LANGUAGE AND SPEECH, Vol. 29, Part 1, 1986

"VOICING" IN ENGLISH: A CATALOGUE OF ACOUSTIC FEATURES SIGNALING /b/ VERSUS /p/ IN TROCHEES*

LEIGH LISKER
University of Pennsylvania
and
Haskins Laboratories

The English category sets /b, d, g/ and /p, t, k/ are now usually referred to as voiced and voiceless stops respectively, although it is recognized that membership in these sets is not entirely determined by whether, according to commonly accepted definitions, a given phonetic element is voiced or voiceless; nor need it even be described as a stop. What is true is that if a phonetic element is phonetically a voiced stop then it will be assigned to the /b, d, g/ set, and if it is a voiceless stop it may, but need not be, assigned to /p, t, k/. A context in which the stop members of the two phonological sets may be distinguished simply on the basis of voicing (as narrowly defined with respect to stop consonants) is between vowels, as for example in the pair *rabid-rapid.* Acoustically, however, as many as 16 pattern properties can be counted that may play a role in determining whether a listener reports hearing one of these words rather than the other. In purely acoustic terms these properties are rather disparate, although most of them show variations that can plausibly be considered to be primarily the diverse effects of a relatively simple difference in the management of the larynx together with the closing and opening of the mouth. This diversity makes it difficult to rationalize a purely acoustic account of the *rabid-rapid* opposition, – i.e., one that makes no reference to the articulatory mechanisms and maneuvers by which the common linguistic effect of varying these acoustic properties might be explained.

Keywords: closure, cues, duration, voicing

INTRODUCTION

If the topic of voicing as a distinctive attribute of speech sounds continues to be a subject of lively interest to students of speech communication, it must be because it continues to provoke new questions or to refuse final answers to old ones. From a strictly phonetic viewpoint it is unclear why the subject of stop voicing should not be considered closed. The acoustic and articulatory bases of the voiced-voiceless difference are fairly well understood, though it is of course true that details of the aerodynamic, physiological, and other aspects of the picture always remain to be clarified. A specified interval of speech signal is readily described as voiced or voiceless on the basis of whether or not it exhibits harmonic patterning that can be attributed to vocal fold vibration. In addition, it is generally agreed that a given phonetic unit is voiced or voiceless depending on whether or not an interval of speech signal with which it is equated is in fact voiced. This

* Preparation of this paper was supported by NICHD Grant HD-01994 to Haskins Laboratories. I want also to express thanks to Arthur S. Abramson and Catherine Browman for helpful criticisms of an earlier draft.

raises the question of selecting the interval over which presence or absence of voicing shall determine whether the phonetic unit is described as voiced or voiceless. For stop consonants the diagnostic interval that linguists usually choose (e.g., the International Phonetic Association) coincides with the interval of articulatory closure. A stop is then "voiced" if the closure is marked by laryngeal buzz, and it is "voiceless" if that interval is devoid of such signal. Aside from the facts that a closure interval may be neither entirely buzzed nor entirely silent, and that auditory judgment and acoustic record may not always agree, it is otherwise not immediately obvious why the subject of stop voicing still draws the amount of attention devoted to it in recent years.

CATALOGUE OF FEATURES

English /b/–/p/ ≠ [b] –[p]

Given the spelling conventions and the definition of stop voicing to which linguists appear generally to subscribe, a phonetic unit represented as [b] is a voiced stop, while [p] stands for its voiceless counterpart. Many languages make contrastive use of stop categories that consistently differ in voicing, e.g., Dutch, Italian, and Hungarian. In these languages phonological sets represented as /b/ and /p/ are regularly [b] and [p], – i.e., they are characterized by voiced and voiceless stoppages of airflow through the vocal tract. But in certain languages, among them English, there are phonological categories, also represented as /b/ and /p/, whose relation to the phonetic categories [b] and [p] is not so straightforward. Since many linguists have long recognized that members of the English "voiced" set are not invariably voiced, – i.e., /b/ may be initially [p] (though more often it is spelled "phonetically" [b̥], with no clear indication of whether the preference for the latter spelling is dictated by phonetic or phonological considerations), and prepausally as well its "voicelessness is marked," i.e., readily detected by ear (Trager and Smith, 1951), it follows that the search for the acoustic properties cueing the /b/–/p/ contrast in English is not necessarily a search for cues to the phonetic feature of stop voicing. What is called the subject of stop voicing in English continues to hold the attention of speech researchers, not because of the problematical nature of the acoustic correlates of a [±voiced] difference, but because the phonological analysis of English yields /p/ and /b/ categories that are phonetically variable in nature and cued by different acoustic properties in different contexts. The "problem" of English stop voicing resides largely in the fact that the observed variability of the /b/–/p/ distinction runs counter to our reasonable expectation that all phonetic elements similarly designated should have some acoustic properties in common.

Medial /b/–/p/ ≈ [b] –[p]

A context in which the contrast between stop members of English /b/ and /p/ seems most nearly to be one involving [b] vs. [p] is medially in words before an unstressed syllable, particularly where the signal preceding and following the closure is voiced. In this context, then, the acoustic features that distinguish the two stop categories can

perhaps be said to serve as cues to the phonetic feature of voicing. This is to say that **if** phoneticians generally agree that, e.g., *rabid* and *rapid* differ in stop voicing alone, then the acoustic properties affecting their identification by listeners can be called cues to stop voicing. As it happens, of the two other features that have traditionally figured in accounts of the English stops, [±aspirated] and [±fortis], there is general agreement that the first of these plays no significant role in differentiating *rabid* and *rapid,* at least in American if not in standard southern British English (Trager and Smith, 1951; Jones, 1956; Bronstein, 1960). As for the second, aside from its controversial nature as a phonetic feature on a par with the others (Lisker, 1963), it appears that linguists are not fully agreed that it applies, – thus, for Trager and Smith (1951) the [p] of *rapid* is fortis, while Heffner (1950) follows Otto Jespersen in describing the American pronunciation of /p/ in words like *rapid* as lenis. Of course, if the durational differences in closure and preclosure intervals between *rabid* and *rapid* are construed as evidence of a [±fortis] distinction, then it must be granted that not all the acoustic cues to the lexical distinction can be, strictly speaking, cues to [±voiced]. Despite these strictures, I find it reasonable to believe that the phonetic basis for the distinction in these trochees is as close to being just a matter of closure voicing as can be found in the language.

Counting the acoustic feature differences

Oddly enough, although in medial position the phonetic difference between English /b/ and /p/ may well be smaller than elsewhere, the number of readily isolated acoustic pattern properties whose variation might be expected to affect the identification of a stimulus as *rabid* or *rapid* is larger. (It far exceeds the six listed in Klatt, 1975, for word-initial but utterance-medial intervocalic position, and is in fact more, by two, than the 14 listed by Edwards, 1981, for the same position, where the phonetic basis for the "voicing distinction" is possibly maximal.) However, this fact is remarkable only if we suppose that the number of phonetic features that differentiate the contrasting sets should directly determine the number of properties that we can isolate and manipulate to linguistic effect. Otherwise it is not so very surprising, since utterance–initial stops cannot be cued by properties of the interval preceding closure (except for the pre-speech silence), nor are they in English regularly cued by any property of the closure interval itself. Of some 16 acoustic properties that cue, or can plausibly be supposed to cue, the identification of a form as *rabid* or *rapid,* seven are to be found in the signal preceding the medial closure, three are closure properties, and the remainder are post-closure.

They are the following:

Closure

1) duration of closure
2) duration of glottal signal
3) intensity of glottal signal

Pre-closure

4) duration of vowel
5) duration of first-formant (*F*1) transition

6) $F1$ offset frequency
7) $F1$ transition offset time (i.e., "$F1$ cutback," or, more precisely, "$F1$ cut forward")
8) timing of voice offset
9) fundamental frequency (F_0 contour)
10) decay time of signal

Post-closure

11) release burst intensity
12) timing of voice onset (VOT)
13) onset of $F1$ transition ("$F1$ cutback")
14) $F1$ onset frequency
15) $F1$ transition duration
16) F_0 contour

This list does not fully exhaust the inventory of properties that possibly affect listeners' labeling behavior, for we might imagine that factors contributing to the "prominence" of the second syllable relative to the first (i.e., the stress contour attributed to the form) could have secondary effects on word identification. A pattern labeled *rapid* might, as a result of acoustic alterations effecting a stress shift, be perceived to include /b/ rather than /p/, since a natural token of the derivative of the first word, *rapidity*, calls for the voiceless aspirate [p^h], whereas a [p] would not be incompatible with an interpretation of the pattern as the word *rabidity*. Nor can we in principle exclude the possibility that still other isolatable acoustic properties, e.g., higher formants, may make contributions to lexical identity, even though such effects might not be readily explained (Lisker, 1975).

In the inventory just listed 16 acoustic properties were enumerated, and several more suggested, but the precise number cannot be taken very seriously, since with respect to some of them it is difficult to decide whether we have one property or more. And while we may decide that we have more than one, at least for purposes of experimentation, they may not be acoustically distinct, to say nothing of whether or not they are subject to independent control by the operator of the human vocal tract. Thus, for example, property no. 12 might be analyzed as two properties, voice-onset time and aspiration (following Klatt, 1975), since a delay in voice onset can be accompanied by a silent interval (per ejective articulation) or by aspiration. On the other hand, items no. 2 and no. 8 are counted as two rather than one, not on an acoustic basis, but only because of a prior segmentation of the speech patterns whereby the test stimuli were partitioned into pre-closure, closure and post-closure intervals. (A similar segmentation underlies the common distinction drawn between the phonetic features of stop voicing and voiceless aspiration in English and some other languages, and their subsequent treatment as independent properties of stop consonants.)

Acoustic properties as context-variable lexical cues

Of the above-listed acoustic properties that might affect the identification of a signal as *rabid* or *rapid,* it is probably true that none is indispensible, while it is possible that several play no significant role in the perception of unedited naturally produced tokens of these words. Thus a reported *rabid* need not mean that the medial closure was voiced (Lisker, 1957), while a long closure duration does not invariably elicit a *rapid* labeling response (Lisker, 1981). At present we may only say that some of the properties demonstrably affect word perception under certain conditions, and that the rest of them are "candidate cues," inasmuch as none has so far been shown to make no contribution to the perception of medial /b/ vs. /p/. To be sure, it cannot in principle be proven that any conceivable acoustic property of a speech or speechlike signal is incapable of affecting the perception of an acoustic signal as a particular linguistic message; on the other hand, we have no right to assume a principle of "once a cue, always a cue." Thus, for example, the linguistic irrelevance of the [±voiced] differences in the case of initial /b/ does not mean that the identification of a stop as /b/ is everywhere unaffected by whether its closure is voiced or voiceless (although it is just this non sequitur that underlies the assertion by Jakobson and Halle, 1956, that the "distinctive feature" distinguishing the category sets /b, d, g/ and /p, t, k/ is one of articulatory force and not voicing). The aim of most research into the processes of speech perception has been to uncover all the acoustic properties that can **somewhere** serve as cues, and not so much to specify the conditions under which any one of them does and does not serve that function, or to assess the likelihood that the conditions under which it is a cue inside the laboratory are met outside it.

A reading of the phonetic literature suggests that the conditions an acoustic property must satisfy in order to qualify as a "cue" do not involve a demonstrable conformity with nature; it is enough that patterns be devised so that manipulating the property effects a significant shift in listeners' word identification, e.g., from *rabid* to *rapid.* There is no absolute requirement, it would seem, that either the constant properties of the test stimuli or the range of values assigned the variables be copied from nature. Thus property no. 1 listed above, the duration of the formantless interval corresponding to oral closure, serves as a cue to the *rabid–rapid* contrast only in the absence of glottal signal over most of that interval, and it may be decisive only when varied over values that exceed the range observed in nature. When glottal signal persists over much of the closure interval, varying the closure duration will have no effect on listeners' word-labeling behavior; at most, some tokens of the reported *rabid* may strike the listeners as having an abnormally long /b/. Nor is the absence of closure voicing enough to ensure that varying closure duration will affect word identification; e.g., a pattern synthesized with very low values of $F1$ offset and onset frequencies (properties no. 6 and no. 14) is likely to be reported as *rabid* no matter how long the closure (cf. Repp, 1978). Under some conditions, then, closure duration operates as a cue to stop "voicing," i.e., lexical identity; otherwise it is a temporal property that is evaluated temporally. It is likely that every other one of the 16 properties enumerated is also of restricted usefulness as a cue to the listener in deciding on the interpretation of a signal as one or the other word.

Acoustic properties as [±voiced] cues

The acoustic properties that can serve as cues to listeners in deciding whether an auditory stimulus is an instance of *rabid* or *rapid* are said to be cues to the /b/–/p/ contrast in medial position because the accepted phonological representations of these forms, /ˈræbɪd/ and /ˈræpɪd/, appear to attribute their phonetic distinctiveness to a phonological contrast between the medial stops. Does it follow, then, that they are cues to the voiced–voiceless distinction as precisely defined? A reasonable answer would be that, if the lexical distinction is equivalent perceptually to a /b/–/p/ difference and that in turn is a matter of closure voicing, then the acoustic cues are cues to the [±voiced] feature. However, it is by no means generally agreed that phonological representations do in themselves amount to claims about the perceptual nature of a phonetic distinction, and it can be argued that the phonetic spellings [ˈɹæːbɪd] and [ˈɹæpɨd] more directly reflect linguists' judgments about its perceptual basis, i.e., that the lexical decision is based on a combined difference of vowel duration and stop closure signal. Moreover, even if we choose to view the *rabid–rapid* distinction as equivalent to a difference in their stop consonants, it can be argued that in order to be counted as a cue to the voicing status of the stop it is not enough that a given acoustic property should significantly determine a listener's lexical decision; it must affect a decision as to whether or not the medial closure was or was not accompanied by laryngeal buzz, i.e., voicing. Thus, for example, a variation in the duration of the [æ] might possibly affect the lexical decision and thus which stop was reported, but need not determine the answer to a question about stop voicing, which involves a judgment that is both auditory and phonetic. It seems quite possible that, of the 16 or more acoustic properties that may help determine the lexical decision, only the three closure characteristics are directly cues to the perceived voicing state of the closure, while the others are cues to that state only in a derivative sense. If the properties of the pre-closure interval are set at values compatible with a [+voiced] closure, they may induce listeners to report hearing a /b/, i.e., the word *rabid,* but it cannot be presumed that they will also lead them to report hearing a voiced closure. They might indeed more consistently report that a stimulus pair, one labeled *rabid* and the other *rapid,* differ in their [æ] durations than in the [±voiced] nature of the medial stop closures. In such a case it would hardly seem appropriate to call the duration of the vowel a cue to the voicing of the stop. (The situation would be analogous to the celebrated cases of *rider–writer* and *ladder–latter* in varieties of American English.)

Acoustic cues → articulatory gesture

If many of the acoustic properties listed above can be considered the consequences of a laryngeal gesture (Lisker and Abramson, 1971; Abramson, 1977; Goldstein and Browman, in preparation) executed in conjunction with labial closure and opening, we may reasonably decide that a speech signal is more simply described as an ensemble of articulatory rather than acoustic events. A signal identified as *rapid,* which can differ acoustically in many ways from one heard as *rabid,* may be said to differ essentially from the latter in that vocal fold vibration is halted for much of the interval of labial closure. Thus, for example, the fact that two pairs of acoustic patterns, one differing

only in closure duration and the other only in release burst intensity, are both interpreted as *rabid* vs. *rapid,* may be explained by the claim that both differences are consequences of a single difference in laryngeal activity. This many–one relation of the acoustic and articulatory differences between *rabid* and *rapid* can be understood to support the view that speech is better described in articulatory than in acoustic terms, – i.e., that the "sounds of speech" as represented in a linguist's phonetic and phonological spellings are connected more directly with articulatory gestures and states than with acoustic properties. This is not to say that charting the connections between articulation and the phonetic features of speech is a trivial matter, only that it is easier than establishing those that relate the latter to the acoustic signal. Interesting evidence recently reported by Flege (1982) shows that the two kinds of English /b/ found initially are frequently produced with glottal closing gestures having the same temporal relation to the supra-glottal articulation, and thus there is an articulatory invariant underlying the allophonic [±voiced] difference.

Articulatory gestures → acoustic cue

Even if it is accepted that speech perception is special in that it involves awareness, not of the acoustic properties, but rather the articulatory gestures that the listener infers from them (Liberman and Mattingly, 1985), it does not follow that phonetic explanation never goes the other way, i.e., that it never seeks to explain articulatory diversity by pointing to a single acoustic consequence. The matter of consonantal voicing provides what appears to be a compelling case, where articulatory gestures of various kinds have been explained as maneuvers all "designed" to produce either voiced or voiceless closures. Thus the longer [æ], as well as the lowered larynx, the raised velum and the generally "laxed" articulation associated with /b/ as against /p/, have all been considered to facilitate the **acoustic feature** of voicing during closure (Bell-Berti, 1975; Halle and Stevens, 1967; Kent and Moll, 1969; Riordan, 1980; Westbury, 1983). Moreover, it does not appear that there is a single laryngeal devoicing gesture for /p/, since the same acoustically silent closure is produced either by abducting the vocal folds or by halting their vibratory movement without very much glottal opening, and indeed, in British English, with glottalization or "glottal reinforcement" (Roach, 1983). In the cases of both the voiced and the voiceless closures, then, it might be argued that the articulatory gestures are many, the "intended" acoustic outcome one.

SUMMARY

The number of acoustic properties that can be manipulated so as to affect the listeners' decision in judging an auditory stimulus as an instance of the English word *rabid* or *rapid* is considerably greater than the number of phonetic features customarily enumerated as the basis on which they are distinguished. At least 16, and quite possibly more, may serve as cues to the lexical distinction. Insofar as the phonetic feature held to be chiefly responsible for the auditory distinctiveness of the two forms is a simple difference in the nature of the signal emitted during the interval of oral closure, to that

extent can the acoustic properties that serve as lexical cues be said to be cues to the contrast between the phonetic categories [b] and [p] and hence, by definition, as cues to the [±voiced] difference. It is reasonable to regard the lexical decision as being equivalent to deciding whether a /b/ or a /p/ was present in the signal, but it is by no means clear whether the lexical decision as between *rabid* and *rapid* is the same as a decision about the acoustic nature of the signal emitted during closure. We may adopt the hypothesis that most of the acoustic properties whose variation affects the *rabid–rapid* decision are the consequences of articulatory maneuvers "designed" either to inhibit or not to inhibit production of voice during the closure interval. If we confine our attention to the larynx as the articulator chiefly responsible for the [±voiced] difference, then those articulatory maneuvers are possibly fewer and more simply described than are the acoustic properties they generate. But if, on the other hand, the nature of the signal emitted during the closure is a major acoustic cue to the lexical distinction (and this seems quite likely so far as naturally produced speech is concerned), and if **all** the articulatory maneuvers said to be associated with voiced versus voiceless stops can be seen as factors determining the [±voiced] feature (i.e., adjustments of glottal area, vocal fold stiffness, larynx height, velar height, and cavity wall tensity), then surely it is the articulatory picture whose relative complexity is to be explained by the acoustic reference. Thus, at least with respect to stop voicing, it does not seem possible to give an adequate account of all the phonetic facts by deciding that, as between its articulatory and acoustic aspects, we can choose one to the exclusion of the other. A purely articulatory account, and a purely acoustical one as well, may appear to gain the simplicity that passes for explanation, but it is at the expense of adequacy.

REFERENCES

ABRAMSON, A.S. (1977). Laryngeal timing in consonant distinctions. *Phonetica*, **34**, 295-303.

BELL-BERTI, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, **57**, 456-461.

BRONSTEIN, A.J. (1960). *The Pronunciation of American English.* New York: Appleton-Century-Crofts.

EDWARDS, T.J. (1981). Multiple features analysis of intervocalic English plosives. *Journal of the Acoustical Society of America*, **69**, 535-547.

FLEGE, J.E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, **10**, 177-192.

GOLDSTEIN, L. and BROWMAN, C.P. (in preparation). Representation of voicing contrasts using articulatory gestures.

HALLE, M. and STEVENS, K.N. (1967). On the mechanism of glottal vibration for vowels and consonants. *Quarterly Progress Report of the Research Laboratory of Electronics, Massachusetts Institute of Technology*, **85**, 267-270.

HEFFNER, R-M.S. (1950). *General Phonetics.* Madison: University of Wisconsin Press.

International Phonetic Association (1949). *The Principles of the International Phonetic Association.* London: Department of Phonetics, University College.

JAKOBSON, R. and HALLE, M. (1956). *Fundamentals of Language.* The Hague: Mouton.

JONES, D. (1956). *An Outline of English Phonetics* (8th ed.). Cambridge: Heffer.

KENT, R.D. and MOLL, K.L. (1969). Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, **46**, 1549-1555.

KLATT, D.H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, **18**, 687-703.

LIBERMAN, A.M. and MATTINGLY, I.G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

LISKER, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, **33**, 42-49.

LISKER, L. (1963). On Hultzén's "voiceless lenis stops in prevocalic clusters." *Word*, **19**, 376-387.

LISKER, L. (1975). Is it VOT or a first-formant transition detector. *Journal of the Acoustical Society of America*, **57**, 1547-1551.

LISKER, L. (1981). On generalizing the *rabid-rapid* distinction based on silent gap duration. *Haskins Laboratories Status Report on Speech Research*, **SR-65**, 251-259.

LISKER, L. and ABRAMSON, A.S. (1971). Distinctive features and laryngeal control. *Language*, **47**, 767-785.

REPP, B.H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, **24**, 471-485.

RIORDAN, C.J. (1980). Larynx height during English stop consonants. *Journal of Phonetics*, **8**, 353-360.

ROACH, P. (1983). *English Phonetics and Phonology.* Cambridge: Cambridge University Press.

TRAGER, G.L. and SMITH, H.L. Jr. (1951). *An Outline of English Structure* (Studies in Linguistics: Occasional Papers, 3). Norman, Oklahoma: Battenburg Press.

WESTBURY, J.R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, **73**, 1322-1336.