ALLOPHONIC AND PROSODIC CUES FOR PARSING SPEECH

Lloyd H. Nakatani, Bell Laboratories, Murray Hill, N. J., U. S. A.

A theory of speech perception must explain how listeners hear discrete words in a continuous acoustic signal. We show that listeners hear words by dividing and combining stretches of the speech stream -- that is, by parsing speech -- into short word-sized portions which are likely to be actual English words. Parsing is done perceptually from allophonic and prosodic cues, not inferentially from syntactic and semantic knowledge. In this view, speech perception goes from continuous speech to discrete words from the bottom up, not from the top down.

Speech is parsed with the aid of allophonic and prosodic variations which function as either fission or fusion cues. Fission cues indicate portions of the speech which are divided by a word boundary. Examples of fission cues are (1) allophonic variations such as aspiration of word-initial voiceless stops, and glottalized onset of word-initial stressed vowels; and (2) prosodic stress and rhythm cues such as consecutive primary stressed syllables which must perforce belong to different words, and a long stressed syllable which is probably a monosyllabic word or phrase-final syllable and therefore must be followed by a word boundary.

Fusion cues, by contrast, cause portions of the speech to fuse perceptually so that a word boundary cannot divide the portions. Examples of fusion cues are (1) allophonic variations such as the syllabic nasal in "maiden" where the /d/ and /n/ are fused, and (2) prosodic stress and rhythm cues such as an unstressed syllable (other than a function word) which must be part of a polysyllabic word, and a short stressed syllable which is probably a non-final syllable of a polysyllabic word formed by fusion with a following unstressed syllable.

Our experiments show that fission and fusion cues are important for parsing speech. But they are not enough. Listeners probably also hear function words and affixes, and use their knowledge of where these sounds occur in English to parse speech. Experiments are planned to see if function words and affixes are used in parsing.