SEGMENTATION ERRORS BY HUMAN LISTENERS: EVIDENCE FOR A PROSODIC SEGMENTATION STRATECY

Sally Butter-field 6 Anne Cutler

MRC Applied Psychology Unit, IS Chaucer Rd., Cambridge CB2 2EF, UK.

1. INTRODUCTION

The recognition of words in continuous speech is made more difficult by the absence of explicit word boundary markers. Psychologists modelling human speech recognition have suggested strategies by which listeners may compensate for the absence of explicit cues by postulating the most likely positions for word boundaries to occur. Cutler and Norris [1] proposed that listeners may base word boundary location in the English language on prosodic structure, by assuming that strong syllables (i.e. syllables containing a full vowel) are highly likely to be word-initial, but weak syllables (i.e. syllables containing a reduced or neutralised vowel) are not. This strategy seems to be very well adapted to the prosodic structure of the English vocabulary and the prosodic structure of actual speech samples. Cutler and Carter [2] showed that the majority of English words do indeed begin with strong syllables, and that in a large corpus of spontaneous British English speech approximately 90% of lexical words had strong initial syllables.

The absence of reliable correlates of the presence of a word boundary makes misperception of word boundary location in speech in principle easy. In the present study we examined listeners' misperceptions of continuous speech in the light of the word boundary location strategy proposed by Cutler and Norris. If listeners are indeed assuming strong syllables to be word-initial and weak syllables to be not word-initial, then we should find that word boundary misperceptions will be very unequally distributed across the four possible types of error. Specifically, we should find that erroneous insertion of a boundary before a strong syllable and erroneous deletion of a boundary before a weak syllable are relatively common, whereas erroneous Insertion of a boundary before a weak syllable and erroneous deletion of a boundary before a strong syllable are relatively rare. We examined both spontaneous and experimentally elicited misperceptions.

2. EVIDENCE FROM SPONTANEOUS SLIPS OF THE EAR

The psycholinguistic literature contains a number of studies of slips of the ear (e.g. Bond 6 Garnes [3], Browman [4]). We examined all the errors listed in these published studies plus all the slips of the ear included in the speech error collection assembled over several years by the second author. Among these we found 111 errors involving misplacement of a word boundary

SEGMENTATION ERRORS BY HUMAN LISTENERS

across at least one syllabic nucleus (that is. we excluded errors in which a boundary was misplaced across only one or two consonants - such as "an ice bucket" + "a nice bucket". Such errors are irrelevant to the present hypothesis.)

Since some slips of the ear involved more than one misplaced boundary (such as "won't bother me" + "lobotomy"). the total number of boundary errors was 139. It can be seen from Table 1 that in this small set of errors precisely the asymmetry predicted by the proposed strategy appears: insertions of a word boundary before a strong syllable and deletions of a word boundary before a weak syllable greatly outnumber Insertions of a boundary before a weak syllable or deletions of a boundary before a strong syllable. In all. there are 95 errors of the types predicted by the hypothesis and only 44 of the types not predicted.

TABLE 1.
WORD BOUNDARY MISPLACEMENTS IN
SPONTANEOUS SLIPS OF THE EAR

Boundary inserted before a strong syllable ("reverse" -> "your purse")	41	95
Boundary deleted before a weak syllable ("she'll officially" -> "Sheila Fishley")	54	
Boundary inserted before a weak syllable ("effective" -> "effect of")	21	44
Boundary deleted before a strong syllable ("any clips" -> "an eclipse")	23	

However, as Cutler and Carter [2] showed, the distribution of strong and weak syllables in spontaneous speech is far from even with respect to word boundaries. Therefore we cannot test the statistical significance of this asymmetry without comparing it to the distribution of strong and weak syllables in word-initial and non-initial position in the actually spoken utterances, in order to determine the pattern of error occurrence as a function of the varying opportunities for errors of different types. Since these errors were originally collected without such an analysis being planned, however, very little of the context in which the errors occurred was actually recorded (or. at least, reported in the publications we consulted). Accordingly, we found it impossible to arrive at an accurate estimate of the opportunities for different types of error in this corpus, which in turn meant that we could not conduct an appropriate test of the statistical significance of the markedly asymmetrical distribution of error types.

Note, however, that since Cutler £ Carter [2] found that about 75% of all strong syllables in a spontaneous speech sample were word-initial, it is reasonable to suppose that most natural speech offers many fewer opportunities for erroneous insertion of a boundary before a strong syllable than for deletion of a boundary before a strong syllable.

In a subsidiary analysis we examined the relative frequency of the words which were actually spoken versus the words' which were erroneously perceived. It may simply be the case that when listeners are presented with an utterance which 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 showed, words with strong initial syllables tend to occur more frequently than words with weak initial syllables.

The frequency analysis is not simple to perform. Firstly, grammatical words such as "the" and "of" have such a high frequency of occurrence that any error which includes a grammatical word which was 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. Therefore, we analysed the frequency of the lexical words alone. since it seems reasonable to suppose that if frequency effects are operative. they should show up in the lexical words.

Secondly, many of the slips of the ear involved proper names, the frequency of which is in the particular context impossible to assess. This reduced the number of error-target pairs in the frequency analysis to 73. Table 2 shows the number of pairs in which the error was more versus less frequent than the actually spoken utterance. There is an overall tendency for boundary insertions to result in higher frequencies, and boundary deletions to result in lower frequencies, as one would expect given that short words are more frequent than long words, and insertions are likely to result in the error having more, shorter words than the target, while deletions are likely to result in fewer, longer words. However, there is no overall tendency for the errors to contain higher frequency words than the targets (if anything, there is a trend in the opposite direction: $z = 1.28$); and there is no significant difference in the frequency effect for the errors predicted by the proposed strategy and the errors not predicted (X^2 (1) = 0.97).

Thus the evidence from spontaneous slips of the ear certainly suggests that listeners rely on a strategy of assuming strong syllables to be word-initial. However, slips of the ear occur infrequently and are difficult to collect; as noted above, they are also in many ways difficult to analyse. In particular, we could not adequately determine the statistical significance of the pattern of boundary misplacements. Therefore we investigated whether the same pattern of results would emerge in an experiment in which listeners were deliberately presented with hard-to-perceive utterances.

SEGMENTATION ERRORS BY HUMAN LISTENERS

TABLE 2.
COMPARATIVE FREQUENCY OF TARCET AND ERROR IN SPONTANEOUS SLIPS OF THE EAR

	Error higher in frequency than target	Error lower in frequency than target
Boundary inserted before a strong syllable	16	7
Boundary deleted before a weak syllable	5	18
Boundary inserted before a weak syllable	8	2
Boundary deleted before a strong syllable	1	16
	30	43

3. EVIDENCE FROM MISPERCEPTIONS OF FAINT SPEECH

predictable utterances (e.g. "rings amused the sultan"), of six ies each, were constructed. Within each utterance strong (S) and (W) syllables alternated. Half of the utterances had the pattern SW and half WSWSWS.

subjects were 18 members of the MRC Applied Psychology Unit Subject who were paid for participating in the experiment for one. hour. All arts were under SO years of age, and reported no hearing problems.

experiment began with a pre-test to estimate each subject's speech ion threshold. Two types of material were used in the pre-test: a read re and a list of spondees (i.e. words with two strong syllables, such toothbrush", "doormat", "workshop"). The read passage was a fairly ex passage containing statistical information. The list of 36 spondees taken from the CID Word Lists (Benson et al. [5]). One obvious item of can vocabulary was replaced ("sidewalk", replaced by "homework").

recorded materials were played over headphones from a Revox B77, cted to a step-attenuator. During the read passage, the attenuator was 20dB and the volume on the tape recorder was adjusted to produce audible speech at the headphones. The subjects were instructed to the volume knob on the tape recorder to the lowest level at which

they could follow the speech. For the spondee lists, the subjects were made familiar with all the items by reading a list and then listening to the items in alphabetical order. The attenuation was reduced so that each spondee was presented 15 dB above the previously established threshold. Subjects were asked to repeat the word heard. Then the subjects were presented with the randomised list, with the first item at least 5 dB above the level set for the read passage. The attenuation was increased 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.

The experiment then consisted of presentation of speech at this estimated threshold. Thus for presentation of the 48 test utterances, the attenuation was sat at the level of the estimated speech reception threshold separately for each subject. The subjects' task was to write down what they thought was said. lThay were asked to insert a dash if they were sure a syllable had bean spoken but they could not report any of it; this enabled us to analyse all responses on which the subjects had correctly determined the number of syllables spoken).

Of the 864 responses (18 listeners x 48 utterances), some were of course entirely correct. Others were non-existent, i.e. on occasion a subject produced no response at all to a given utterance. In very many cases the response consisted of a few syllables only. Although it was usually fairly obvious which syllables in the target utterance were being reproduced (for example, "wrinkle" as a response to "rings amused the sultan" is presumably based on the first two syllables), we decided to omit such cases from the analysis. We confined the analysis to responses which preserved the number of syllables (six) in the target utterance. Nearly half of the responses (414 in total) fell into this category, and 168 of these contained word boundary misplacements. Some responses contained more than one boundary misplacement, so the total number of errors available for analysis was 257. The distribution of these errors across the four possible error classes is shown in Table 3. It can be seen that, again, the pattern predicted by the proposed strategy emerges: erroneous insertions of a word boundary before a strong syllable and deletions of a word boundary before a weak syllable greatly outnumber insertions of a boundary before a weak syllable or deletions of a boundary before a strong syllable. Since the distribution of strong and weak syllables across the stimuli was known, we were able to compare the observed distribution of errors with the distribution which could be predicted from the available opportunities for each kind of error respectively. There were significantly more errors of the types predicted by the proposed strategy, and fewer errors of the types not predicted (x^2 (1) = 63.1. p < .001).

SEGMENTATION ERRORS BY HUMAN LISTENERS

TABLE 3.
WORD BOUNDARY MISPLACEMENTS
IN FAINTLY PERCEIVED SPEECH

Boundary inserted before a strong syllable ("sons expect enlistment" —> "some expect a blizzard**)	144	196
Boundary deleted before a weak syllable ("achieve her ways instead" —> "a cheaper way to stay")	52	
Boundary inserted before a weak syllable ("dusty senseless drilling" —> "thus he sent his drill In")	48	61
Boundary deleted before a strong syllable ("soon police were waiting" —> "soon to be awakened")	13	

Again, we compared the relative frequency of error and target: there was no overall tendency for errors to be of higher frequency than targets (z * 0.37). and no significant difference in frequency effects for the error types predicted and not predicted by the proposed strategy (x^2 (1) = 1.18).

4. CONCLUSION

Listeners' misperceptions both in spontaneous conversation and in experimental presentations of faint speech show an asymmetry in the misplacement of word boundaries: erroneous insertion of a boundary before a strong syllable and erroneous deletion of a boundary before a weak syllable are relatively common, while erroneous insertion of a boundary before a weak syllable and erroneous deletion of a boundary before a strong syllable are relatively rare. The pattern does not simply reflect a preference for higher over lower frequency responses. This asymmetry is precisely a* predicted by Cutler and Norris* [1] proposed strategy : listeners assume that strong syllables are most likely to be the initial syllables of lexical words.

5. ACKNOWLEDGEMENTS

This research was supported by a grant from the Alvey Directorate to Cambridge University, the Medical Research Council and STC Technology Limited. We thank Ian Nimmo-Smith for statistical advice.

TABLE 4.
COMPARATIVE FREQUENCY OF TARGET AND ERROR
IN FAINTLY PERCEIVED SPEECH

	Error higher in frequency than target	Error lower in frequency than target
Boundary inserted before a strong syllable	49	35
Boundary deleted before a weak syllable	12	17
Boundary inserted before a weak syllable	10	7
Boundary deleted before a strong syllable	1	8
	72	67

6. REFERENCES

[1] A CUTLER & D NORRIS.'The Role of Strong Syllables in Segmentation *for* Lexical Access',J Exp Psychol: Hum Perc £ Perf, 14 pp. 113-121 (1988)

[2] A CUTLER S D M CARTER.'The Predominance of Strong Initial Syllables in the English Vocabulary',Computer Sp & Lang. 3 (1988)

[3] Z S BOND S S CARNES.'Misperceptions of Fluent Speech', in R. Cole (Ed.) Perception and Production of Fluent Speech. Hillsdale. NJ (1980)

[4] C E BROWMAN.'Tip of the Tongue and Slip of the Ear: Implications for Language Processing', UCLA Working Papers in Phonetics, 42 (1978)

t5j R W BENSON. H DAVIS. C E HARRISON. I J HIRSCH. E G REYNOLDS. & S R SILVERMAN,'C.I.D. Auditory Tests YM and W-2'.Journal of the *Acoustical Society of America,* 23 p719 (1951)