glossa

AN INTERNATIONAL JOURNAL OF LINGUISTICS

Published twice a year, in April and October, for the GLOSSA SOCIETY, Simon Fraser University.

ANNUAL SUBSCRIPTION RATES
Institutional Subscriptions: $8.00 Canadian
Personal Subscriptions: $7.00 Canadian
All orders placed through an agency: $8.00 Canadian

All rates include postage.
Orders should be addressed to the Business Manager.

MANUSCRIPTS

Manuscripts (articles, discussions, notes, reviews) may be written in any language using Latin or Cyrillic alphabets. They should be addressed to the Editorial Board.

All correspondence should be sent to:
GLOSSA
Department of Modern Languages
Simon Fraser University
Burnaby, B.C.
Canada
V5A 1S6

Lithographed in Canada by Price Printing

LEIGH LISKER, Haskins Laboratories and
University of Pennsylvania

ON "EXPLAINING" VOWEL DURATION VARIATION*

Studies of vowel duration have yielded two well-known and generally accepted formulations: 1) vowel duration varies directly with degree of opening; 2) vowel duration depends on the following consonant, most strikingly on its voicing state. The first relation is explained as a mechanical effect of the relatively large mass of the mandible; the second is ascribed either to the alleged "fortisness" of the voiceless consonants or to a presumed need to defer consonant closure while the larynx is adjusted for consonant voicing. All the explanations offered for the two relations are plausible, but none fails to present certain difficulties when considered within a more general phonetic framework.

As they go about investigating physical aspects of speech communication, phoneticians are most interested, like other linguists, 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 which nevertheless display regularities of occurrence that preclude 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 unmotivated property must be explained away, and this enterprise demands a more strenous 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.

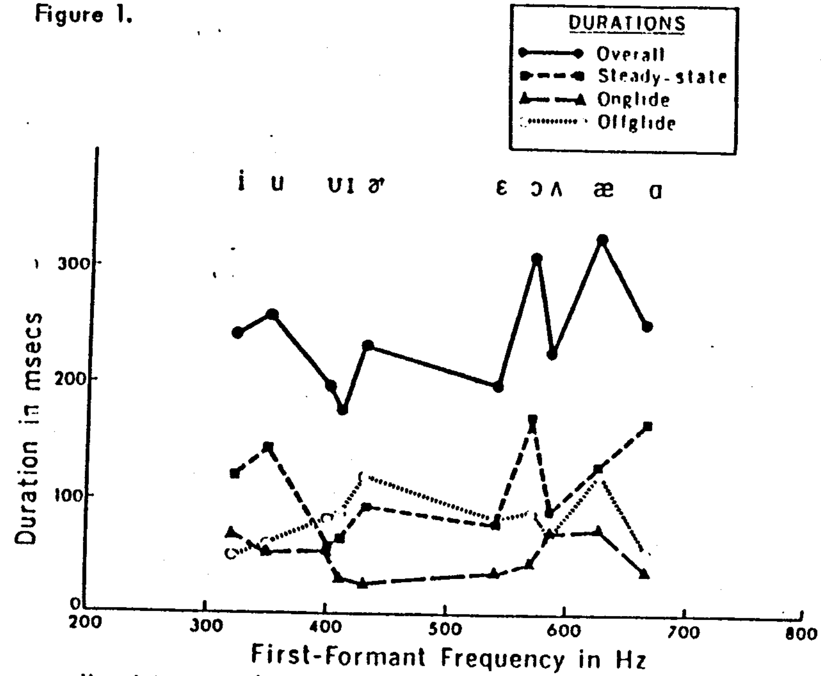
Studies of vowel duration have resulted in two well-known and generally accepted formulations: 1) that the duration of the acoustic segment associated with a vowel depends significantly on the degree of opening of the vowel, and 2) that the duration depends also on certain properties of a following consonant. For these relations several explanations have been offered and apparently accepted, all of them reasonable, but none without 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 as to their precise meaning for more spontaneous speech behavior.

Data supporting the statement that for English (and we shall here restrict ourselves to that language[1]) vowel duration is related directly to degree of mouth opening have been reported by many workers, – House and Fairbanks (1953), Peterson and Lehiste (1960), House (1961), and Sharf (1962), to name a few.[2] 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.

VOWEL DURATION vs FIRST-FORMANT FREQUENCY

Figure 1.



Vowel durations (mean of 5 talkers) from Peterson-Lehiste 1960, 1961
First-formant frequencies from Lehiste-Peterson 1961

In Figure 1 the topmost curve connects points representing mean vowel durations reported by Peterson and Lehiste (1960) as functions of representative first-formant values provided by the same authors in a subsequent paper (Lehiste and Peterson, 1961). If the vowels traditionally called "short" and "long" (or "lax" and "tense") are taken separately, we can see a tendency for duration to increase with increasing first-formant frequency; at least [ɪ] and [ʊ] are shorter on the average than [ʌ][3], and [i] and [u] are likewise shorter than either [ɔ] or [æ]. The picture is spoiled a bit by [ɑ], which according to these data is no longer than [u], but there are other studies, that of Sharf (1962) for example, which show [ɑ] to be longer than [u].[4] It might be remarked here that the relation of [ɑ:] to [ɑ] is consistent with Perkell's (1969) X-ray finding that although the tongue is higher for [æ] than for [ɑ] the mandible is lower for the former.

The data represented by the three lower curves of Fig. 1 are somewhat more troublesome than the one representing overall duration. If we suppose that the low vowels are longer simply because the mandible is moving as fast as it can, but that it cannot move 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 be produced with longer glides and shorter steadystate intervals than the high vowels. But the Lehiste-Peterson data of 1961 show instead an absence of any systematic difference in glide durations for low as against high vowels. Moreover they show quite clearly that the greater overall duration of the low vowels, or at any rate of [ɔ] and [ɑ], is primarily a greater duration of their steadystate intervals. In fact, the mean onglide duration for the vowel [i] is about 65 msec, while for [æ] it is 75 msec; for [ɑ], quite inexplicably, the onglide duration is only 35 msec![5] By contrast, the steadystate intervals appear to be 120, 130 and 170 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] (240 and 330 msec respectively), for example, while at the same time the latter vowel is on the average produced with a steadystate interval lasting as long as 130 msec. Probing the Peterson-Lehiste data a bit more we find that the vowel for which the sum of the on- and offglide durations is *least* is [ɑ], while it is greatest for [æ]. These two vowels can hardly be said to differ radically with respect to either tongue height or degree of mandibular lowering. However the failure to find high vowels with regularly shorter glides than low vowels is not inconsistent with findings such as those of Sussman, MacNeilage and Hanson (1973), which 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 (1963, 1967) has talked about,[6] provides some basis for expecting little or no systematic difference in the durations of high and low vowels. If there *is* in fact such a difference, it remains difficult to maintain that this difference is a *necessary* one because of a mechanical constraint on tongue and jaw velocities, given the long steadystate intervals reported by Lehiste and Peterson (1961). Without the Lehiste-Peterson analysis of vowels into glide and steadystate subsegments the explanation of duration variation might be unquestioned; with their analysis, and with the studies of articulatory velocities, it seems to me untenable. One might at least as plausibly assert that the low vowels, produced with more movement of jaw and perhaps tongue as well, involve a greater expenditure of "articulatory energy", and if the often stated belief that this results in a greater segment duration is assumed to be true,[7] then the greater steadystate duration of the low vowels is a natural consequence. One might even hazard the guess, though there is little to support it, that there exists a perceptual requirement that for the low vowels the formant pattern be maintained in a target region over a longer interval because of the more

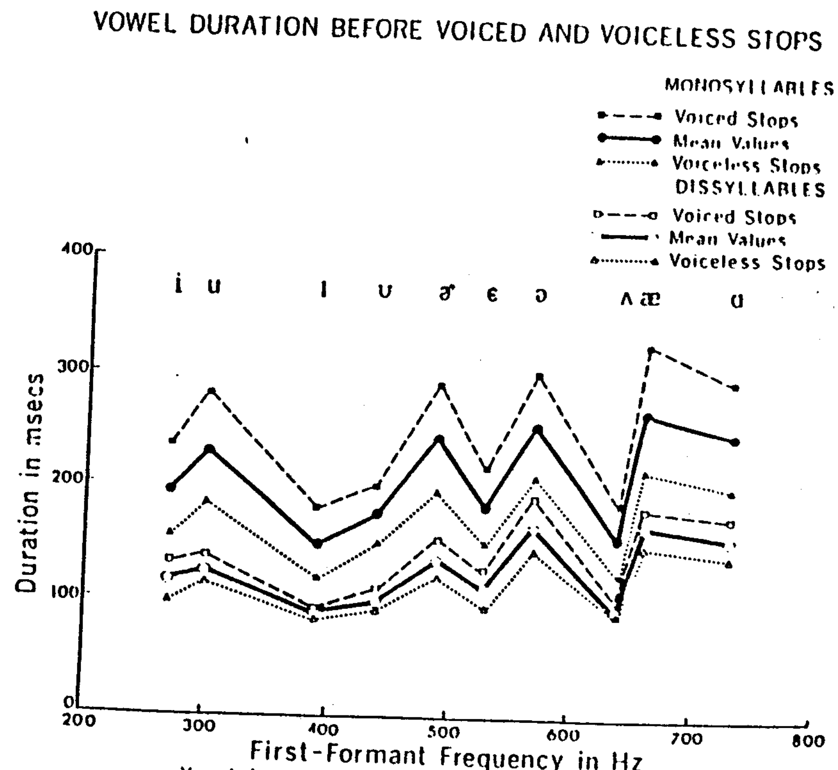


Figure 2. Vowel durations from Sharf 1962
First-formant frequencies from Peterson-Barney 1952

extensive formant shifts during the onset and offset intervals. But it is dangerous to talk too soon of *necessity* in connection with any feature of the speech signal because of alleged limitations on either articulation or perception; so far all we have is evidence that some sort of relation between duration and vowel height *exists*, none that this relation *must be*.

The notion that the longer vowels are necessarily longer, owing to the mechanical inertia of jaw or jaw and tongue, runs into other difficulties when we consider vowel duration variation ascribable to differences of context. Figure 2 shows data from Sharf (1962) that are very much in agreement with those of Peterson and Lehiste (1960). (Observe that |ʌ| is here one of the longer vowels.) What is remarkable is how the durational relations among the various vowels are maintained, whether the vowels are in monosyllabic or dissyllabic words, and whether they are followed by voiced or voiceless stops. The magnitude of these context effects is at least as great as that attributed to degree of oral 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, or at least seriously modified, 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 vowel duration in relation to number of syllables there is a certain amount of literature,[8] but there has not, to my knowledge, been any attempt to fashion an explanation, and we 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.[9] These are:

1) 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.[10]

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

3) Vowels are shorter before voiceless consonants because those consonants are fortes, and fortisness involves the earlier onset of articulatory closure.

4) Vowels before voiceless consonants are shorter because the strong closure gesture is accomplished more rapidly, again because of the fortis nature of those consonants.

Of these explanations the first 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 defined so as to be measurable, should characterize only one class of sequences, those consisting of a syllabic nucleus and a following consonant or consonant cluster. The second explanation has, I suppose, the greatest currency at the moment, mainly because it is advanced in The Sound Pattern of English (Chomsky and Halle, 1968, p. 301). As a serious explanation of the relation between vowel duration and following consonant it suffers on several counts. First of all, 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 adjustment in rough synchrony with the closure for voiceless stop production.[11] 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 the vowel is before voiced than before voiceless stops, but rather how much longer it is before voiced stops than before nasal consonants.[12] On this latter question the literature provides conflicting answers;[13] I should assert that the most likely one is that the incremental duration as one goes from nasal consonant to voiced stop context 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 ones. In fact, if we look at data such as Sharf's (1962) it appears that the longer a vowel is preceding voiceless stops the greater the duration added before voiced stops. Sharf's data suggest that the relation between duration before voiceless stops and the durational increment is very nearly linear.[14]

The argument that vowel lengthening is required for stop voicing, even if accepted, ought to raise another kind of question, namely this: what is the warrant for assuming that the maintenance of glottal pulsing during occlusion is at all essential to the production of normal intelligible English /b,

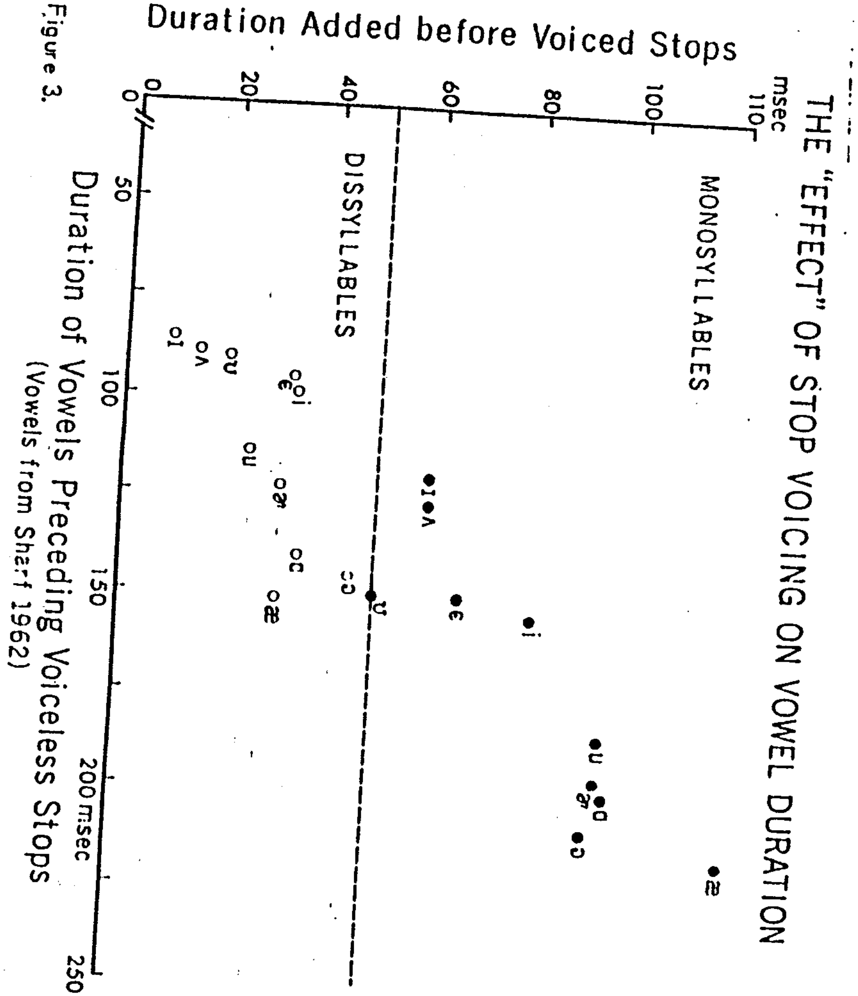


Figure 3.

d,g/? There is in fact a good deal of evidence to indicate that it is not at all necessary. Nor has it ever been shown, as far as I know, that glottal pulsing is most "successfully" maintained during stop closure in precisely those contexts in which vowels show the greatest effect of a following voiced stop.

Explanations 3 and 4, both of which postulate shortening before the voiceless consonants because of their claimed greater force of articulation, find confirmation in electromyographic and velocity measurement data[15] that show voiceless stop closures begin earlier and are executed more rapidly than are 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 when at the same time it is one of the properties defining that feature. Since those who appeal to a fortis-lenis distinction typically insist on its independent and distinctive status in English, it would appear as though the earlier and more rapid closure for the voiceless stops is claimed at one and the same time to be a physiologically determined effect (i.e. a property to be explained away) and a linguistically motivated, hence learned and deliberately imposed,aspect of the articulatory maneuver.

Explanations of vowel duration differences before voiced and voiceless consonants that invoke the fortis-lenis dimension may reflect a general prejudice against giving first place to that phonetic difference between the two classes of consonants for which the evidence is incontrovertible,[16] namely the difference in laryngeal state. The assumption is made, and there is no serious attempt to justify it, that the articulatory program for a consonant involving oral occlusion is independent of whether the consonant 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 fact must be explained. But the explanation based on the asserted fortisness of the voiceless stops says no more than that these consonants are produced with an earlier and more rapid closure than the voiced stops, i.e. they are more forcefully articulated. The concomitant laryngeal change, abduction of the arytenoids, is tacitly taken to be secondary to the supraglottal event. However, one might as reasonably suppose that it is the laryngeal gesture that determines the timing of stop 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 articulatory closure for the voiced stop. There is, however, a reason that can be offered as to why the closure for the voiceless stop should occur not long after the devoicing gesture begins, and that is that the phonetic result would otherwise be, not a sequence of vowel + voiceless stop, but rather vowel + aspiration + voiceless stop, i.e. a phonetic sequence unacceptable as normal English. *Why* such a phonetic output is unacceptable to native English speakers is a question that the phonetician is, happily, not expected to answer.

NOTES

*-A somewhat shorter version of this paper was read before the meeting of the Linguistic Society of America held at San Diego, California, in December 1973.

1-Discussion here is limited to data on English, but the explanations offered to account for the relations reported for English can only be valid if they hold true for speech generally.

2-More extensive bibliographies covering work on English and several other languages are provided in Delattre (1962), Chen (1970) and Lehiste (1970).

3-This relation seems to be at variance with another observation regularly reported in the literature, namely that the shorter the vowel the more centralized it is. Thus if [ʌ] is a more central vowel than either [ɪ] or [ʊ] we might expect to find that it is also shorter. An answer to the apparent anomaly might be that the relation between shortening and centralization holds only for vowel tokens which are phonologically the same.

4-The literature on vowel duration provides no guidance as to just how reliably any pair of vowels differing in degree of opening will conform to the rule relating this difference to one in duration. A careful inspection of the literature, even no more than a comparison of Figs. 1 and 2 of this paper, will reveal inconsistencies, – i.e. [ʌ] > [ɛ] according to Peterson-Lehiste (1960), but the relation is reversed in Sharf (1962).

5-The values given in this paragraph were calculated by combining the figures reported in Peterson-Lehiste (1960) with those in Lehiste-Peterson (1961).

6-The notion of "target undershoot", particularly if linked as it is by Lindblom (1967, p. 24) to the inertial properties of the lower jaw, accounts for the body of observations that inspired it in less satisfactory fashion than its general acceptance implies. Peculiarly enough, it is Lindblom himself who, in describing what he calls undershoot effects, notes that "in most cases there is little undershoot in F_{1m} [i.e. first-formant frequency measured where it presumably most closely approximates the posited target value – LL] observable even at short durations of the vowel . . . " (Lindblom 1963, p. 1776). Moreover, Stevens and House (1963) have also noted that the first formant reaches generally higher frequencies in the environment of the voiceless consonants, where we can suppose the vowels to be relatively short. If we can hold to the assumption that a higher first formant means greater vowel opening, then it would seem we must believe that with increasing openness the duration of the vowel increases (i.e. "intrinsic duration"), but with that decreasing duration (due to consonant context) the degree of opening may also increase!

7-This relation is reported in Jakobson, Fant and Halle (1951, p. 36) and in Chomsky and Halle (1968, pp. 324-325). However it seems not to be claimed that greater length implies greater force of articulation, i.e. that low vowels are tenser than high, or that fricatives are more fortis than stops. (See also fn. 10).

8-See the brief review in Lehiste (1970, pp. 40-41).

9-An admirably concise discussion of the explanations then considered is provided in Delattre (1962). A fuller treatment that includes more recent suggestions is in Lehiste (1970, pp. 19-27).

10-Belasco (1953) asserts that "the anticipation of a consonant requiring a 'strong' force of articulation will tend to shorten the preceding vowel since more of the total energy needed to produce the syllable is concentrated in the consonant". Thus, for two vowels differing in duration, the longer will require more energy because the following consonant needs less; at least we may infer that Belasco asserts this to be true for two vowels that are otherwise the same.

11-See Lisker et al (1970). A short comment in Chen (1970, pp. 151-152) is difficult to interpret, but perhaps can be understood to say that no EMG evidence of glottal readjustment for voiced stops was turned up in his study.

12-In the fully public literature only voiced and voiceless *oral* seem ever to have been compared. However, Halle and Stevens (1967) have more properly (to my mind) compared vowel durations before /b/ and /m/ and before /d/ and /n/.

13-Thus Belasco (1953) and Halle and Stevens (1967) report longer vowels before the oral stops, while Delattre (1962) states without comment that vowels are shorter before oral than before nasal stop consonants. House and Fairbanks (1953) present data indicating that the durational relation differs for the different vowels: thus [a] is on the average 32 msec longer before the oral stops, but [i o u] are of about the same duration before the oral and nasal consonants. Lehiste (1970, p.24) reports that "the influence of homorganic nasals differed but little from that of voiced stops".

14-This is not very different from the relation reported in Klatt (1973).

15-Thus Sussman, MacNeilage and Hanson (1973), and Raphael (1970; in press).

16-Though perhaps Chen (1970) would demur (see fn. 11).

REFERENCES

Belasco, S. (1953) Influence of force of articulation consonants on vowel duration. *JAcS* 25.1015-1016.

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonantal environment. *Phonetica* 22. 129-159.

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English* New York, Harper and Row.

Delattre, P. (1962) Some factors of vowel duration and their cross-linguistic validity. *JAcS* 34.1141-1143.

Halle, M. and Stevens, K.N. 1967. On the mechanism of glottal vibration for vowels and consonants. *MIT QPR* No. 85. 267-271.

House, A.S. (1961) On vowel duration in English. *JAcS* 33. 1174-1178.

House, A.S. and Fairbanks, G. (1953) The influence of consonant environment upon the secondary acoustical characteristics of vowels. *JAcS* 25.105-113.

Klatt, D. (1973) Interaction between two factors that influence vowel duration. *JAcS* 54.1102-1104.

Lehiste, I. (1970) Suprasegmentals. Cambridge, MIT Press.

Lehiste, I. and Peterson, G.E. (1961) Transitions, glides, and diphthongs. *JAcS* 33.268-277.

Lindblom, B. (1963) Spectrographic study of vowel reduction. *JAcS* 35.1773-1781.

Lindblom, B.E.F. (1967) Vowel duration and a model of lip mandible coordination. *Royal Institute of Technology, Speech Transmission Laboratory, Stockholm. QPRS* 4/1967.

Lisker, L., Sawaşhima, M., Abramson, A.S. and Cooper, F.S. (1970) Cinegraphic observations of the larynx during voiced and voiceless stops. *Status Report on Speech Research, Haskins Laboratories*. SR 21/22.201-210.

Peterson, G.E. and Lehiste, I. (1960) Duration of syllable nuclei in English. *JAcS* 32.693-703.

Raphael, R. (1970) Electromyographic investigation of the articulatory basis of vowel duration differences before voiced and voiceless consonants. *JAcS* 48.120(A).

Raphael, L. (*in press*) The physiological control of durational differences between vowels preceding voiced and voiceless consonants. *Journal of Phonetics*.

Sharf, D. (1962) Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5.26-30.

Stevens, K.N. and House, A.S. (1963) Perturbation of vowel articulations by consonant context: an acoustical study. *J. Speech and Hearing Res.* 6.111-128.

Sussman, H.M., MacNeilage, P.F. and Hanson, R.J. (1973) Labial and mandibular movement dynamics during the production of bilabial stop consonants: preliminary observations. *J. Speech and Hearing Res.* 16.397-420.

(Received April 1974)