Rapid vs. Rabid: A Catalogue of Acoustic Features That May Cue the Distinction*

Leigh Lisker+

ABSTRACT

In American English, initial /bdg/ often lack the acoustic feature taken as the defining feature of voiced stops; intervocally before unstressed vowel /ptk/ lack aspiration, without which initial stops are not labeled "ptk." Initially the two categories differed in the timing of vocal fold adduction and onset of fold vibration; several acoustic cues, all tied to the VOT difference, have been studied. Medially there is also a difference in the management of the larynx, though it results in a phonetically simpler contrast, one of voicing with no accompanying difference in aspiration. Acoustically, however, the list of features that play, or might plausibly play a role is quite large. The word pair rapid–rabid, for example, might be affected by the following: 1) presence/absence of low frequency buzz during the closure interval; 2) duration of closure; 3) $F_1$ offset frequency before closure; 4) $F_1$ offset transition duration; 5) $F_1$ onset frequency following closure; 6) $F_1$ onset transition duration; 7) [æ] duration; 8) $F_1$ "cutback" before closure; 9) $F_1$ cutback following closure; 10) VOT cutback before closure; 11) VOT delay after closure; 12) $F_0$ contour before closure; 13) $F_0$ contour after closure; 14) amplitude of [ɪ] relative to [æ]; 15) decay time of glottal signal preceding closure; 16) intensity of burst following closure. Even if some of these should turn out to be perceptually negligible, enough of them surely have cue value to make it a formidable task to justify preferring an acoustic to an articulatory account of the distinction between the two English words.

If stop voicing continues to be a subject of lively interest to students of speech, it must be because it continues to provoke new questions or to refuse final answers to old ones. Perhaps this is because the stops of American English are not well chosen as the object of investigation whose purpose is to construct or test hypotheses concerning the perception of some single phonetic feature difference such as voicing. The phonetic differences between the American English phonemes /b/ and /p/, /d/ and /t/, and /g/ and /k/ are several, and we do violence to the internationally accepted definition of the term "voiced stop" if we call the phoneme set /bdg/ voiced, particular-

*This paper was presented at the 94th Meeting of the Acoustical Society of America, Miami Beach, December 1977.

+Also University of Pennsylvania.

Acknowledgment: The support of the National Institute of Child Health and Human Development is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-54 (1978)]
ly if attention is focused on these categories in initial position. To refer
to the acoustic features by which /bdg/ are distinguished from /ptk/ as cues
to stop voicing is to produce a terminological muddle on two counts: 1)
initial /bdg/ need not be, and often are not, produced with glottal pulsing
before release, so that they may be voiceless in the sense of the word as used
by the International Phonetic Association; and 2) the phonemes /bdg/ and /ptk/
are distinguished in different ways in different contexts, and constancy with
respect to the phonetic feature of voicing is not a property of the
phonologically contrasting sets.

Putting aside the matter of terminological purity and any question of
cross-language validity, it does seem strange to construct hypotheses of stop
voicing perception on the basis of the initial stops of American English,
where the data derived from studying these events cannot represent the full
range of phenomena that a theory of American English stop production and
perception must encompass to be adequate. On the basis of traditional
phonetic descriptions of these stops, and they seem to be taken seriously by
most of us, it is hard to understand how any single detector, or detector-
pair, could yield "outputs" that match the labeling behavior of English-
speaking listeners. A detector that fires in response, let us say, to a
periodically-excited transition following closure can tell the host-listener
that a /bdg/ has occurred, but what inhibits a like response to medial /ptk/?
The context-dependent nature of the phonetic differences between the two stop
category sets of American English is a very old story for linguists, and it
was in recognition of this fact that they were long disposed to find that the
basis of the phonemic distinction was not one of voicing at all, but of
something else they called "force of articulation." Voicing, it was sometimes
said, is irrelevant to the contrast. More recently, a dimension of relative
voice onset time (VOT) has been promoted as a measure by which to describe the
difference between aspirated and unaspirated initial stops and the difference
between voiced and voiceless medial stops, thereby avoiding recourse to the
different and less accessible level of description at which something like
"articulatory force" might be discovered and measured.

Medially in words before unstressed vowels American English /bdg/ are
most often voiced, in the strict IPA sense, particularly where the signal both
preceding and following closure is voiced. Just as commonly in that context,
members of the other category set show voiceless closures. Here then is the
place where the acoustic features that serve to distinguish the two sets of
phonemes can be said to cue stop voicing. Oddly enough, although in this
position the phonetic difference between the sets is considered to be smaller
than it is in initial position, the number of acoustic pattern features whose
manipulation may affect the labeling of a stoplike interval is much larger.
However, it is only odd if we suppose that the number of phonetic features
that differentiate the contrasting sets should determine the number of
acoustic features we can isolate and manipulate to linguistic effect. Otherwise it is not particularly surprising: the initial stops cannot be cued
by features preceding closure, nor are they usually cued by any feature of the
closure interval itself. Of sixteen acoustic features that can, or can
plausibly be supposed to serve as cues in the identification of members of a
word pair like rapid-rabid, seven are to be found in the signal preceding the
medial closure. These are: 1) duration of the [æ] vowel; 2) F₁ closing
transition duration; 3) F₁ offset frequency before closure; 4) F₁ "cutback"
Figure 1: Responses of American listeners to stimuli derived from naturally produced tokens of the words rabid and rapid. Waveforms were edited to vary closure duration from 30 to 150 msec, in 15 msec steps. Closure intervals were either acoustically blank (-buzz) or entirely filled by laryngeally produced signal (+buzz) derived from the originally recorded token of rabid.

Figure 2: Responses of 12 listeners, all native speakers of American English, to stimuli derived by editing naturally produced tokens of caliber and caliper. All closure durations tested, except for 120 msec in the case of the caliper-derivatives and 130 msec for the caliper-derivatives, were acoustically blank.
Figure 3: Responses of English-speaking listeners to stimuli derived from four naturally produced tokens of rapid. Closure durations were set at three values: 133 msec (an appropriate value for /p/), 87 msec (an appropriate /b/ value), and 60 msec (a value appropriate for /b/ and inappropriate for /p/).

Figure 4: Responses of six English-speaking subjects to stimuli, all with acoustically silent closures, derived from naturally produced tokens of locker and lager. The data represented in upper and lower panels are the same; the lower display shows that the six subjects could be divided into two groups with rather different response patterns.
before closure; 5) timing of voice offset before closure; 6) F₀ contour before closure; 7) decay time of glottal signal at closure. Another feature, the intensity balance of the [ə] and [ɪ] vowels, which may affect stress-placement judgments and secondarily the evaluation of certain VOT values, is also restricted, obviously, to medial position. Two features of the closure interval itself, its acoustic nature and its duration, also play little or no role in initial position. The remaining six features, which are measures of the signal from release to following vowel, are simple mirror images of the features of the closing transition already mentioned.

Of the sixteen features that might affect the identification of a signal as rapid or rabid, it is probably true that none is indispensable and that several play no significant role in the perception of unedited naturally produced tokens of these words. If we are talking of acoustic cues, we do not limit ourselves to the perceptual evaluation of normally produced signals, and in particular we do not refrain from treating as independent variables features that are not independent in natural speech. If my reading of the experimental phonetic literature is correct, the conditions that an acoustic feature must satisfy in order to be called a "cue" do not involve a demonstrable conformity with nature; it is enough that patterns be devised so that manipulating the single feature affects a significant shift in listeners' word identification—say from "rapid" to "rabid." There is no requirement, it seems, that constant features of the test stimuli be copied from nature.

If the medial stops of rapid and rabid are correctly called voiceless and voiced respectively, it should suffice that they differ acoustically only over the interval corresponding to closure. In fact, we can delete the buzz from the closure of a naturally produced token of rabid to elicit "rapid" judgments. Moreover, a normal token of rapid with low-frequency buzz replacing its silent closure will be identified as "rapid," if the buzz is carefully tailored to its context. Figure 1 shows what listeners reported when tokens of the two words were provided with silent and buzzed closure intervals over various durations. Buzz-filled closures elicited mainly "rabid," with no effect of duration worth mentioning. Silent intervals were most often interpreted as /p/, although not when the closure was very brief. Thus it appears that buzz of any duration is incompatible with /p/, and that silent closures longer than about 100 msec preclude "rabid" responses.

Similar results were obtained by the same kind of editing of other naturally produced word pairs. In Figure 2, for example, data are the responses to stimuli derived from productions of the words caliber and caliper, where the closures were intended to be all silent, and the effect of varying these silent intervals was the object of interest. The unexpected results for caliper with 120 msec of silence and for caliber with 130 msec of silence is explained very simply; I made a mistake in fabricating these stimuli, providing them with buzzed instead of silent closures. However, this inadvertently obtained corroboration of the earlier finding, namely that the acoustic nature of the closure interval can determine listeners' responses, does not mean that we cannot also fail to obtain the same results with other natural tokens of words of this kind. When four different productions of rapid were tested with silent intervals copied from nature, they yielded the results shown in Figure 3. Two tokens were never heard as anything but "rapid," although the shortest duration tested was shorter than any /b/-
duration observed in the speaker's productions of rabid; one token of rapid was ambiguous for the shortest duration tested; only one was heard more often as "rabid" (66 percent) for this same duration. Possibly still shorter durations would have elicited more "rabid" judgments. Still other word pairs, locker-lager for example (Figure 4), resisted this treatment: locker with zero closure duration was still identified almost 50 percent as "locker," and lager with 140 msec of silent interval was reported equally often as "lager" and "locker." This result, possibly peculiar to the /g/-/k/ contrast, raises some questions for future discussion.

Measuring the cue values of the features preceding and following closure by presenting separately the first and second syllables of the natural speech disyllables gives data that are not easy to interpret; the part of rapid preceding closure is often not clearly heard to terminate in a /p/-closure, while the signal following closure is generally reported to begin with "b." It seems reasonable to suppose that in natural productions of a word like rapid, everything up to closure may be ambiguous so far as to whether a /b/ or a /p/ closure was executed, and that what follows closure is unambiguously /b/ if heard in isolation. Buzz in the intervening interval is decisive, while a short silent interval may or may not have the same outcome as buzz.

These results with natural speech can be described as less than clearcut, presumably because natural productions that are linguistic and even phonetic repetitions can differ in acoustic features that can be made to bear a greater perceptual load than they ordinarily have before mutilation of the signal. By pure synthesis it is possible to obtain results to indicate that silent intervals of any duration can be compatible with "b"; if the offset and onset frequencies of F1 are very low, only "rabid" will be reported. In such patterns, the feature of [æ]-duration, which under other conditions may be decisive, can have no important effect on word identification.

In summary, in medial position the basis of the contrast between stop category sets is better identified with what happens in the closure interval than in other positions, but events elsewhere can make the closure interval alone insufficient to explain labeling behavior. The ensemble of features, spread over two syllables, shows a degree of disparity at the purely acoustic level that seems strange, given that they all affect the same phonetic judgment. However, they can all be referred back to a single crucial articulatory difference in the management of the larynx.