaries are inserted before weak syllables, the weak syllable is more often interpreted as a grammatical, or function, word. Again the difference is significant (with correction for continuity, $\chi^2$ [1] = 52.13, $p <$ .001), and the word class difference is separately significant for insertions before strong versus weak syllables: $z = 8.33, p <$ .001 for strong syllables, $z = 1.79, p < .04$ for weak syllables.

TABLE 2
WORD BOUNDARY INSERTIONS AND DELETIONS BEFORE STRONG VERSUS WEAK SYLLABLES IN SPONTANEOUS SLIPS OF THE EAR

| | Before strong | Before weak |
|---|---|---|
| Boundary insertions | 90 | 45 |
| Boundary deletions | 68 | 107 |

TABLE 3
OCCURRENCE OF LEXICAL VERSUS GRAMMATICAL WORDS FOLLOWING INSERTED WORD BOUNDARIES BEFORE STRONG VERSUS WEAK SYLLABLES IN SPONTANEOUS SLIPS OF THE EAR

| | Before strong | Before weak |
|---|---|---|
| Lexical | 85 | 16 |
| Grammatical | 5 | 29 |

Thus both types of prediction from the rhythmic segmentation hypothesis are supported by the data from spontaneous slips of the ear. Word boundaries tend to be inserted more often before strong syllables than before weak, but deleted more often before weak syllables than before strong; boundaries inserted before strong syllables produce lexical words, while boundaries inserted before weak syllables produce grammatical words.

As we pointed out above, both these findings are counterintuitive given the relative proportions of strong and weak syllables indicated by Cutler and Carter's corpus analysis. Moreover, note that just over half of all errors occurred before strong syllables, although Cutler and Carter estimated that only 39% of all syllables in typical English speech are strong. This again is consistent with the hypothesis that speech segmentation is primarily driven by hypotheses about strong syllables, with the interpretation of weak syllables being to a certain extent subordinate (cf. the even more radical proposals to this effect made by Grosjean & Gee, 1987).

It might be argued that two errors in the same utterance are not independent of one another—for instance, deletion of one word boundary may require insertion of a boundary elsewhere. We therefore examined only the *first* error in each utterance (although it

is, of course, not necessarily the case that the first error in the utterance as reported is actually the first error made by the listener, since earlier word assignments may be revised as a consequence of later ones). However, this analysis produced exactly the same pattern, with insertions more common before strong syllables than before weak ($z = 3.22$, $p < .001$), and deletions more common before weak syllables than before strong ($z = 2.16$, $p < .02$).

The predictions of the rhythmic segmentation hypothesis are, therefore, strongly supported by the pattern of boundary misperceptions in slips of the ear. But might the same pattern be predicted by an alternative hypothesis? Consider the possibility that slips of the ear tend to result from application of the listener's inferential abilities to imperfectly interpretable input and that the pattern we found in the data simply falls out of the statistical properties of the vocabulary which the listener accesses in reconstructing the input. We know that the vocabulary contains more words beginning with strong syllables than with weak; the efficiency of the rhythmic segmentation hypothesis is built upon that very fact. Might it be the case that listener misperceptions involve no segmentation process at all, but merely misselection from a heavily skewed vocabulary?

Two versions of this alternative hypothesis seem possible. First one might claim that when listeners find an utterance for some reason difficult to perceive, they attempt to construct a *plausible* hypothesis as to what it was. This suggestion seems to us reasonable and consistent with self-reports from listeners (especially from listeners with hearing loss). Another version of the hypothesis, however, might be that listeners who cannot interpret all or part of an input choose candidate words *randomly* from the lexicon.

Each version of the alternative hypothesis translates on the face of it into a prediction about frequency effects. If plausibility drives listener hypotheses, then there should be an overall tendency for errors to contain words of higher frequency than the input contained. On the other hand, if listener reconstructions represent random choice in the lexicon, the fact that the great majority of words have very low frequencies should produce the opposite result: words in errors, being randomly selected, should tend to be of lower frequency than words in the input.

In practice, however, we will not expect a negative frequency effect because it has long been known that conditions of difficulty for listeners tend to produce high-frequency responses. Studies of the perception of speech in noise show very robust frequency effects (Broadbent, 1967; Howes, 1957; Savin, 1963), which seem most consistent with an explanation in terms of criterion bias, i.e., a readiness to accept a high-frequency word on the basis of scantier acoustic evidence than would be required for a low-frequency response (Broadbent, 1967; Luce, 1986a).

To test for the presence of frequency effects in the present corpus, we analysed the frequency of words in errors versus input utterances. This analysis presented several problems. First, there was the problem presented by grammatical words. These have such a high frequency of occurrence that any error which includes a grammatical word not present in the input will necessarily have a higher overall frequency than the input, while any error which omits a grammatical word present in the input will necessarily have a lower overall frequency than the input.[2] It seems reasonable to suppose, however, that if frequency effects are operative they will be in evidence in the lexical words; we therefore chose to avoid this problem by analyzing lexical words only.

Second, many of the slips of the ear involved proper names, either in the input or

[2] In fact errors containing a lower number of function words than were in the actual utterance outnumbered errors containing a higher number of function words.

in the error. The frequency of these is impossible to assess, since details of the listeners and the speech context are unavailable. Omitting all errors involving proper names (as well as errors in which only grammatical words were involved) reduced the total of 246 utterance pairs to 165. Table 4 shows the number of errors of each of the four types in which the lexical words were of higher versus lower frequency than the lexical words in the input.

It can be seen that there is a strong tendency for boundary insertions to result in words of higher frequency, while boundary deletions result in words of lower frequency. This is as expected: short words tend to be of higher frequency than long words, and boundary insertions tend to result in errors containing more, shorter words than the input contained, while boundary deletions tend to result in errors with fewer, longer words than the input. However, there is no overall difference in frequency between error and input (in 81 pairs the error words are higher in frequency than the input words, in 84 pairs they are lower; $z = 0.16$), and there is no significant difference in the frequency pattern for errors predicted by the rhythmic segmentation hypothesis versus errors not predicted (with correction for continuity, $\chi^2$ [1] = 0.83).

We conclude, therefore, that the rhythmic effects which we observe in the distribution of juncture misperceptions do not simply fall out of the statistical distribution of the English vocabulary given either selection of plausible words or selection of random words. Of course, they do represent the statistical distribution of English in the sense that rhythmic segmentation works well for listeners precisely because it accurately reflects the probable structure of spoken input. But the pattern we find in slips of the ear strongly suggests that it is rhythmic structure which guides the hypothesis which listeners form when an input is difficult to interpret: strong syllables are hypothesised to be initial syllables of lexical words, while weak syllables are hypothesised to be non-initial syllables, or grammatical words.

TABLE 4
COMPARATIVE FREQUENCY OF LEXICAL WORDS IN INPUT VERSUS ERROR IN SPONTANEOUS SLIPS OF THE EAR, SEPARATELY FOR WORD BOUNDARY INSERTIONS AND DELETIONS BEFORE STRONG VERSUS WEAK SYLLABLES

| | Error higher in frequency than input | Error lower in frequency than input |
|---|---|---|
| Boundary insertions before strong syllables | 33 | 12 |
| Boundary insertions before weak syllables | 23 | 6 |
| Boundary deletions before strong syllables | 7 | 32 |
| Boundary deletions before weak syllables | 18 | 34 |

Thus the evidence from natural slips of the ear solidly supports the predictions of the rhythmic segmentation hypothesis. In the second part of this study we attempted to induce misperceptions in the laboratory. The laboratory study provides a control for the natural study, in that in the laboratory we can precisely constrain the characteristics of the input. The characteristics of the input in the natural corpus (very little of which was actually collected by us) cannot be fully ascertained. Although we can estimate the likely rhythmic structure of a spontaneous speech sample, based on Cutler and Carter's corpus analysis, we cannot be sure of the accuracy of this estimate. Moreover, we know very little of the semantic and pragmatic constraints which may have affected natural errors. These problems can be overcome by eliciting in the laboratory juncture misperceptions of the kind listeners so often make spontaneously.

## 2. LABORATORY-INDUCED MISPERCEPTION: FAINT SPEECH

Misperceptions can be induced in the laboratory by presenting listeners with speech which for any reason is difficult for them to hear; filtering and noise-masking are fre-

quently-used methods of making speech perception hard. In a previous study (Smith, Cutler, Butterfield, & Nimmo-Smith, 1989) we established that speech rhythm is highly resistant to noise-masking. However, both noise-masking and filtering interact with the spectral characteristics of the speech signal (Miller & Nicely, 1955): any given noise signal, or any particular filter, will affect some speech sounds more than others, and concurrent masking can result in a percept which is a spurious combination of characteristics of the masker and the speech (Gordon-Salant & Wightman, 1983). To avoid this confounding factor we chose to make our speech signals very hard for listeners to hear by presenting them very faintly, just at the level at which listeners could hear about 50% of presented input. Although differing intrinsic intensity of speech sounds will inevitably mean that some sounds become actually fainter than others, listener familiarity with characteristic intrinsic intensities should imply that the *subjective* reduction in intensity is equivalent for all sounds; most importantly, however, there will be no interference from concurrent sound. Since listeners differ in auditory sensitivity, choice of this method required us to pretest subjects individually to establish speech reception thresholds.

## *Method*

### *Subjects*

Eighteen experienced members of the Applied Psychology Unit subject panel took part, for payment, in the experiment. All were under 55 years of age, and none reported problems with their hearing.

### *Materials and Procedure*

*a. Pre-test.* Subjects were tested individually. For each subject, a pretest was conducted to estimate speech reception threshold. In this pretest, subjects were presented with speech material of the type used in speech reception threshold tests by audiologists. A short passage and a list of spondees (i.e., words with two strong syllables, such as "toothbrush," "doormat," "workshop") were recorded by a phonetically trained male speaker of Southern British English. The passage was fairly complex text containing statistical information. The list of 36 spondees was taken from the CID W-1 and W-2 Word Lists (Benson, Davis, Harrison, Hirsch, Reynolds, & Silverman, 1951). One obvious item of American vocabulary was replaced ("sidewalk" was replaced with "homework").

These recorded materials were played over Sennheiser HD 420 SL headphones from a Revox B77 tape recorder connected to a step-attenuator. The passage was played first, with the attenuator set at 20 dB and the volume on the tape recorder adjusted to produce clearly audible speech at the headphones. The subjects were instructed to adjust the volume knob on the tape recorder to the lowest level at which they could follow the speech. At the end of the passage they were asked a few general questions to confirm that they had been able to follow the speech at the level they had chosen.

The subjects were then familiarized with the set of spondees; they read the list through and then listened to the same items, in alphabetical order, at a level 15 dB above the threshold they had previously established. Subjects repeated each word as they heard it. They were then presented with a randomised list of the same items, with the first item at least 5 dB above the level previously established for the read passage. The attenuation was increased (i.e., the volume reduced) by 3-dB steps for each three items until three words were not repeated or repeated incorrectly. Then the attenuation was decreased by 1-dB steps until an item was repeated correctly. If the subject repeated 50% of the following items correctly, this level was taken as the estimated speech reception threshold; if not, then the threshold seeking phase was continued until the end of the list.

*b. Experiment*

Forty-eight unpredictable sequences of six syllables ("soon police were waiting"; "conduct ascents uphill") were constructed. Each sequence had an alternating stress rhythm of strong (S) and weak (W) syllables. In half the cases the rhythm was SWSWSW ("soon police were waiting"); in the other half it was WSWSWS ("conduct ascents uphill"). These manipulations resulted, obviously, in exactly equal numbers of strong and weak syllables in the sequences as a whole as well as in each syllable position. Note that each of the two chosen rhythmic structures will allow very many different possible divisions into words, and each is a very common pattern in English; thus rhythmic patterns alone could not afford our subjects much information about what words had occurred.

Two further factors were varied systematically in the materials. One was where word boundaries occurred with respect to the rhythm. One-third of the sequences had only weak word-initial syllables ("conduct ascents uphill"; "sons expect enlistment"—note that although in the latter example the very first syllable is strong, the first syllable is to a certain extent irrelevant, since subjects have no choice about whether or not it is word-initial). A further one-third had only strong word-initial syllables ("dusty senseless drilling"; "an eager rooster played"); and the remaining third had a mixture of strong and weak word-initial syllables ("soon police were waiting"; "achieve her ways instead"). Roughly equal numbers of strong and weak syllables were word-initial versus non-word-initial.

The remaining factor was the nature of the vowel in the strong syllables. These were chosen from a set of three phonetically short vowels (/ɛ/, /ɪ/, /ʌ/) and a set of three phonetically long vowels (/e/, /i/, /u/). One quarter of the utterances contained all long vowels in the strong syllables ("soon police were waiting"); one quarter contained all short vowels ("conduct ascents uphill"); and the remaining half contained a mixture of long and short vowels ("achieve her ways instead"). The weak vowels were mostly schwa. (The vowel length factor was included as part of another study for which these materials were used: Smith, Cutler, Butterfield, & Nimmo-Smith, 1989.) The 48 sequences are listed in the Appendix.

The sequences were recorded (by the same speaker who recorded the materials for the pretest) such that the peak level of the strong syllables was at approximately −12 dB on the VU meter of the tape recorder. Each sentence was repeated; the speaker's voice gave prior to each trial the number (from 1 to 48) of the trial, and, prior to each repetition, the word "again." Both the number and the word "again" were recorded several dB above the level of the experimental item.

For subjects 1–4, the experimental materials were presented with a further attentuation of 2 dB beyond the level of the estimated speech reception threshold; however, this procedure yielded rather few complete responses. Accordingly, for subjects 5–18, the attenuation was left at the level of the estimated speech reception threshold for each subject. The subjects were told that they would be listening to "speech that is difficult to hear clearly." Their task was to write down what they thought was said. They were asked to insert a dash if they were sure a syllable had been spoken but they could not report any of it; this enabled us to analyze all responses on which the subjects had correctly determined the number of syllables spoken.

The predictions from the rhythmic segmentation hypothesis are the same as they were for the natural slips of the ear: boundary insertion errors should be more common before strong than before weak syllables, while boundary deletion errors should be more common before weak than before strong syllables; insertion errors before strong syllables should produce lexical

words, while insertion errors before weak syllables should produce grammatical words. The laboratory-induced corpus differs from the natural corpus in that we have complete knowledge of the rhythmic and contextual characteristics of the input.

### *Results and Discussion*

Of the 864 responses (18 listeners × 48 input sequences), some were entirely correct and some were entirely missing (i.e., listeners produced no response). Many other responses consisted of a few syllables only. Although it was usually obvious which syllables in the input were being interpreted (for instance, "super" as a response to "soon police were waiting" is presumably based on the first two syllables), we decided to omit such cases from the analysis. We confined the analysis to responses in which both the number of syllables (six) and the rhythmic pattern of the input were correctly preserved; 369 of the responses satisfied these criteria, and 168 of these contained boundary misplacements. Some responses contained more than one boundary misplacement. The total number of boundary errors was 264; 55 of the analysed responses contained one or more dashes; there were 52 dashes replacing strong syllables and 53 replacing weak syllables. (The symmetry of these numbers arises from the fact that most such responses contained two dashes, e.g., "The deaths are just––" as a response to "Cadets are just unfit." This response was classified as a single boundary misplacement, at the second syllable.)

Table 5 gives examples of the type of complete responses produced to the faint speech. Examples of all four types of boundary error were produced. Thus reporting "conduct ascents uphill" as "the doctor sends her bill" involves inserting boundaries before the strong syllables and deleting boundaries before the weak syllables, while reporting "an eager rooster played" as "a new resolve again" involves inserting boundaries before the weak syllables and deleting boundaries before the strong syllables. However, as Table 6 shows, the types of error were again unevenly distributed. Word boundary insertions were more common before strong than before weak syllables, whereas word boundary deletions were more common before weak than before strong syllables.

TABLE 5
EXAMPLES OF MISPERCEPTIONS IN FAINTLY HEARD SPEECH

| Input | | Error |
|---|---|---|
| Conduct ascents uphill | → | The doctor sends her bill |
| Soon police were waiting | → | Soothe the least where waiting |
| Music's even paces | → | Music seen in phases |
| Sons expect enlistment | → | Some expect a blizzard |
| An eager rooster played | → | A new resolve again |
| Achieve her ways instead | → | A cheaper way to stay |
| Angels pinned beneath it | → | Angels pin their needles |
| Dusty senseless drilling | → | Thus he sent his drill in |
| Conduct ascents uphill | → | A duck descends some pill |
| Soon police were waiting | → | Soon to be awakened |
| Music's even paces | → | Music her replaces |
| Sons expect enlistment | → | Suns expectant listen |

We subjected the laboratory-produced data to the same analyses which we had conducted on the natural misperceptions. First we analysed the relative frequency of the different error types. In this case, however, we compared the observed frequen-

TABLE 6
WORD BOUNDARY INSERTIONS AND DELETIONS BEFORE STRONG VERSUS WEAK SYLLABLES IN FAINTLY HEARD SPEECH

| | Before strong | Before weak |
|---|---|---|
| Boundary insertions | 146 | 49 |
| Boundary deletions | 17 | 52 |

cies with expected frequencies which we generated on the basis of the actual properties of the input; we computed the actual frequency of strong versus weak initial versus noninitial syllables in the input, and prorated this across the 1845 syllables available for analysis (369 responses times five syllables—i.e., excluding the first syllable in each input sequence). Once again, the predicted interaction is significant (with correction for continuity, $\chi^2$ [1] = 59.13, $p < .001$), and the syllable strength difference is separately significant for insertions versus deletions: $z = 6.87$, $p < .001$ for insertions; $z = 4.09$, $p < .001$ for deletions.

Analysis of only the first error in each response once again produced the same pattern of effects: insertions were more common before strong syllables ($z = 5.18$, $p < .001$), but deletions were more common before weak syllables ($z = 4.0$, $p < .001$).

There were no significant differences in the frequency of errors as a function of whether the input sequence had the SWSWSW rhythmic pattern (86 responses with boundary misplacement) or the WSWSWS pattern (82 responses), nor were there significant effects of vowel length.

As with the natural slips, there are word class differences in the words which are postulated following erroneously inserted boundaries in subjects' responses. In this analysis it was necessary to exclude those responses classified as insertions where the postboundary word was represented by a dash (e.g., "music sees—faces" as a response to "music's even paces"); there were also some nonword responses. But as Table 7 shows, when the boundary precedes a strong syllable, the following word in the response is more often a lexical word, whereas when the boundary precedes a weak syllable, the following word is more often a grammatical word (with correction for continuity, $\chi^2$ [1] = 58.11, $p < .001$; $z = 7.64$, $p < .001$ for insertions before strong syllables; $z = 3.75$, $p < .001$ for insertions before weak syllables).

An analysis of lexical word frequencies was carried out in the same way as for the natural misperceptions, excluding proper names and function words as before.[3] There was again no significant tendency for errors to differ in word frequency from inputs, as Table 8 shows ($z = 0.08$); also, there was again no difference in frequency effects for the error types predicted versus not predicted by the rhythmic segmentation hypothesis (with correction for continuity, $\chi^2$ [1] = 0.84).[4]

Thus the analyses of the laboratory-induced misperceptions show just the same pattern as we found with natural slips of the ear. However, with this laboratory corpus we can carry out further analyses which we could not apply to the natural slips. First, because we have repeated measures on

[3] In this case responses containing a higher number of function words than were in the stimulus outnumbered responses containing a lower number of function words.

[4] Some evidence of a tendency towards higher frequency words in the responses than in the stimulus materials appeared when we analyzed responses which had the correct number of syllables and rhythm, and contained errors, but involved no boundary misperceptions (e.g., "a better budget ship" as response to "a better budget shift"). In 48 such lexical substitutions, 29 were words higher in frequency than the input, while 17 were lower frequency words (and there were two ties). This difference does not quite reach significance ($z = 1.62$, $p < .055$), but the ratio of 1.71:1 higher to lower is noticeably larger than the ratio of 1.06:1 obtained in the boundary misplacement errors consistent with the rhythmic segmentation hypothesis. It might be argued, therefore, that there *is* a tendency for frequency effects to operate under conditions of perceptual uncertainty. Missegmentations, however, do not *result from* the frequency bias—by contrast, the effect of the prior operation of rhythmically based segmentation is to constrain the candidate set within which the frequency bias can operate and hence to *obscure* frequency effects where missegmentation has occurred. That is, if the candidate set defined by the segmentation procedure contains the presented stimulus item, it may fail to be chosen if a higher-frequency candidate is available. If the candidate set does not contain the stimulus item, frequency effects on what is chosen will be solely determined by the frequency characteristics of the candidate set, which may well contain only words which are lower in frequency than the stimulus.

TABLE 7
OCCURRENCE OF LEXICAL VERSUS GRAMMATICAL WORDS FOLLOWING INSERTED WORD BOUNDARIES BEFORE STRONG VERSUS WEAK SYLLABLES IN FAINTLY HEARD SPEECH

| | Before strong | Before weak |
|---|---|---|
| Lexical | 103 | 8 |
| Grammatical | 18 | 33 |
| Nonsense word or dash | 25 | 8 |

TABLE 8
COMPARATIVE FREQUENCY OF LEXICAL WORDS IN INPUT VERSUS ERROR IN FAINTLY HEARD SPEECH, SEPARATELY FOR WORD BOUNDARY INSERTIONS AND DELETIONS BEFORE STRONG VERSUS WEAK SYLLABLES

| | Error higher in frequency than input | Error lower in frequency than input |
|---|---|---|
| Boundary insertions before strong syllables | 48 | 37 |
| Boundary insertions before weak syllables | 10 | 7 |
| Boundary deletions before strong syllables | 2 | 10 |
| Boundary deletions before weak syllables | 12 | 16 |

both subjects and items, we can check the consistency of our findings across both samples. Of the 18 subjects, 16 produced more errors predicted by the rhythmic segmentation hypothesis than not predicted (one produced seven unpredicted errors to six predicted errors, while the remaining subject was a tie); this distribution is significantly unlikely to have arisen by chance ($z = 3.4$, $p < .001$). Separately by types of error, 17 subjects produced more boundary insertions before strong than before weak syllables, with one tie (statistical evaluation unnecessary), and 14 subjects produced more deletions before weak than before strong syllables, two subjects produced a difference in the opposite direction, and there were two ties ($z = 2.75$, $p < .01$). Similarly, 30 of the 48 items elicited more errors predicted by the rhythmic segmentation hypothesis than not predicted, with 14 items eliciting more unpredicted than predicted errors, and four ties; again, the difference is significant ($z = 2.26$, $p < .02$). Separately by error types again, 27 items elicited more insertions before strong than before weak syllables, 15 items the reverse, with six ties ($z = 1.7$, $p < .05$), and 17 items elicited more deletions before weak than before strong syllables, eight items the reverse, with 23 ties ($z = 1.6$, $p < .055$). These results indicate that the pattern we have found is highly consistent over both subjects and items.

Second, we can undertake a more stringent test of the alternative hypothesis. According to this hypothesis, the pattern of responses simply falls out of the distribution of strong and weak syllables in the vocabulary, so that misperception of, say, *cadets* as *the deaths* occurs when a subject perceives the vowels accurately and then chooses (quasi-)randomly from the lexicon, in which words with [ɛ] in the first syllable greatly outnumber words with schwa in the first syllable and [ɛ] in the second. We can test the hypothesis by examining whether there is in fact any direct relationship between such asymmetries in the vocabulary and the asymmetries of response type which we observed.

The test is tractable because our stimulus materials contained only six strong vowels. We can be confident that for every one of these vowels, there will be more words with the vowel in stressed initial position than in second position preceded by a weak syllable; that is the way English is. However, the size of the asymmetry is likely to vary from vowel to vowel. If the alternative hypothesis is correct, the larger the asymmetry in the vocabulary, the larger should be the tendency toward errors consistent with the dominant pattern. Accordingly, we compared the actual distribution of these six vowels in the English vocabulary with the likelihood of each kind of boundary misplacement error occurring with each vowel.

Table 9 displays the relevant data. Rows 1–4 show the adjusted frequencies of each

TABLE 9
BOUNDARY MISPERCEPTIONS IN FAINTLY HEARD SPEECH AS A FUNCTION OF THE VOWEL IN THE STRONG SYLLABLE ADJACENT TO THE BOUNDARY (ROWS 1–4), WITH THE RELEVANT STATISTICS FOR THE DISTRIBUTION OF THOSE VOWELS IN THE BRITISH ENGLISH VOCABULARY (ROWS 5 AND 6)

| Vowel / Proportionally adjusted errors | [ɛ] | [e] | [ʌ] | [u] | [ɪ] | [i] |
|---|---|---|---|---|---|---|
| Boundary insertions before strong syllables | .106 | .067 | .127 | .144 | .185 | .185 |
| Boundary deletions before strong syllables | .011 | .040 | .011 | .005 | .012 | .009 |
| Boundary insertions before weak syllables | .028 | .062 | .061 | .062 | .028 | .069 |
| Boundary deletions before weak syllables | .081 | .062 | .033 | .009 | .050 | .033 |
| Ratio of strong-initial words to weak-initial words with strong second syllables | 4.388 | 6.836 | 13.419 | 4.419 | 5.747 | 5.894 |
| Ratio of weak-final words with strong penultimate syllable to strong-final words | 1.099 | 1.051 | .949 | .472 | .978 | .440 |

kind of boundary misplacement as a proportion of the opportunities available in our stimulus materials for that kind of error with that vowel. The adjustment was necessary because the opportunities were not exactly matched across vowels (they were matched across long versus short vowels). For instance, there were eleven words in the materials in which a syllable with [ɛ] was preceded by a weak syllable: *cadets, ascents, expect,* etc. Thus the opportunity for boundary insertions before syllables with [ɛ] was 11 (words) × 18 (subjects) = 198. The actual number of boundary insertions before syllables with [ɛ] was 21, so the adjusted figure in row 1 is 21/198 = .106. There were 10 words in the materials (excluding words in string-initial position) in which [ɛ] occurred in a word-initial syllable: *senseless, men, went,* etc., giving an opportunity for boundary deletion before a syllable with [ɛ]; and two boundary deletions in fact occurred before syllables with [ɛ], giving an adjusted figure of .011. There were eight words in the materials with [ɛ] in penultimate position, giving an opportunity for insertion of a boundary before a weak syllable following a syllable with [ɛ]: *senseless, tender, lessons,* etc.; the number of boundary insertions before weak syllables after syllables with [ɛ] was four, so the adjusted figure is .028. Finally, there were thirteen words in the materials (excluding words in string-final position) in which [ɛ] occurred in a word-final syllable, which offered the opportunity for deletion of a boundary before a weak syllable after a syllable with [ɛ]: *cadets, ascents, men, went,* etc. The actual number of boundary deletions before a weak syllable after a syllable containing [ɛ] was 19, so the adjusted figure is .081. The figures for the other five vowels were calculated in analogous fashion.

Rows 5 and 6 in the table contain the relevant lexical statistics. Since all our subjects, and the speaker who recorded our materials, spoke British English, the relevant vocabulary is British English, and we computed[5] these statistics from the *Longmans Dictionary of Contemporary English* (Procter, 1975). The figure in row 5 for each vowel is the ratio of words in the vocabulary with that vowel in initial syllable (examples for [ɛ]: *beg, chest, feather, residence, verisimilitude*) to words in the vocabulary with that vowel in the second syllable, preceded by a weak syllable (e.g., *affect, forgetful, suggestible, togetherness*). As predicted, this ratio is quite variable across the six vowels. The figure in row 6 is the ratio of words with the relevant vowel in penultimate syllable position, followed by a weak syllable (e.g., *feather, convention, disinfectant, superintendant*) to words with the same vowel in word-final syllable position (e.g., *beg, chest, affect,*

[5] Actually, James McQueen did this.

*disinfect, superintend*). This ratio can be seen to be rather less variable.

The principal prediction from the alternative hypothesis is that the greater the asymmetry between strong-initial words and weak-initial words, the greater the degree to which boundary insertions should outnumber boundary deletions before strong syllables; thus the larger a vowel's ratio in row 5, the more boundary insertions (row 1) should outnumber boundary deletions (row 2) before strong syllables containing that vowel. In fact, there is no significant correlation between these two ratios (and the $r$ value is negative, i.e., in the wrong direction): $r\,[5] = -.31$, $p > .5$.

A subsidiary prediction could be derived from the alternative hypothesis concerning boundary misplacements before weak syllables: boundary deletions should outnumber boundary insertions before weak syllables most heavily where schwa-final words most outnumber strong-final words. In other words, the larger a vowel's ratio in row 6, the greater the ratio of deletions (row 4) to insertions (row 3) before weak syllables. However, this relationship is also statistically unreliable: $r\,[5] = .71$, $p > .1$.

These findings give no support to the alternative hypothesis.[6] Neither in the pattern of frequency effects, nor in the relationship to vocabulary distributions, is there any sign that the pattern of juncture misperceptions simply arises from the structure of the vocabulary given selection of plausible or random word candidates. Thus in laboratory-induced juncture misperceptions, as in natural slips of the ear, we find the patterns predicted by the rhythmic segmentation hypothesis: listeners tend to insert boundaries before strong syllables and delete them before weak syllables; boundaries inserted before strong syllables tend to produce lexical words while boundaries inserted before weak syllables tend to produce grammatical words. That is, the rhythmic properties of the input guide listeners' hypotheses about the placement of lexical boundaries in imperfectly perceived speech.

[6] The pattern of errors across vowels does not correlate with the raw lexical statistics for each vowel, either.

## General Discussion

The rhythmic segmentation hypothesis proposes that listeners processing spoken English operate on the assumption that strong syllables are highly likely to be the initial syllables of lexical words, whereas weak syllables are most probably not word-initial or, if word-initial, are more likely to be grammatical words.

This hypothesis accurately accounts for the juncture errors which listeners produce when speech input is hard to perceive; the same patterns arise in naturally occurring errors and in laboratory-induced errors.

The original motivation for the rhythmic segmentation hypothesis was the assumption that to a certain extent speech is *always* hard to perceive. Speakers do not (in English, at least) produce consistent and reliable cues to the presence of a boundary at the word level. Yet listeners need to be able to locate word boundaries, because otherwise they cannot recognize what the speaker has said; recognition involves matching the input to stored representations, and our stored representations—our lexical memory—must contain discrete entries. We do not have the infinite storage space which would be required to contain a representation of every utterance with which we might possibly be presented. Therefore segmentation is a necessary operation; and the absence of reliable boundary cues makes it a difficult one.

Some psycholinguistic models of speech recognition assume, however, that segmentation is in practice not a problem, on the grounds that in the temporal flow of the recognition process the successful recognition of a word will ensure that whatever immediately follows that word will be known to be word-initial. The most explicit proposal of this type was made by Cole and Jakimik (1978), who proposed that recognition of

spoken utterances proceeds in strictly temporal order, so that "one word's recognition automatically directs segmentation of the immediately following word" (1978, p. 93). Such models will, of course, often work very well. There are utterances, for example, which admit of only one segmentation throughout; thus "some guy should now call" is easy to segment because the cross-boundary phoneme sequences [m g], [ai š] and [au k] do not occur word-internally in English (Lamel & Zue, 1984; Harrington, Johnson, & Cooper, 1987). In such cases the Cole and Jakimik model will produce perfect segmentation; and indeed, it will work well whenever the speech input is clear enough for the listener to be able to recognize each word as it is presented.

Unfortunately, however, these ideal conditions do not always exist. First, speech signals rarely offer only a single segmentation. In typical English speech the majority of words are monosyllabic (Cutler & Carter, 1987); and most monosyllabic English words do not become unique until at or after their final phoneme (Luce, 1986b)—*ball* is fully realized in *bald, bald* in *balderdash,* and so on. Thus it is not surprising to find that words—especially monosyllabic words—are in fact often not recognized until after their acoustic offset. Postoffset recognition has been demonstrated both with laboratory-produced (i.e., carefully read) speech (Grosjean, 1985), and, to an even greater extent, with spontaneously produced speech (Bard, Shillcock, & Altmann, 1988; Shillcock, Bard, & Spensley, 1988).

Second, speech signals are not always fully clear. Background noise, distance between speaker and listener, distortion of the speaker's vocal tract, foreign accents, slips of the tongue—all these, and similar factors, conspire to make the listener's phonetic interpretation task harder. Clearly, if listeners cannot be certain about the phonetic structure of the speech input, and if even phonetic certainty does not allow unambiguous word identification, the "automatic" segmentation proposed by Cole and Jakimik will quickly break down.

In fact, it is precisely under conditions of phonetic uncertainty that rhythmic segmentation proves most useful. Researchers in the field of automatic speech recognition (e.g., Shipman & Zue, 1982) have in recent years developed systematic representations of phonetic uncertainty, by replacing fully specified phonetic transcription with transcription in which only general classes of phoneme are provided; these may be broad classes (glide, nasal, stop consonant, etc.), or they may be rather more constrained (voiceless stop, back vowel, etc.). Two recent investigations using such imperfectly specified input have provided impressive support for the rhythmic segmentation hypothesis. In the first study, Briscoe (1989) implemented four lexical segmentation algorithms and tested their performance on a (phonetically transcribed) continuous input, using a 33000-word lexicon. The algorithms postulated potential lexical boundaries: (a) at the end of each successfully identified word (Cole & Jakimik's proposal); (b) at each phoneme boundary; (c) at each syllable onset; and (d) at each strong syllable onset (Cutler & Carter's rhythmic segmentation algorithm). The measure of performance was the number of potential lexical hypotheses generated (the fewer the better). With fully specified phonetic input all algorithms performed reasonably well. However, significant differences between the algorithms emerged when some or all of the input was incompletely specified; most noticeably, both the word-by-word algorithm and the phoneme-based algorithm suffered a severe performance decrement, generating huge numbers of potential parses of incomplete input. Far better performance was produced by the algorithms which constrained possible word onset positions in some way, and the more specific the constraints, the better the performance: the rhythmic segmentation algorithm performed best of all with the incomplete input.

In the second study, Harrington, Watson, and Cooper (1989) compared the rhythmic segmentation algorithm with a segmentation algorithm based on permissible phoneme sequences (Lamel & Zue, 1984; Harrington, Johnson, & Cooper, 1987), using as a metric the proportion of word boundaries correctly identified in a 145-utterance corpus. With partially specified input, phoneme sequence constraints proved virtually useless, but the rhythmic segmentation algorithm still performed effectively (in fact, it detected more word boundaries with the partially specified input than the phoneme sequence constraints had detected with fully specified input).

Thus the rhythmic segmentation hypothesis has proved not only amenable to implementation but also much more successful at locating word boundaries than alternative algorithms. Of course, its full implementation in an automatic speech recognition system would depend on front-end discrimination of strong versus weak syllables; we note, however, some encouraging preliminary results from the computer speech recognition literature which suggest that such discrimination is achievable (Sholicar & Fallside, 1988; Harrington, 1990).

The rhythmic segmentation hypothesis has found, therefore, a widely varying range of supporting evidence. Its appropriateness to the characteristics of English speech is assured in that its formulation is based on Cutler and Carter's (1987) distributional analysis. It accounts for the reported results from Cutler and Norris' (1988) word spotting task. It has performed well in comparison with alternative algorithms when implemented for computational studies. And as the present study has demonstrated, it correctly predicts the pattern of juncture misperceptions which occur in the recognition of continuous speech, both in spontaneous slips and in laboratory-induced misperception of faintly presented speech. An alternative hypothesis, in contrast, makes predictions about frequency effects and about correlations with vocabulary patterns which are unsupported by the data.

Faint speech, of course, is not the only laboratory method for eliciting misperceptions. We would predict that the same patterns would occur in the perception of noise-masked speech. And a recent pilot study in our laboratory has discovered the same pattern also in the perception of time-compressed speech. In this pilot experiment, part of a larger study by Young, Altmann, Cutler, and Norris, 30 listeners were presented with forty 18-syllable sentences of reasonable complexity at compression rates of 40 and 50%. Even with this high degree of compression, the listeners' comprehension was remarkably well preserved; the mean number of words correctly reported was over 70%. Elicitation of juncture misperceptions was not the primary purpose of this experiment. However, a number of such errors did occur in the subjects' responses, and of these errors there were 79 of the types predicted by the rhythmic segmentation hypothesis, but only 16 of the types which the hypothesis does not predict ($z = 6.36$, $p < .001$). Once again, in other words, the rhythmic segmentation hypothesis correctly predicts the boundary misperceptions which listeners make when listening conditions are difficult.

Even an algorithm as well adapted to the structure of the vocabulary as rhythmic segmentation, however, remains only a heuristic approximation; it does not always produce perfect results. Listeners apparently realize that rhythmic segmentation is of great use when perception is difficult; but it is by their misperceptions, i.e., the ways in which the rhythmic segmentation algorithm has misled them, that we can discern its operation. Since these occasional malfunctions do not seem to stop listeners relying on rhythmic segmentation, we must assume that their reliance is based on past experience: in general, the algorithm works very well indeed. One has to be a lert to spot it going wrong.

## APPENDIX

### *Experimental Sequences, Faint Speech Study*

Soon police were waiting
Dusty senseless drilling
Achieve her ways instead
Angels pinned beneath it
Rust presents a nuisance
Lou's bereaved disgraced him
A rustic settled hill
Within reviewed results
The newsmen seemed delayed
Making tinsel keyrings
The music's even pace
Tim approved results of
Trusting tender viewers
Machines create duress
Never just convict them
The eastern news remained
Distrust pretend balloons
Blinking lunar pulses
The hunters went fulfilled
Debates are grim relief
And cleaning Mabel's pets
Pete's display corrects it
Between secure campaigns
Eager rooster playing
Conduct ascents uphill
An eager rooster played
Sons expect enlistment
Music's even paces
Collect enough adrift
Depict a tool discussed
Cadets are just unfit
Readers playing lessons
Jets adjust equipment
Rings amused the sultan
Instruct the men confused
Mean baboons detained him
They're making wrinkled jeans
Leaders' claims expect it
The blinking lunar pulse
A better budget shift
Leaks reduced the traces
Butler's sense eclipsed them
Rust unchecked removes it
Includes serene refrains
Ornate distinct machines
The trusting slender loons
Better budget system
Hay begins beneath it

## REFERENCES


AHLBERG, J., & AHLBERG, A. (1982). *The ha ha bonk book*. London: Penguin.

BARD, E. G., SHILLCOCK, R. C., & ALTMANN, G. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, **44**, 395–408.

BENSON, R. W., DAVIS, H., HARRISON, C. E., HIRSCH, I. J., REYNOLDS, E. G., & SILVERMAN, S. R. (1951). C.I.D. Auditory tests W-1 and W-2. *Journal of the Acoustical Society of America*, **23**, 719.

BOND, Z. S. (1973). Perceptual errors in ordinary speech. *Zeitschrift für Phonetik*, **26**, 691–694.

BOND, Z. S., & GARNES, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

BRISCOE, E. J. (1989). Lexical access in connected speech recognition. *Proceedings of the 27th Congress, Association for Computational Linguistics, Vancouver* (pp. 84–90).

BROADBENT, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, **74**, 1–15.

BROWMAN, C. P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics*, **42**, 1–149.

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.

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.

COLE, R. A., & JAKIMIK, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.) *Strategies of Information Processing*. London: Academic Press.

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

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.

DUNKLING, L., & WRIGHT, G. (1987). *A Dictionary of Pub Names*. London: Routledge & Kegan Paul.

GARNES, S. (1977). *Folk etymologies as lexicalised slips of the ear*. Paper presented at the Twelfth International Congress of Linguists, Vienna.

GARNES, S., & 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).

GARNES, S., & 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.

GORDON-SALANT, S. M., & WIGHTMAN, F. L. (1983). Speech competition effects on synthetic stop-vowel perception by normal and hearing-impaired listeners. *Journal of the Acoustical Society of America,* 73, 1756–1765.

GROSJEAN, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics,* 38, 299–310.

GROSJEAN, F., & GEE, J. (1987). Prosodic structure and spoken word recognition. *Cognition,* 25, 135–155.

HARRINGTON, J. M. (1990). The acoustic basis of the distinction between strong and weak vowels. *Proceedings of the Third Australian International Conference on Speech Science and Technology, Melbourne,* (pp. 342–347).

HARRINGTON, J., JOHNSON, I., & COOPER, M. (1987). The application of phoneme sequence constraints to word boundary identification in automatic, continuous speech recognition. *Proceedings of the First European Conference on Speech Technology, Edinburgh* (Vol. 1, pp. 163–167).

HARRINGTON, J. M., WATSON, G., & COOPER, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language,* 3, 367–382.

HOWES, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America,* 29, 296–305.

LAMEL, L., & ZUE, V. W. (1984). Properties of consonant sequences within words and across word boundaries. *Proceedings of the 1984 International Conference on Acoustics, Speech and Signal Processing* (pp. 42.3.1–42.3.4).

LUCE, P. A. (1986a). *Neighborhoods of Words in the Mental Lexicon.* Ph.D. dissertation, Indiana University.

LUCE, P. A. (1986b). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics,* 39, 155–158.

MILLER, G. A., & NICELY, P. E. (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America,* 27, 338–353.

PROCTER, P. (Ed.) (1975). *Longmans Dictionary of Contemporary English.* London: Longman.

REES, N. (1979). *Graffitti lives, OK.* London: Unwin.

REES, N. (1980). *Graffitti 2.* London: Unwin.

REES, N. (1981). *Graffitti 3.* London: Unwin.

REES, N. (1982). *Graffitti 4.* London: Unwin.

SAVIN, H. B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America,* 35, 200–206.

SHILLCOCK, R. C., BARD, E. G., & SPENSLEY, F. (1988). Some prosodic effects on human word recognition in continuous speech. *Proceedings of SPEECH '88 (Seventh Symposium of the Federation of Acoustic Societies of Europe), Edinburgh* (pp. 819–826).

SHIPMAN, D. W., & ZUE, V. W. (1982). Properties of large lexicons: Implications for advanced isolated word recognition systems. *Proceedings of the 1982 International Conference on Acoustics, Speech and Signal Processing, Paris* (pp. 546–549).

SHOLICAR, J. R., & FALLSIDE, F. (1988). A prosodically and lexically constrained approach to continuous speech recognition. *Proceedings of the Second Australian International Conference on Speech Science and Technology, Sydney* (pp. 106–111).

SMITH, M. R., CUTLER, A., BUTTERFIELD, S., & NIMMO-SMITH, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research,* 32, 912–920.

(Received November 2, 1990)
(Revision received March 21, 1991)