

Some effects of speaking rate on the production of /b/ and /w/

Joanne L. Miller

Department of Psychology, Northeastern University, Boston, Massachusetts 02115

Thomas Baer

Haskins Laboratories, 270 Crown St., New Haven, Connecticut 06510

(Received 4 October 1982; accepted for publication 18 January 1983)

One of the acoustic properties distinguishing the syllable-initial stop consonant /b/ from the semivowel /w/ is the duration of the initial formant transitions; syllables beginning with /b/ have shorter transitions than those beginning with /w/. This experiment investigated the way in which the transition durations of /b/ and /w/ change as a function of speaking rate by examining tokens of /ba/ and /wa/ produced by four male speakers. At any given speaking rate the /wa/ transitions were, on average, longer than the /ba/ transitions, although pooled across rates, the distributions of transition duration for /ba/ and /wa/ were overlapping. In addition, the magnitude of the difference between average /ba/ and /wa/ transition durations increased with decreases in speaking rate. This is because as rate of speech decreased so that syllable duration increased, there was little change in the initial transition duration of /ba/, but a considerable increase in the initial transition duration of /wa/. Given the overall pattern of results, the transition duration that could optimally distinguish /ba/ from /wa/ was not constant, but increased with syllable duration. This is in accord with Miller and Liberman's (1979) finding that when listeners identify /ba/ and /wa/ on the basis of transition duration, they do so in relation to the duration of the syllable.

PACS numbers: 43.70.Bk, 43.70.Dn, 43.70.Ve

INTRODUCTION

It appears from the literature that a primary difference in the acoustic realization of syllable-initial stop consonants and semivowels is the abruptness of onset, with the onset of a stop consonant being more abrupt than the onset of a semivowel (e.g., Dalston, 1975; Fant, 1960; Fischer-Jørgensen, 1954; O'Connor *et al.*, 1957). Moreover, numerous studies have shown that at least one aspect of syllable onset, the duration of the initial formant transitions, is perceptually relevant for this phonetic distinction: Syllables with short initial transitions are perceived as beginning with a stop consonant, whereas those with long transitions are perceived as beginning with a semivowel (e.g., Cooper *et al.*, 1976; Liberman *et al.*, 1956; Schwab *et al.*, 1981).

Recently, Miller and Liberman (1979) have shown that the duration of the initial formant transition sufficient to perceptually differentiate a stop consonant from a semivowel is not absolute. Focusing on the stop consonant /b/ and the semivowel /w/ in the syllables /ba/ and /wa/, they found that as the overall syllable duration was lengthened, an increasingly longer transition was required to perceive /wa/, as opposed to /ba/. They interpreted this effect as an adjustment for changes in rate of speech. That is, listeners were presumably adjusting for those changes in transition duration that arise from alterations in speaking rate by treating transition duration in relation to overall syllable duration when making a phonetic decision. Thus, as overall syllable duration increased, as would occur if rate of speech decreased, the duration of the transition that distinguished /b/ and /w/ increased.

This interpretation of the perceptual findings rests on two assumptions about speech production. The first, which has considerable empirical support, is that a change in

speaking rate does alter the overall duration of the syllable (e.g., Gay, 1978; Kozhevnikov and Chistovich, 1965; Peterson and Lehiste, 1960; Port, 1981). The second is that the alteration in syllable duration is due not only to a change in the duration of the vowel nucleus, but also to a change in the duration of the initial formant transitions. The few available data that bear on this assumption provide support for it. In particular, Gay (1978) has reported that the initial transitions for /b/ (in intervocalic position) are somewhat longer at slower rates of speech. And recently, Soli (1982) has shown that the initial transitions for the semivowel /j/ (in syllable-initial position) lengthen considerably as speech is slowed and the syllable becomes longer. The present study was undertaken to gather additional data relevant to this assumption. Specifically, we examined the manner in which the initial formant transition durations for /ba/ and /wa/ varied as a function of speaking rate.

I. METHODS

A. Subjects

The subjects were four male speakers whose native language was American English, AA, LL, JS, and DZ. None were aware of the specific purposes of the experiment.

B. Data collection

Each speaker was asked to produce numerous tokens of /ba/ and /wa/ across a wide range of speaking rates. The specific procedure for obtaining these tokens was as follows. The speaker was seated in a sound-treated room in front of a microphone. On the table in front of him was a silent flashing light metronome. For a given setting of the metronome, the speaker was instructed to produce the syllable (/ba/ or

/wa/) in time with each successive flash. He was specifically told to vary the rate of the individual syllables in a manner that was commensurate with the overall rate specified by the metronome, but not to exaggerate the syllables at the slower metronome rates. Finally, he was instructed to produce each syllable as a distinct utterance, that is, to separate successive syllables with a voicing break. Six settings of the metronome were used, ranging from 40 to 192 flashes per minute. The settings were presented twice, first in descending and then in ascending order. At each setting, the subject first produced a sequence of /ba/'s and then a sequence of /wa/'s. A total of 860 syllables was recorded for subsequent analysis, 232 for AA, 182 for LL, 231 for JS, and 215 for DZ. For each subject, approximately half of the tokens were /ba/ and half were /wa/. Informal listening tests indicated that all syllables were unambiguously perceived as /ba/ or /wa/, as intended by the speaker.

C. Acoustic measurements

A wideband (300-Hz) spectrogram was made of each syllable. A Voiceprint Laboratories Sound Spectrograph was used to analyze the speech of AA, LL, and DZ, and a Kay Sonograph was used for the speech of JS.

Two measurements were made for each syllable. The first was the duration of the syllable, defined as the period between syllable onset and syllable offset. The syllable onset was defined as the first observable glottal pulse or, in the case of those /ba/'s for which a release burst was evident, the burst onset. The syllable offset was defined as the last observable glottal pulse. The second measurement was the duration of the initial transition. This acoustic segment was defined as the period between syllable onset and the termination of the first-formant (F1) transition, that is, the point at which the transition attained its steady-state value. Two aspects of this definition require explanation. The first has to do with focusing on the first formant, rather than on one of the higher formants. Preliminary inspection of the spectrograms revealed that the observable formant transitions (in particular, transitions of F1 and F2) terminated at approximately the same point in time. It was also apparent that across all subjects and rates, the F1 transition was the easiest to discern. Thus this formant was chosen for measurement purposes. The second issue concerns treating the onset of the syllable as the onset of the transition. For some /ba/'s there was a short burst before any discernible formant transition, and for some /wa/'s (especially at the slower speaking rates) there was a short period of relatively steady-state formant before the transition proper began. However, for most tokens neither the burst (for /ba/) nor the steady-state segment (for /wa/) was clearly evident, and the demarcation between the offset of the burst or steady-state segment and onset of the transition proper for those tokens containing these properties was very difficult. Thus for the purposes of this experiment we elected to treat any burst or initial steady-state as part of what we will call the transition.

The actual measurements were made by a highly trained technician who was not aware of the purposes of the experiment. The procedure was first to draw vertical lines on

each spectrogram at syllable onset and offset, according to the criteria described above. Next, a horizontal line that extended from the onset to the offset of the syllable was drawn through the center of the steady-state portion of F1. Then an oblique line was drawn as a visual best fit through the F1 transition proper, excluding any preceding burst or steady state. The intersection of the horizontal and oblique lines was taken as the point of termination of the F1 transition. The measurements of syllable duration and transition duration were made to the nearest millimeter, which corresponds to 7.35 ms for the tokens analyzed with the Voiceprint spectrograph and 8.17 ms for the tokens analyzed with the Kay spectrograph.

II. RESULTS AND DISCUSSION

Consider first the syllable duration measurements, graphically displayed for the individual speakers in Fig. 1. Each speaker's graph consists of two interleaved histograms, one for all productions of /ba/ and one for all productions of /wa/. Each histogram displays the percentage of syllable tokens that fall within 50-ms syllable duration intervals. It is apparent that our procedure was successful in generating syllables that spanned a wide range of syllable durations, ranging roughly from 100 to 700 ms. It is also evident that the distributions for /ba/ and /wa/ are very similar, although there is a slight tendency for the /wa/ distributions to be displaced toward longer syllable durations, especially for AA and JS.

Next turn to the transition duration measurements, displayed in a similar fashion in Fig. 2. Each histogram shows the percentage of syllable tokens that fall within 10-ms transition duration intervals. Unlike the case of syllable duration, transition duration differs considerably for /ba/ and /wa/. As expected, the /ba/ distributions are centered at shorter transition durations than the /wa/ distributions. Note also that the distributions for /ba/ are typically narrower than those for /wa/, and that for each subject the two distributions are overlapping.

Finally, Fig. 3 shows how transition duration varies with syllable duration. For each subject, for each of the two syllables, the mean transition duration for all tokens within a given 50-ms syllable duration interval is plotted as a function of the midvalue of that interval; error bars indicate plus and minus one standard deviation from the mean. In addition, a linear regression line has been fit to each set of mean data. The pattern of results is similar for all four subjects. For any given syllable duration, the mean transition duration for /wa/ is greater than that for /ba/. However, the magnitude of the difference is not constant, but increases as syllable duration increases. This is due to the fact that transition duration behaves very differently across rates for the two syllables. For /ba/, there is relatively little change in transition duration as the syllable duration increases. Thus, in contrast to Gay's (1978) findings, it appears from our data that the increase in syllable duration for /ba/ is due almost entirely to an increase in the posttransition, relatively steady-state segment of the syllable. For /wa/, however, the increase in syllable duration is also due to a considerable

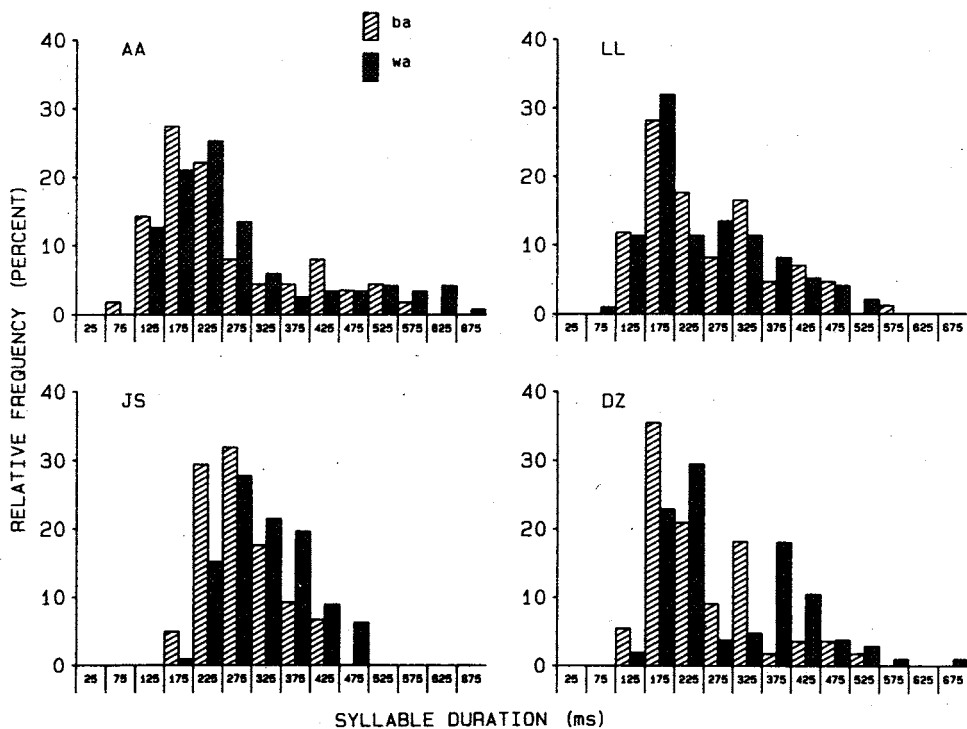


FIG. 1. For each of two syllable types, /ba/ and /wa/, the percentage of syllable tokens with syllable durations within successive 50-ms intervals, for each of four speakers. The numbers along the abscissa represent the midpoints of the intervals.

increase in transition duration, from approximately 40 to 150 ms across the range of syllable durations produced. This increase in transition duration for /w/ is in accord with Soli's (1982) data on the semivowel /j/.¹

The most important issue for our purposes is whether the pattern of results obtained in the present study is consistent with Miller and Liberman's (1979) finding, namely, that

as the syllables from a /ba-wa/ series varying in transition duration were lengthened, the transition duration at the /b-/w/ category boundary increased. Consider the present production data pooled across all four speakers, shown in Fig. 4. It is apparent that a fixed criterion transition duration, which would be represented by a horizontal line on the graph, would not most efficiently differentiate /ba/ from

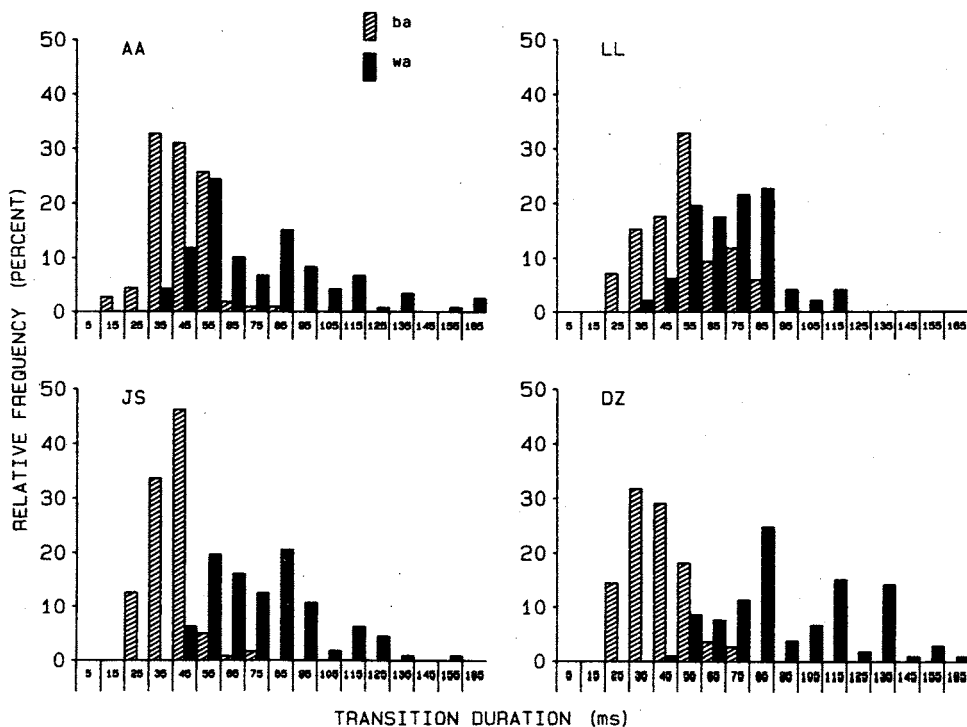


FIG. 2. For each of two syllable types, /ba/ and /wa/, the percentage of syllable tokens with transition durations within successive 10-ms intervals, for each of four speakers. The numbers along the abscissa represent the midpoints of the intervals.

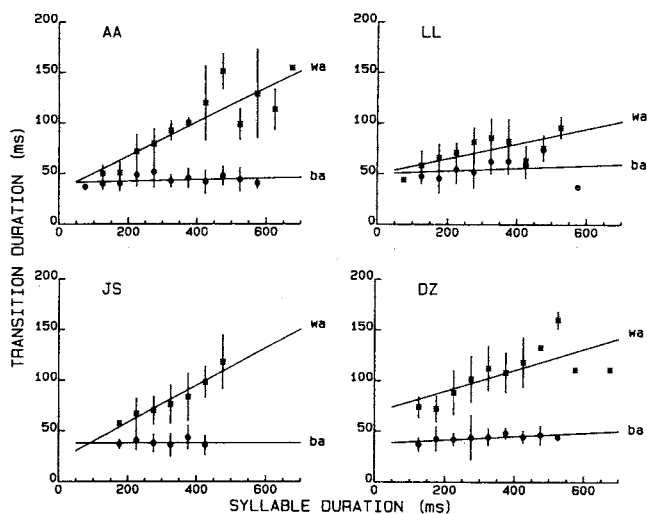


FIG. 3. Mean transition duration in ms, of syllables within successive 50-ms syllable duration intervals, for each of four speakers. The data points are plotted at the midpoint of the 50-ms interval, and the error bars indicate plus and minus one standard deviation from the mean. Linear regression lines are fit to each set of mean data.

/wa/: If the criterion duration were short enough to optimally distinguish */ba/* from */wa/* at fast rates of speech, many */ba/*'s produced at slow rates of speech would be identified erroneously as */wa/*. But if the criterion duration were sufficiently long to adequately differentiate */ba/* from */wa/* at slow rates, many */wa/*'s produced at fast rates would be identified mistakenly as */ba/*. Clearly, fewer misclassifications would occur if the criterion transition duration for differentiating */ba/* from */wa/* were relative rather than fixed, becoming increasingly longer as the speaking rate decreased and the overall syllable duration became longer. Of course, this is precisely what was found by Miller and Liberman.

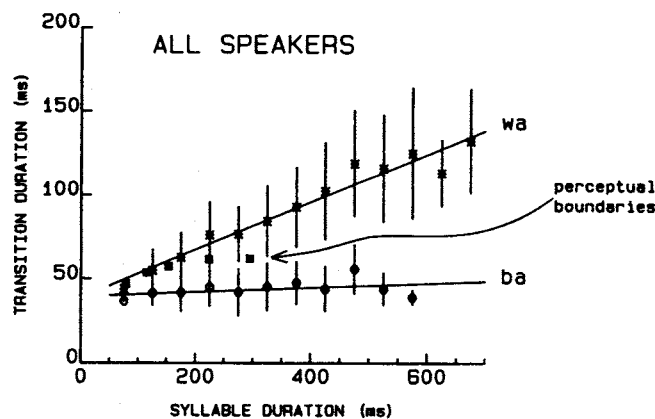


FIG. 4. Mean transition duration in ms, of syllables within successive 50-ms syllable duration intervals for pooled data from four speakers. The data points are plotted at the midpoint of the 50-ms interval, and the error bars indicate plus and minus one standard deviation from the mean. Linear regression lines are fit to each set of mean data. Also plotted on the figure are the perceptual */b/-/w/* category boundaries for syllables with overall durations of 80, 116, 152, 224, and 296 ms, from Miller and Liberman (1979) (see footnote 2).

Thus the overall pattern of the production data is fully in accord with the earlier perceptual data.

It is also possible to assess the match between the actual transition durations measured in the current study at different rates of speech and the */b/-/w/* boundary values reported by Miller and Liberman. The perceptual boundary values they obtained for the five syllable durations tested, 80, 116, 152, 224, and 296 ms, are plotted with the pooled production data on Fig. 4.² The simplest prediction we can make regarding the fit between the two sets of data is that the perceptual boundary will be equidistant from the */b/* and */w/* regression lines. Since linear regression lines were fit to the data, the boundaries should fall on a straight line that bisects the angle between the two regression lines. The obtained perceptual boundaries, especially those for the 80-, 224-, and 296-ms syllables, lie quite close to such a line, although the boundary values appear to be increasing in a negatively accelerated fashion, rather than linearly. One possible reason for the discrepancy is that a curvilinear, and not a linear function, better reflects the variation in */wa/* transition durations across rates, such that the predicted boundary values also increase in a nonlinear manner. Given the variability in the production data, it is difficult to choose between the two types of function, and thus the issue must remain for the present unresolved. Finally, we note that the perceptual boundary value appears to be approaching an asymptote at the longest syllable durations tested. The existence of such a limiting value of transition duration at the perceptual boundary is reasonable vis-à-vis the production data. That is to say, the listener can effectively differentiate */b/* and */w/* if the criterion transition duration becomes longer as the syllable duration increases to about 300 ms, but remains constant thereafter.

It is clear that there is a close correspondence between the way in which alterations in speaking rate affect the transition durations for */b/* and */w/* and the way in which listeners use this acoustic property to distinguish the two phonetic segments across a range of speaking rates. In essence, the perceptual mechanism of the listener operates in such a manner that transition duration can be used to distinguish */b/* and */w/* across speaking rates, even though it does not itself remain constant across rate variation. An issue that immediately arises is the nature of the perceptual mechanism responsible for this type of rate-dependent processing. Given the close relationship between the acoustic consequences of production and perception, it is tempting to propose that the perceptual mechanism is one which is specialized for speech, operating in accord with articulatory principles. However, there are recent data that raise questions about an interpretation based on specialized speech processing. Specifically, Carrell, Pisoni, and Gans (1980) have reported results analogous to those of Miller and Liberman (1979) for nonspeech, sine-wave analogues of */ba/* and */wa/*. On the basis of these findings, they argued that the effect of syllable duration on the perception of */ba/* and */wa/* reflects not a specialized adjustment for articulatory rate *per se* but the operation of the auditory system, which in general treats stimulus onsets in relation to total stimulus duration. One problem with this interpretation, however, is that although it can account for

the similarity in the influence of stimulus duration on the perception of /ba/ and /wa/ and their nonspeech analogs, it does not account for the close correspondence between the perceptual effects for the speech stimuli and the speech production data. Thus the nature of the underlying mechanism remains an open issue.

Finally, it is interesting to note that recent developmental research indicates that young prearticulate infants, like adults, treat transition duration in relation to syllable duration when distinguishing tokens of /ba/ and /wa/ (Eimas and Miller, 1980). Thus the mechanism that allows the listener to accommodate for changes in speaking rate, whether speech-specific or not, appears to be operative in early infancy, prior to the acquisition of language.

ACKNOWLEDGMENTS

This research was supported by NIH Grant NS 14394 and RCDA NS 00661 to J. L. M., Biomedical Research Support Grant RR 07143 to Northeastern University, and NIH Grants HD 01994 and NS 13617 to Haskins Laboratories.

¹We should point out that the increase in transition duration for /w/ was primarily due to an increase in the duration of the transition proper, and not to an increase in the duration of any preceding steady-state segment that was included in our measure of transition duration. Soli's (1982) data on /j/ provide corroboration of this observation, in that his measurements of transition duration were based solely on the transition proper, excluding any preceding steady-state segment.

²Each of the syllables in the /ba-wa/ series used by Miller and Liberman (1979) possessed 16 ms of steady-state first-formant information prior to the onset of the formant transition. This constant 16 ms was not included in their specification of transition duration. Our measure of transition dura-

tion, however, did include any steady-state formant segment at syllable onset. Thus in order to properly compare the perceptual boundaries with the production data we added 16 ms to each of the values reported by Miller and Liberman (1979, Fig. 4).

- Carrell, T. D., Pisoni, D. B., and Gans, S. J. (1980). "Perception of the duration of rapid spectrum changes: Evidence for context effects with speech and nonspeech signals," *J. Acoust. Soc. Am. Suppl. 1* **68**, S49.
- Cooper, W. E., Ebert, R. R., and Cole, R. A. (1976). "Perceptual analysis of stop consonants and glides," *J. Exp. Psychol.: Hum. Percept. Perform.* **2**, 92-104.
- Dalston, R. M. (1975). "Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults," *J. Acoust. Soc. Am.* **57**, 462-469.
- Eimas, P. D., and Miller, J. L. (1980). "Contextual effects in infant speech perception," *Science* **209**, 1140-1141.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).
- Fischer-Jørgensen, E. (1954). "Acoustic analysis of stop consonants," *Misc. Phonet. ii*, 42-59.
- Gay, T. (1978). "Effect of speaking rate on vowel formant transitions," *J. Acoust. Soc. Am.* **63**, 223-230.
- Kozhevnikov, V., and Chistovich, L. (1965). *Speech: Articulation and Perception*, translated by Clearinghouse for Federal Technical and Scientific Information JPRS 30543 (Washington, DC).
- Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956). "Tempo of frequency change as a cue for distinguishing classes of speech sounds," *J. Exp. Psychol.* **52**, 127-137.
- Miller, J. L., and Liberman, A. M. (1979). "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Percept. Psychophys.* **25**, 457-465.
- O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1957). "Acoustic cues for the perception of initial /w,j,r,l/ in English," *Word* **13**, 25-43.
- Peterson, G., and Lehiste, I. (1960). "Duration of syllabic nuclei in English," *J. Acoust. Soc. Am.* **32**, 693-703.
- Port, R. F. (1981). "Linguistic timing factors in combination," *J. Acoust. Soc. Am.* **69**, 262-274.
- Schwab, E. C., Sawusch, J. R., and Nusbaum, H. C. (1981). "The role of second-formant transitions in the stop-semivowel distinction," *Percept. Psychophys.* **29**, 121-128.
- Soli, S. D. (1982). "Structure and duration of vowels together specify fricative voicing," *J. Acoust. Soc. Am.* **72**, 366-378.