On "Explaining" Vowel Duration Variation*

Leigh Lisker+
Haskins Laboratories, New Haven, Conn.

As they go about investigating physical aspects of speech communication, phoneticians, like other linguists, are most interested in trying to identify those properties that serve a distinctive function. If a property is determined to do so—to be, in other words, "linguistically relevant"—then this finding in itself constitutes the explanation for its presence or absence in any particular piece of speech behavior. But the interest in such properties is, for the phonetician, only one side of the coin; the other side comprises those properties that have little or no apparent cue value for the linguistic identification of an utterance, but that nevertheless display regularities of occurrence that prevent our dismissing them out of hand as simply "noise in the channel." In the case of these latter properties the phonetician is also interested in devising explanations, and here explanations are in a sense much more interesting: whereas the distinctive property is explained by its linguistic function, the linguistically irrelevant property must be explained away, and such an enterprise demands a more strenuous exercise of ingenuity. Most often the explanation offered appeals to mechanical or other physiological constraints on the human organism. Temporal phenomena have frequently been the object of this kind of attention, and among these, certain regularities of vowel duration have been an especially favored topic.

Studies of vowel duration have resulted in two well-known and generally accepted formulations: 1) that the duration of the acoustic segment associated with a vowel depends to a significant extent on the degree of opening of the vowel, and 2) that the duration depends also on the nature of a following consonant. For these relations several explanations have been advanced and apparently accepted, all of them reasonable, but all with certain weaknesses when considered within a more general phonetic framework. A consideration of these explanations, both as to their presuppositions and their implications over and above the particular phenomena they were designed to explain, suggests strongly that, ad hoc to begin with, they have yet to be subjected to the critical testing that must precede their inclusion in any well-integrated theory of speech production. Moreover, the fact that the underlying measurement data derive uniformly from speech samples of a narrowly restricted kind, while it does not relieve us of the duty of trying to explain them, does at the same time raise a question about their precise implication for more spontaneous speech behavior.


†Also University of Pennsylvania, Philadelphia.

Data supporting the statement that for English (and let us restrict ourselves to that) vowel duration is related directly to degree of opening have been reported by many workers—House and Fairbanks (1953), Peterson and Lehiste (1960), and Sharf (1962), to name a few.\(^1\) The relation reported has been understood as a mechanical effect due to a temporal constraint on the movement of the relatively large mass of the lower jaw, with that of the tongue sometimes also implicated: if open or low vowels involve more jaw movement than do the close vowels, then the greater so-called "intrinsic duration" of the former is a natural consequence, provided we believe that in speech we regularly operate close to the limits set by the physical constraints on the mechanism. Lehiste, in her 1970 review of the literature on vowel duration, says very bluntly that "The greater length of low vowels is due to the greater extent of the articulatory movements involved in their production" (p. 19). If we can take the frequency of the first formant as a reasonably good acoustic index of vowel opening, we can see just how closely duration and opening are related.

In Figure 1 the topmost line represents mean vowel durations reported by Peterson and Lehiste (1960) as functions of representative values of first-formant frequencies from Peterson and Barney (1952). If the short and long vowels are taken separately, we can, I think, see a tendency for duration to increase with increasing first-formant frequency; at least [I] and [u] are shorter on the average than [ʌ], and [i] and [u] are likewise shorter than [o] and [æ]. The picture is spoiled a bit by [a], which is no longer than [u], but other studies, that of Sharf's (1962) for example, show [a] longer than [u]. Note that the relation of [æ] to [a] is consistent with Perkell's (1969) X-ray finding that although the tongue is higher for [æ] than for [a] the mandible is lower for the former.

The data represented in the three lower lines of Figure 1 are more troublesome. If we are to suppose that the low vowels are longer simply because the mandible is moving as fast as it can, but that it can't manage to cover the required distance in as short a time for those vowels as it does for the high vowels, then we should expect the low vowels to have longer glides and shorter steady-state intervals than the high vowels. But the data from Lehiste and Peterson (1961) show instead the absence of any systematic difference in glide durations for low as against high vowels, and show quite clearly that the greater overall duration of the low vowels, or at any rate of [o] and [a], is primarily a greater duration of their steady-state intervals. In fact the average onglide duration for the vowel [i] is given as 67 msec, while that for [æ] is 73 msec, with [a] only 36 msec. By contrast the steady-state intervals are reported as 120, 132, and 169 msec respectively. It is difficult to see just how a mechanical constraint operates to yield a 90 msec difference in the overall durations of [i] and [æ], while at the same time the latter vowel is on the average produced with a steady-state interval lasting as long as 132 msec. Probing a bit further we find that the vowel for which the sum of the on- and offglide durations is least is [a], while it is greatest for [æ]. These two vowels can hardly be opposed along the vowel height dimension. But the failure to find high vowels with

---

\(^1\) This paper was prepared before the publication of Dennis Klatt's short paper in the Journal of the Acoustical Society of America (54, 1102-1104, 1973), "Interaction between two factors that influence vowel duration," in which data that agree substantially with earlier studies are reported.
VOWEL DURATION vs FIRST-FORMANT FREQUENCY

DURATIONS
- Overall
- Steady-state
- Onglide
- Offglide

First-formant frequencies from Lehiste-Peterson 1961.

FIGURE 1
regularly shorter glides than low vowels is consistent with findings reported by
a number of investigators, most recently by Sussman, MacNeilage, and Hanson
(1973), that show the velocity of jaw movement to vary directly with its total
displacement in vowel-stop and stop-vowel sequences.

This finding, taken together with the capacity of the system for "target
undershoot" and articulatory compensation that Lindblom (1967) has talked about,
provides a basis for expecting little or no systematic difference in the dura-
tions of high and low vowels. If there is in fact such a difference, it remains
nevertheless difficult to maintain the necessity for this difference because of a
mechanical constraint on tongue and jaw velocities, given the long steady-state
intervals reported by Lehiste and Peterson (1961). Without the Lehiste-Peterson
analysis of vowels into glide and steady-state subsegments the explanation of
duration variation would be unquestioned; with their analysis, and with the
studies on articulatory velocities, it seems to me untenable. One might perhaps
better assert that the low vowels, produced with more movement of jaw and per-
haps tongue as well, require for perception of the intended vowel that the form-
ant pattern be maintained in a target region over a longer interval "because of"
the more extensive formant shifts during the onset and offset glide intervals.
But it is dangerous to talk too early of the necessity for any feature. Perhaps
it is not a matter of perceptual need at all, but only that since a stronger ac-
tion is required for the low vowels it is not totally unexpected, on the basis of
other phonetic facts, to find that the articulators are maintained longer on tar-
get.

The notion that the longer vowels are necessarily longer due to the mechani-
cal inertia of jaw or jaw and tongue runs into other difficulties when one con-
siders vowel duration variation ascribable to differences of context. Figure 2
shows data from Sharf (1962) that are very much in agreement with those of
Lehiste and Peterson (1961). (Observe that [a] here is one of the longer vowels.)
What is remarkable is how the durational relation among the various vowels is
maintained, whether the vowels are in one- or two-syllable words, and whether
they are followed by voiced or voiceless stops. The magnitude of these context
effects is at least as great as that ascribed to degree of opening. Thus, for
example, if it is maintained that [æ] is longer than [I] because of the inertia
of the mandible, this constraint is nevertheless suspended if to the monosyllabic
word containing [æ] a second syllable is added, or if instead of a following
voiced stop there is a voiceless one.

On the general subject of duration in relation to number of syllables there
has not, to my knowledge, been any attempt to fashion an explanation, and I will
pass on to the relation between vowel duration and the nature of the following
consonant, for which the literature provides no fewer than four explanations.
These are:

1) Vowels are shorter before voiceless consonants because those conso-
nants are fortis, and fortisness involves the earlier onset of art-
iculatory closure.

2) Vowels before voiceless consonants are shorter because the strong
closure gesture is accomplished more rapidly, again because of the
fortis nature of those consonants.
VOWEL DURATION BEFORE VOICED AND VOICELESS STOPS

MONOSYLLABLES

Voiced Stops
Mean Values
Voiceless Stops

DISSYLLABLES

Voiced Stops
Mean Values
Voiceless Stops

Duration in msecs

First-Formant Frequency in Hz

Vowel durations from Sharf 1962
First-formant frequencies from Peterson-Barney 1952

Figure 2
3) Vowels are lengthened before voiced stops to allow time for laryngeal readjustment needed if voicing is to be maintained during oral closure.

4) Vowels are longer before voiced and shorter before voiceless consonants according to a rule of constant energy expenditure for the syllable, longer vowels and voiceless consonants both being more costly in articulatory energy.

Of these explanations the last is flawed by the absence of an agreed-upon measure of overall articulatory energy, as well as any rationale for supposing that constancy of energy expenditure, even if measurable, should characterize only one class of sequences, those consisting of a syllabic nucleus and following consonant or consonant cluster. Explanation 3 has, I suppose, the greatest currency at the moment, mainly because of The Sound Pattern of English (1968). As a serious explanation of the relation between vowel duration and following stop it suffers on several counts. First, the available electromyographic and fiberoptic data provide little indication of laryngeal change before voiced stops, but they do indicate that the arytenoid cartilages are subject to an adjustment in rough synchrony with the closure for voiceless stop production. Moreover, the hypothesis that lengthening is required for the maintenance of glottal pulsing throughout the interval of oral closure should be tested, not by asking how much longer is the vowel before voiced than before voiceless stops, but rather how much longer is the vowel before voiced stops than before nasal consonants. The answer to the second question is that the incremental duration is essentially zero. If it is still insisted that vowel lengthening is required to allow time for glottal readjustment before voiced stop closure, then this would seem to imply that the shorter vowels ought to show a greater increase in duration than the longer vowels. But in fact, from Sharf's (1962) data (Figure 3) it appears that the longer a vowel is preceding voiceless stops, the greater the durational increment added before voiced stops. According to Sharf's data the relation between duration before voiceless stops and the durational increment appears to be linear.

Explanations 1 and 2, both of which postulate a shortening before the voiceless consonants because of their claimed greater force of articulation, find confirmation in electromyographic and velocity measurement data that show voiceless stop closures to begin earlier and to be executed more rapidly than closures for the voiced stops. It is odd, however, that this advancement in the timing of closure can be said to be explained by the fortisness feature. This may reflect the prejudice against giving first place to that feature difference between voiced and voiceless stops for which the evidence is incontrovertible—namely the difference in laryngeal state. The assumption is made, and there is no serious attempt made to justify it, that the articulatory program for a consonant involving oral occlusion is independent of whether the stop is voiceless, oral and voiced, or nasalized and voiced. Since the programs for voiced and voiceless stops are clearly not the same when they follow a stressed vowel, this failure must be explained. But the explanation based on the assumed fortisness of the voiceless stops says no more than that the voiceless stops are produced with an earlier and more rapid closure than the voiced ones. The concomitant laryngeal change, abduction of the arytenoids, is tacitly taken to be secondary to the supraglottal event. However, one might just as reasonably suppose that the laryngeal gesture determines the timing of the closure, and that there is no ground for assuming a priori that the onset of arytenoid abduction should follow a temporal program identical with the one that determines the time of closure for the
THE "EFFECT" OF STOP VOICING ON VOWEL DURATION

Duration Added before Voiced Stops

Duration of Vowels Preceding Voiceless Stops

(Monosyllables)

(Vowels from Sharf 1962)
voiced stop. However, there is a reason that can be advanced as to why the closure for the voiceless stop should occur not long after the devoicing gesture begins: that is, that the phonetic result would otherwise be not a sequence of vowel + voiceless stop, but rather, vowel + aspiration + voiceless stop—a phonetic output unacceptable as normal English. Why such an output is unacceptable is a question that the phonetician is not in a position to answer.

REFERENCES


