ORCHESTRATING ACOUSTIC CUES TO LINGUISTIC EFFECT

LEIGH LISKER

University of Pennsylvania
and
Haskins Laboratories

ABSTRACT

A most convincing way to demonstrate that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with respect to that property, but in nature such pairs are most unlikely. The English words rapid and rabid are a minimal pair at the level of the segmental phoneme, and are near minimal at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. Three properties lend themselves to simple waveform editing: the duration of the stressed vowel, the duration of the closure, and the glottal buzz vs silence of the closure signal. Listener responses to stimuli having natural values of these properties show that, with a single exception, there was no decisive effect on word identification produced by a shift in the value of any one property. At least two properties had to be changed to achieve any significant effect.

Phonetic research nowadays considers the processes involved in speech communication from a wide variety of perspectives, but a central concern remains that of identifying and characterizing those features of the speech processes that serve a message-differentiating function. The phonetic analysis of a speech signal into a temporal sequence of "sounds," as well as the decomposition of those sounds into features, provide a framework within which to specify the distinctive properties that determine a particular interpretation of the signal. A coherent account of a given speech event, considered as representative of a set of linguistically equivalent events, states the interrelations among physiological, anatomical and acoustic patterns, and the nature of their connection to the listener responses they elicit. By far the most attention has been given to finding the acoustic cues to the linguistic message conveyed by a vocal tract emission. The search has involved the acoustic analysis of signals, the selection of promising cue candidates, and the empirical assessment of their cue value by the methods of speech synthesis. Such evaluation of a feature's cue value typically has involved the use of sets of acoustic patterns designed to maximize the likelihood that the feature of interest will affect listeners' response behavior. The number of acoustic pattern features that have been found to have measurable cue value is uncertain, and presumably with continued research along established lines that number will only increase. Clearly it is easier to show that a feature has cue value than to justify a claim to the contrary (the famous unprovability of the null hypothesis).

Most of the acoustic cues so far uncovered are referred to as segmental cues, or even as cues to particular phonetic features of segments. The experimental data supporting their identification are derived via some variant of the linguist's "minimal pair" test. A most convincing way to show that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with respect to that property, but in nature such pairs are very unlikely. The English words rapid rabid make a minimal pair at the level of the segmental phoneme, and an almost minimal one at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. It is not certain, however, that any one of them is an independent cue, i.e. one that is capable of signaling a lexical distinction by itself. Even if a given acoustic property can be shone to have such power to affect perception, it need not be true that this property functions independently in nature.

Here I want to report some listener responses to sets of stimuli derived by waveform editing of some naturally produced tokens of rapid and rabid. Three properties served as experimental variables: the duration of the closure interval, the glottal buzz/silence difference during closure, and the duration of the pre-closure vowel. Unlike many tests of this kind, in which the values assigned a variable range over a span in steps of a size designed to establish category boundaries, in the tests reported on here each variable was given just two values, each chosen on the basis of naturalness.

A token of each of the sentences I think it's rapid and I think it's rabid was recorded by a speaker of a central eastcoast variety of American English and stored on computer. A waveform editing program was applied to produce a total of sixteen different acoustic patterns. The durations of the intervals corresponding to the labial closures were set to values of 60 and 120 msecs, these being typical of /b/ and /p/ closures for the speaker. The closure intervals were either acoustically blank or filled with buzz derived from the original /b/ closure. The pre-closure intervals, from the cessation of the noise interval marking the /s/ preceding the target word to the beginning of closure, were set at the following values:-for derivatives of rabid: 270 msec, the original value, and 230 msecs; for rapid derivatives: 190 msecs, the original value, and 230 msec. The common value of 230 msecs was selected because it fell within the range of natural values for both words in the sentence context used. (Shortening the pre-closure span to a duration of 190 msecs effected a noticeable shift in vowel quality.) A test order in which each of the sixteen stimuli was presented five times, i.e. a random order of eighty items, was presented to twelve native American English speakers, all linguistically and phonetically naive. Each test item was composed of an acoustically invariant carrier I think it's followed by the target word to be identified. Listeners' responses were the following:

Source: rabid	Operation	% "rabid"
1)	none	100
2)	-voicing	95
3)	+long closure	100
4)	-long vowel	100
5)	+long closure -long vowel	97
6)	-voicing -long vowel	93
7)	-voicing +long closure	15
8)	-voicing +long closure -long vowel	8

Source: rapid	Operation	% "rapid"
1)	none	100
2)	+long vowel	100
3)	-long closure	98
4)	+voicing	12
5)	+long vowel -long closure	62
6)	+long vowel +voicing	10
7)	-long closure +voicing	13
8)	+long vowel -long closure +voicing	8

For each of the variables a change to a value not normally associated with the original stimulus type has, with one exception, no great effect on labeling behavior. Only when glottal buzz replaces the silence of the /p/ closure is there a decided shift to "rabid" judgments.

It does not follow, of course, that the three features are of negligible importance for the perception of the two words. Thus a combination of devoicing and lengthening of the /b/ closure elicited an overwhelmingly "rapid" response, a result in conformity with earlier findings. A shortening of the /p/ closure together with a lengthening of the preceding vocalic interval yielded mostly "rabid" responses. Original "rapid" was heard largely as "rabid," while "rabid" went to "rapid" when all three variable features were assigned values appropriate to the competing form.

The results summarized above indicate that an acoustic feature to which cue value has been attributed does not always effect a significant effect on linguistic labeling behavior; its effect is quite context-dependent. Indeed it may well be, in the case of certain properties, that the context in which it can be decisive can only (?) be contrived in the laboratory. The status of an acoustic feature of speech is therefore very different from that of a phonetic feature, which we generally suppose to possess the power, for at least some natural phonetic system, to mark differentially some words from others, and to do this independently of other phonetic features.