Orchestrating Acoustic Cues to Linguistic Effect*

Leigh Lisker†

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 versus 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. Adding glottal buzz to the /pJ/ closure led listeners to report the word “rabid.” To transform original “rabid” to “rapid,” at least two properties had to be changed to achieve any significant effect, one of these necessarily the replacement of closure buzz by silence.

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 identical events, states the mutual interrelations among physiological, anatomical, and acoustic patterns, and how they relate to the listener responses they elicit. By far the greatest attention has been given to finding the acoustic cues to the linguistic message conveyed by a speech signal. The search has involved the analysis of speech signals, the selection of promising cue candidates, and the experimental assessment of their cue value by the methods of speech synthesis and tests of perception. Such evaluation of a feature’s cue value typically has involved the use 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 determined to have cue value is not known with certainty, 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 cues to particular phonetic features of segments. The experimental data supporting their identification are derived by means of some variant of the linguist’s “minimal pair” test. 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 unlikely. The English words *rapid* and *rabid* make a minimal pair at the level of the segmental phoneme, and almost minimal 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, that is, one that can signal a lexical distinction independently of its acoustic context. Furthermore, even if a given property can be shown to have such power to affect listener responses, it need not be true that this property functions independently in natural speech.

In the following 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 closure duration interval, the glottal buzz/silence difference during closure, and the duration of the steadystate /a/ vowel. Unlike many tests of this kind, in which the values assigned a variable are altered in steps of a size designed to establish one or more category boundaries, in the test here reported each variable was given just two ( for one variable three) values, these being chosen on the basis of naturalness. This does not mean that the stimuli thus derived can be called natural, only that the values assigned the experimental variables were found in natural utterances recorded by the talker who served as the signal source.

From a set of recordings of the expressions *I think it's rapid* and *I think it's rabid* produced by a male speaker of a central East Coast variety of American English a typical token of each sentence was selected, digitized, and stored on computer. A waveform editing program was applied to produce a total of sixteen stimuli having acoustic differences restricted to the intervals corresponding to the final words of the utterances. The durations of the acoustic intervals corresponding to the labial closures were given two values, 60 and 115 ms, these being values typical of the talker's productions of the /p/ and /b/ closures in the particular context. 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 the labial closure, were set to the following values: for derivatives of *rabid* the mean /b/ value of 270 ms and a shorter duration of 230 ms; for *rapid* derivatives the mean /p/ value of 190 ms and an increased duration of 230 ms. The common value of 230 ms was chosen because it fell within the range of natural values for both words in the sentence context used. (Shortening the pre-closure span of *rabid* to 190 ms, the mean /p/ duration, effected a noticeable shift in vowel quality.) A test order in which each of the sixteen stimuli was presented five times, that is, 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.

The figure illustrates the perceptual effects of the several shifts in the values of the three acoustic properties. The upper panel represents changes in the perception of the original *rabid* token, while the lower panel shows the effect of editing the original *rapid*. 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 variables are otherwise 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. But a shortening of the /p/ closure together with a lengthening of the preceding vocalic interval still yielded mostly "rapid" judgments. Original "rapid" was heard largely as "rabid," while "rabid" went to

*Lisker*
"rapid," when all three variable features were assigned values appropriate to the competing form.

**STEPWISE FEATURE MODULATION**

Vowel Duration (VwD) Closure Duration (CD)

± Closure Buzz

(N = 60; 12 Ss x 5 Trials)

<table>
<thead>
<tr>
<th>Key</th>
<th>% “rabid”</th>
<th>% “rapid”</th>
</tr>
</thead>
<tbody>
<tr>
<td>RABID</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>RAPID</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

**Stimulus:** [rəbid] [rəbid] [rəbid] [rəbid] [rəpid] [rəpid] [rəpid] [rəpid] [rəpid]

**Operations:** None - VwD + CD - VwD - Buzz - VwD + CD - Buzz

**RAPID → RABID**

<table>
<thead>
<tr>
<th>Stimulus:</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
<th>[rəpid]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operations:</td>
<td>None</td>
<td>- VwD</td>
<td>- CD</td>
<td>+ VwD</td>
<td>- CD</td>
<td>+ Buzz</td>
<td>+ VwD</td>
<td>- CD</td>
</tr>
</tbody>
</table>

The results summarized above indicate that changing the value of an acoustic feature to which cue value has been attributed does not always produce 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 they can be decisive can only (?) be contrived in the laboratory. The status of an acoustic property 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 systems, to

_Orchestrating Acoustic Cues_
mark differentially some words from others, and to do this independently of other phonetic features.

ACKNOWLEDGMENT

This work was supported by NICHD Grant HD-01994 to Haskins Laboratories.

FOOTNOTES

*Also University of Pennsylvania