Perception of Second-Formant Transitions in Non-Speech Patterns

We have begun pilot tests comparing the ability of subjects to discriminate differences in direction and extent of the second formant transition when this transition is embedded in a two-formant pattern (speech) or a one-formant pattern (non-speech).

Direction and extent of second-formant transition have been shown to be an important acoustic cue for perception of place of articulation of stops. Liberman et al. (1957), using synthetic two-formant patterns, demonstrated that in a speech context the second-formant transition is perceived categorically; subjects are able to discriminate easily patterns differing in frequency of onset of the second-formant transition when the two stimuli fall in different phoneme categories, while stimuli falling within the phoneme boundary are largely indiscriminable. As categorical perception has not been demonstrated along a non-speech acoustic continuum, this result suggests the existence of separate processing systems for speech and non-speech. In particular, it has been hypothesized that speech is processed by reference to articulation and that categorical perception reflects the categorical nature of the commands to the articulators. However, the absence of data on the discriminability of transition-like non-speech signals makes it impossible to rule out a purely auditory explanation of the peaks in discrimination at the phoneme boundaries.

The single-formant patterns used in this pilot study, which were produced on the Pattern Playback are duplicates of the second-formants from the stimuli used in the 1957 study. The isolated formant sounds like a whistle; it is clearly non-speech.

There are thirteen stimuli each consisting of a transition (50 msec.) followed by a steady state portion (200 msec.) centered at 1800 cps.; the starting point of the transition varies in steps of 120 cps between 1080 cps and 2520 cps.

The stimuli were arranged in ABX triads for the discrimination tests. One-step, two-step, and three-step comparisons were constructed (33 comparisons in all) and all four combinations of each comparison were included (ABA, ABB, BAA, BAB). Eight tapes were prepared, each with a different randomization of the thirty-three comparisons. In four of the eight tapes the patterns are reversed so that the transition follows the steady-state portions of the signal, providing a comparison of the discriminability of the transitions in initial and final position in the signal. Each of three subjects went through the set of eight tapes four times, yielding for each comparison 32 judgments per subject.

The most interesting finding thus far is that the patterns in which the transitions precede the steady state are indiscriminable while the same patterns reversed are discriminated well above chance level. For these latter, there was a decrease in errors as the number of steps between stimuli increased; there was no such improvement for transitions in initial position. This result does, indeed, suggest that speech and non-speech signals are processed differently, since it is known that (voiced) stop consonants in initial position are more easily identified than their unreleased counterparts in final position in the syllable, the opposite of what appears to be true for the non-speech case. We intend to investigate this position effect experimentally using two-formant (speech) patterns.

The data are ambiguous with respect to the question of categorical or continuous perception. Some peaks appear, but these are not readily explained on theoretical grounds and may reflect faulty stimuli or simply the great variability resulting from the difficulty of the discrimination. Since the judgments are so difficult, we have decided to make up new tests using the computer-controlled synthesizer, permitting better control of all stimulus parameters.

Emily Kirstein