Proceedings of The Institute of Acoustics

SEGMENTAL CUES TO SYNTACTIC STRUCTURE.

DONIA R. SCOTT AND ANNE CUTLER

LABORATORY OF EXPERIMENTAL PSYCHOLOGY, UNIVERSITY OF SUSSEX.

In recent years a number of studies in English speech perception have been concerned with finding acoustic correlates to syntactic boundaries. This topic is an important one for speech understanding since syntactic knowledge provides the simplest form of constraint on word sequences and since the meaning of any utterance is directly related to its syntactic structure. To date, work has been primarily concerned with the relationship between suprasegmental aspects of the speech wave (variations in the FO contour, amplitude and duration) and the presence of major syntactic boundaries (MSB's). Such studies have found that fall-rises in the FO contour, pauses, and the duration of phrase-final stressed syllables and of inter-stress intervals can provide valuable cues to the presence of a MSB (1 - 4). However, prosody is not the only possible acoustic cue to syntactic boundaries -- It is well known that speakers of English also produce segmental features as markers of MSB's. Phonological rules which normally apply across word boundaries are often inhibited when a word boundary is also a MSB. Examples of this are the rules of palatalization and flapping. Palatalization refers to the changing of an intervocalic /t/ to /tʃ/ or /d/ to /dʒ/ when followed by a /j/, giving /mɛtʃ(j)u/ instead of /mɛt(j)u/ or /dɪdʒ(j)u/ instead of /dɪd(j)u/. Flapping refers to intervocalic /t/'s or /d/'s in a falling stress pattern being changed to a flap (/ɾ/), thereby giving, for example, the same pronunciation of the words litre and leader (/liɾɾ/). The application of flapping and palatalization, however, are blocked by the presence of a MSB between the stop and following segment.

One of the major problems with these and other segmental features which are produced differentially as a function of the syntactic environment in which they occur (eg. intrusive r) is that, unlike the prosodic features, they differ across dialects. For example, r-intrusion only occurs in non-rhotic accents of English; intervocalic alveolar stops are commonly flapped in American, but not British, English; palatalization of /t/ and /d/ occurs in both American and British English. Segmental aspects of the speech wave can only be useful as syntactic markers to any speaker-independent, continuous speech recognition system if there is a significant relationship between the presence or absence of certain segmental features and the presence or absence of specific syntactic boundaries.

This paper describes an experiment which examines whether the presence or absence of an intervocal flapped or palatalized /t/ or /d/ can provide English listeners with a valuable cue to the presence or absence of a phrase boundary and whether the efficacy of such a cue varies according to whether the particular segmental distinction exists in the listener's own dialect. We looked at twelve syntactically ambiguous sentences. Six of these sentences contained a flapping environment at a potential phrase boundary. The other six contained a palatalization environment at a potential phrase boundary.

SEGMENTAL CUES TO SYNTACTIC STRUCTURE.

We had a native American produce the alternative readings of each sentence. He was told to flap or palatalize the /t/ or /d/ in the readings where there was not a phrase boundary between the two words and to release the stop when the phrase boundary was present. He was also instructed not to pause at the phrase boundaries, and not to produce contrastive stress or intonational cues. His productions were recorded and later judged by the two authors according to these criteria. Of several productions of each of the twenty-four sentences, we chose the best version for the experiment. Two additional versions of each of the chosen productions were then created. In one (the 2-word version), the two words of the relevant flapping or palatalization environment were replaced by the same two words from the alternative reading of the sentence. In the other (the consonant version), only the /t/ or /d/ of the flapping or palatalization environment was cross-spliced from the alternative reading. These two new versions, along with the original, were resynthesized with a new 'neutral' intonation contour, which was the arithmetic mean of the intonation contours of the two alternative readings produced by the speaker.

Three groups of subjects participated in the experiment. One group (the American group) consisted of 11 native speakers of American English who had been in Britain for less than two months at the time of the experiment. A second group (the British group) consisted of 11 native British speakers. These two groups were tested at Sussex University. A third group (the Chicago-British group) consisted of 14 Britons who had been living in the U.S. for some time (mean = 4.5 years). All members of this group were resident in Chicago and were tested at the University of Chicago. The subjects listened to all of the six versions we had created of each of the twelve sentences, and classified each utterance in terms of the two alternative meanings. Each subject heard each version of each sentence three times.

Since it was possible that alternative readings of the sentences were not equally likely for the different groups, a further test was devised in which people were asked to rate each of the sentences for plausibility. Two new groups provided rating controls for the American and British groups. The Chicago-British group acted as their own rating controls. The British raters found two of the sentences to be heavily biased towards one meaning; for the American raters, three of the sentences were found to be heavily biased; for the Chicago-British group none of the sentences were found to be biased.

For each group, responses to only those sentences which were judged to be ambiguous in the rating test were submitted to statistical analysis. Subjects' responses were scored in terms of the number of times (out of a total of three) that the reading containing a phrase boundary between the two critical words was chosen for each stimulus item. These scores were subjected to separate analyses of variance. The results of these analyses showed that for the American group, both flapping and palatalization provided cues to the absence of a phrase boundary. Stimuli which contained a flapped or palatalized /t/ or /d/ at the ambiguous phrase boundary received fewer phrase boundary judgements than stimuli which were identical in every way except for a released /t/ or /d/ at the ambiguous boundary. These differences were significant on Min F'. For the British group, neither flapping nor palatalization was a useful cue to the presence or absence of the phrase boundary. The main effect to flapping found

with the American group was also displayed by the Chicago-British group: sentences with flapping attracted significantly fewer phrase boundary judgements than sentences without flapping. But, as with the British group, palatalization was not an effective cue for the Chicago-British group.

A possible explanation of these results can be found in the subjects' own productions. Before they listened to the tape, we recorded each subject reading each sentence set in two different contexts: one which suggested the phrase-boundary reading, and one which suggested the no-phrase boundary reading. Preliminary results of transcriptions of each subject's utterances by one of the authors and two phonetically-trained colleagues reveal that the native American group used both flapping and palatalization differentially in the different syntactic conditions. They never flap when a phrase boundary is present, but do so over 40% of the time when there is no phrase boundary present; they palatalize in over 40% of the no-phrase-boundary readings, but in only about 15% of the phrase-boundary readings. The British group flapped neither when the phrase boundary was present nor when it was absent and palatalized about equally often in both conditions. The Chicago-British group were more or less the same as the British group: they palatalized equally often in the two conditions and flapped in only 2 of the 66 no-phrase-boundary readings. This group, however, differed from the British group in that whereas the British speakers always produced clear /t/'s instead of flaps, the Chicago-British speakers often produced /d/'s where the Americans flapped. This suggests that the speech of the Chicago-Britishers may be undergoing the process of acquiring the flap feature. To Britishers, flaps sound more like /d/'s than /t/'s.

The results of this study show that segmental differences can be used to parse English utterances. The presence of a flap should be a reliable cue to the absence of a MSB for the machine speech recognizer since this feature is only ever present when there is not a following MSB. The presence of palatalization, however, will be a less reliable cue except in cases where the input is restricted to American English.

These results also suggest that speech synthesis programs which include rules for generating segmental markers of MSB's should produce more natural-sounding and intelligible speech than those which do not include such rules. Indeed, Klatt has a flapping rule in his synthesis program and remarks of it and others like it that they are "...extremely important. They are not "sloppy speech" rules, but rather rules which aid the listener in hypothesizing the location of word and phrase boundaries" (5). A perceptual evaluation of Klatt's synthetic speech has shown that the inclusion of rules which modify the realization of a segment as a function of syntax and segmental context is of significant importance for naturalness and intelligibility (6).

References

1. 't Hart, J. and Cohen, A. (1973), "Intonation by rule: a perceptual quest", J. Phon., pp. 309-327.

2. Lehiste, I., Olive, J.P., and Streeter, L. (1976), "Role of duration in disambiguating syntactically ambiguous sentences", J. Acoust. Soc. Am. 60, pp. 1199-1202.

3. Streeter, L. (1978), "Acoustic determinants of phrase boundary perception", J. Acoust. Soc. Am., 64, pp. 1582-1592.

4. Scott, D.R. (1982), "Duration as a cue to the perception of a phrase boundary", J. Acoust. Soc. Am., 71, pp. 996-1007.

5. Klatt, D.N. (1979), "Synthesis by rule of segmental durations in English sentences", Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen. pp. 290-297.

6. Carlson, R., Granstrom, B., and Klatt, D.H. (1979), "Some notes on the perception of temporal patterns in speech", Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen. pp. 260-267.