



# Effects of fundamental frequency on medial and final [voice] judgments

Wendy A. Castleman and Randy L. Diehl\*

*Department of Psychology, University of Texas, Austin, TX 78712, U.S.A.*

*Received 20th January 1995, and in revised form 12th February 1996*

---

Experiment 1 evaluated the effects of  $f_0$  variation on [voice] judgments in utterance-final position. Analogous to the effects of varying  $F_1$  reported in earlier studies, a lower steady-state  $f_0$  during the preceding vowel and a lower offset  $f_0$  proximal to consonant closure contributed additively to more [+voice] identification responses. The similarity between the effects of  $f_0$  and  $F_1$  variation extends the parallel between the effects of these variables observed in other utterance positions, and suggests that a low  $f_0$  and a low  $F_1$  contribute to a single integrated perceptual correlate of [+voice] consonants. Experiment 2 investigated the effective domain in which  $f_0$  variation influences utterance-medial [voice] judgments for VCV stimuli whose primary [voice] cues are consistent with a stressed-unstressed disyllabic pattern. There was an increase in [+voice] identification responses as a function of a lower steady-state and offset  $f_0$  in the first syllable and as a function of a lower onset  $f_0$  in the second syllable. A comparison of the results with earlier findings suggests that the domain of  $f_0$  influence on [voice] judgments may be determined by the syllable-affiliation of the target consonant.

© 1996 Academic Press Limited

---

## 1. Introduction

Among the world's languages, differences in fundamental frequency ( $f_0$ ) in the vicinity of the consonant are a widely attested correlate of the [voice]<sup>1</sup> distinction, with  $f_0$  values being lower for [+voice] than [–voice] consonants. These differences are reliably observed in the vowel immediately following the consonant (e.g., House & Fairbanks, 1953; Lehiste & Peterson, 1961; Mohr, 1971; Hombert, 1978; Reinholt Petersen, 1983) and have also been reported in the pre-consonantal vowel (Kohler, 1982; Silverman, 1987; but see Mohr, 1971, and Gruenenfelder & Pisoni, 1980, for negative results). Correspondingly, in perceptual studies using synthetic or digitally manipulated natural speech, a lower  $f_0$  near the consonant has been shown to increase [+voice] judgments. Most of these perceptual studies (e.g., Fujimura, 1971;

\* Please address all correspondence to: Randy L. Diehl Department of Psychology, 330 Mezes, University of Texas at Austin, Austin, TX 78712, USA.

<sup>1</sup>The expression “[voice]” is used here as an abbreviation for “[+voice]/[–voice].” The square brackets indicate a phonological feature distinction; without brackets, the term “voicing” refers to vocal fold vibration.

Chistovich, 1969; Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981; Massaro & Cohen, 1976, 1977) have focused on the [voice] distinction in utterance-initial prestressed<sup>2</sup> position. However, several studies (Derr & Massaro, 1980; Gruenenfelder & Pisoni, 1980; Kohler, 1985; Kohler & van Dommelen, 1986) have examined preconsonantal  $f_0$  effects for consonants in medial or final poststressed position.

For the initial prestressed case, there is evidence that the only part of the  $f_0$  contour that influences [voice] judgments among English-speaking listeners is the  $f_0$  value at voicing onset (Massaro & Cohen, 1976; Haggard *et al.*, 1981). Recently, Diehl & Molis (1995) replicated this finding for the medial [voice] distinction when the syllable durations and primary [voice] cues were consistent with a perceptual interpretation of the target consonant as syllable-initial and prestressed (see below).

As to whether the effect of the preconsonantal  $f_0$  is similarly localized for poststressed medial or final consonants, the evidence is somewhat mixed and difficult to interpret. Gruenenfelder & Pisoni (1980) compared a flat  $f_0$  contour to a falling contour with starting and ending values equidistant (in Hz) above and below the flat contour. Despite having the same average  $f_0$  as the flat contour, the falling contour yielded more [+voice] judgments. This suggests that the  $f_0$  values nearest the consonant may have had the predominant effect on perceived [voice] status. Consistent with this interpretation, Kohler (1985) found that a final upward deflection in an otherwise flat preconsonantal  $f_0$  contour increased [-voice] judgments (see also Kohler & van Dommelen, 1986), whereas a final downward deflection increased [+voice] judgments. Moreover, increasing the frequency extent of the downward deflection produced a further sizable increase in [+voice] responses, whereas increasing the duration of the  $f_0$  deflection had an inconsistent effect on labeling performance. In contrast to Kohler's (1985) findings, Derr & Massaro (1980) reported that either an upward or downward  $f_0$  deflection produced more [+voice] judgments than a stationary  $f_0$  contour, suggesting that the property relevant to perception of the [voice] distinction was the presence or absence of change throughout the  $f_0$  contour rather than the  $f_0$  values proximal to the consonant.

The present study examined further the role of  $f_0$  in perception of the [voice] distinction in utterance-final and -medial poststressed position. The study was motivated on the basis of the following considerations. Stevens & Blumstein (1981) claimed that a main, perceptually relevant correlate of the [voice] distinction is the presence ([+voice]) or absence ([-voice]) of low frequency periodic energy in or near the consonant constriction interval. They further claimed that the low frequency property of [+voice] stops may be analyzed into at least three phonetically distinct subproperties—voicing during the consonant constriction interval, a low first formant ( $F_1$ ) frequency near the constriction interval, and a low  $f_0$  in the same vicinity. According to these claims, which we refer to jointly as the “low frequency hypothesis,” a low  $f_0$  near the consonant produces more [+voice] judgments owing to its contribution to a single integrated perceptual quality corresponding to the low frequency property. One prediction of the low frequency hypothesis is that two stimuli in which separate subproperties of the low frequency

<sup>2</sup>The terms “prestressed” and “poststressed” are used in this paper to mean “occurring in initial position of a stressed syllable” and “occurring in final position of a stressed syllable,” respectively.

property are positively correlated (i.e., the subproperties are either both present or both absent) will be more distinguishable than two stimuli in which the subproperties are negatively correlated. This prediction was recently supported for stimulus arrays involving orthogonal variation in either  $f_0$  and voicing duration, or  $F_1$  and voicing duration (Diehl, Castleman, & Kingston, 1995; Kingston & Diehl, 1995). (For additional discussion of the low frequency hypothesis, see Diehl & Kingston, 1991; Kingston & Diehl, 1994.)

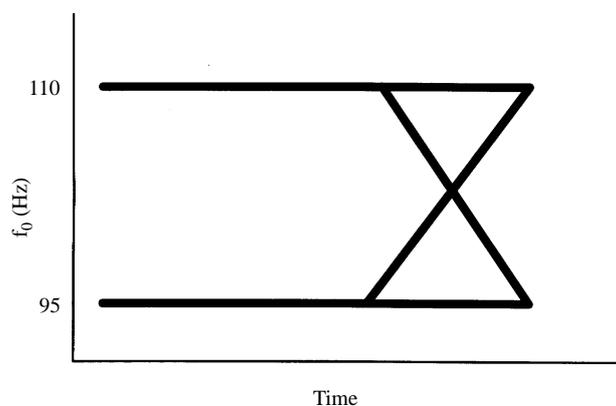
Another prediction of the low frequency hypothesis is that the effects on [voice] judgments of varying either  $f_0$  or  $F_1$  should pattern in similar ways for a given utterance position and stress pattern. There is some evidence consistent with this prediction. Recall that in initial prestressed position, the only part of the  $f_0$  contour that affects English [voice] judgments is the  $f_0$  value at voicing onset (Massaro & Cohen, 1976; Haggard *et al.*, 1981). The effect of  $F_1$  on English [voice] judgments in this utterance position appears to be similarly limited to the region of voicing onset (Lisker, 1975; Summerfield & Haggard, 1977; Kluender, 1991). In the case of utterance-final poststressed consonants, the influence of  $F_1$  on [voice] judgments appears to be less localized. Summers (1988) varied the duration and steady-state  $F_1$  value of the vowel in /bVb/-/bVp/ syllables and found that a lower steady-state  $F_1$  yielded more final [+voice] identification responses. [This perceptual experiment followed an earlier production study (Summers, 1987) showing that, before [+voice] consonants,  $F_1$  is lower throughout most of the vowel.] Castleman, Hughes, & Diehl (unpublished data) replicated and extended Summers's perceptual results. In one experiment, modeled directly after that of Summers (1988), subjects identified synthetic /bab/-/bap/ syllables in which vowel duration and steady-state  $F_1$  value were varied independently. In a second experiment, the vowel duration variable was replaced by an  $F_1$  offset frequency variable. As expected, a longer vowel produced more final [+voice] labeling responses. Of more interest for our present purposes, both a lower steady-state  $F_1$  and a lower offset  $F_1$  yielded significantly more [+voice] responses, and the effects of these two variables were additive. By the low frequency hypothesis, we would predict that comparable effects on [voice] judgments should be observed in utterance-final poststressed position if the preceding  $f_0$  steady-state and offset values are manipulated. This prediction was tested in Experiment 1. In Experiment 2, we examined the effect on [voice] judgments of varying  $f_0$  at several locations within vowel-consonant-vowel (VCV) stimuli. The goal was to determine the perceptually effective domain of  $f_0$  influence for the [voice] distinction in utterance-medial poststressed position.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Stimuli

Four seven-item stimulus series, each ranging perceptually from /bap/ to /bab/, were prepared using the Klatt88 software synthesizer implemented on a DEC VAXstation 3500 computer. The stimuli within a series varied in duration from 180 ms to 360 ms, in 30-ms steps. Each item consisted of a 35-ms interval of linear rising formant transitions (onset frequencies:  $F_1 = 350$  Hz,  $F_2 = 900$  Hz,  $F_3 = 1700$  Hz,  $F_4 = 2500$  Hz), followed by a variable length interval of steady-state



**Figure 1.** Schematic diagram of the  $f_0$  contours used in Experiment 1. The four contours represent all combinations of the binary  $f_0$  values (110 Hz or 95 Hz) in the steady-state region of the syllable and the syllable offset.

formant frequencies ( $F_1 = 750$  Hz,  $F_2 = 1500$ ,<sup>3</sup>  $F_3 = 2500$ ,  $F_4 = 3000$ ), which was in turn followed by linear falling transitions terminating at the values of the onset frequencies. To create convincing final stops, it was found to be advantageous to use  $F_1$  and  $F_4$  transition durations of 20 ms, while the  $F_2$  and  $F_3$  transition durations were 35 ms. Formants were excited by a periodic source throughout the duration of the stimuli. The four series differed according to  $f_0$  contour (see Fig. 1), which began with a steady-state value of either 95 Hz or 110 Hz, and then either maintained that value throughout the stimulus or else changed linearly to the other  $f_0$  value (i.e., 110 Hz or 95 Hz) during the final 70 ms of the stimulus.<sup>4</sup> The series are referred to in terms of their initial and final  $f_0$  values as: *High-High* (110 Hz–110 Hz), *High-Low* (110 Hz–95 Hz), *Low-High* (95 Hz–110 Hz) and *Low-Low* (95 Hz–95 Hz).

### 2.1.2. Subjects and procedure

Twenty-four experimentally naive undergraduate students, enrolled in an introductory psychology course at the University of Texas at Austin, served as subjects. All were native speakers of American English and reported having normal hearing.

Subjects identified 12 randomized blocks of the 28 test stimuli (7 stimulus durations  $\times$  2 steady-state  $f_0$  values  $\times$  2 offset  $f_0$  values) by pressing either of two response keys corresponding to the final consonant /p/ or /b/. They were given up to 2 s to respond, after which another 1 s elapsed before the next stimulus token was presented. The stimuli, stored on a PC, were output at a 10 kHz sampling rate via a 16-bit D/A converter, low-pass filtered at a 4.9 kHz cut-off frequency, and

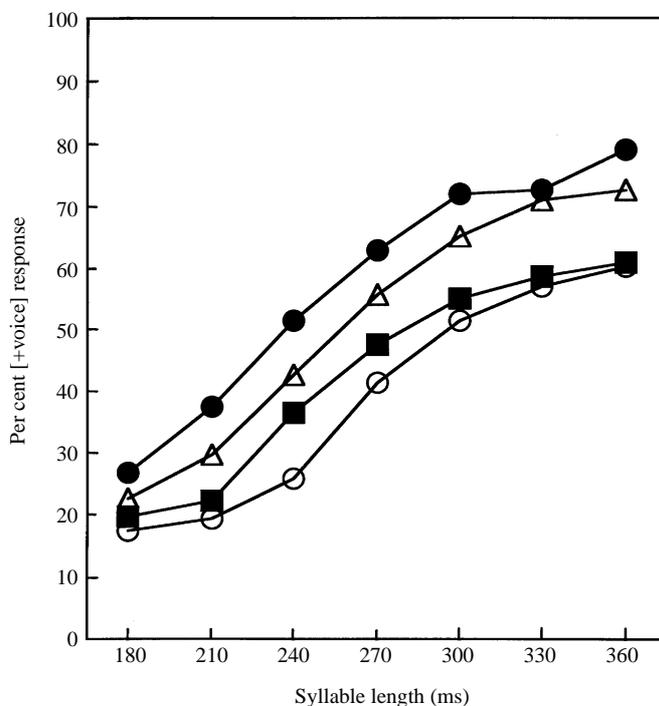
<sup>3</sup> Although the  $F_2$  value used for the steady-state /a/ in Experiment 1 is considerably higher than the average adult male value for that vowel reported by Peterson & Barney (1952) and by Syrdal (1985, from the Texas Instruments database collected by R.G. Leonard), it is well within the range of adult male  $F_2$  values for American English /a/ recently reported by Hillenbrand, Getty, Clark, & Wheeler (1995).

<sup>4</sup> The stimulus values used in Experiments 1 and 2 are broadly consistent with [voice]-conditioned  $f_0$  perturbations observed in natural utterances of American English speakers. For example, Hombert, Ohala, & Ewan (1979) reported that  $f_0$  in CV syllables was about 15–18 Hz higher following the release of /p/ than following the release of /b/, that the  $f_0$  contours following /b/ and /p/ were rising and falling, respectively, and that these [voice]-conditioned perturbations lasted as long as 100 ms. Although the frequency and temporal characteristics of the  $f_0$  contours used here are probably more typical of syllable-initial than syllable-final consonants (Kohler, 1982; Ohde, 1984), it was considered desirable for purposes of comparison to use symmetrical  $f_0$  values across the two syllable positions (see Experiment 2).

presented to subjects binaurally over Beyer DT-100 earphones at a peak level of 78 dB SPL. Subjects were seated at separate response stations in a double-walled sound-attenuated chamber.

## 2.2. Results and discussion

The results of two subjects were not included in the analysis: one had identified virtually all the stimulus tokens as containing a final /p/; the other misinterpreted the instructions and reversed the response categories. Fig. 2 displays the percentage of final [+voice] (i.e., /b/) responses for the four stimulus series as a function of stimulus duration. As expected, [+voice] judgments tended to increase at longer stimulus durations, consistent with earlier findings (e.g., Denes, 1955; Raphael, 1972; Summers, 1988; Fischer & Ohde, 1990).<sup>5</sup> More relevant to the purposes of the



**Figure 2.** Listener identification functions for the four stimulus series of Experiment 1. ○, High-high; ■, high-low; △, low-high; ●, low-low.

<sup>5</sup> It will be noted that while the identification functions are monotonic with respect to syllable duration, their slopes are quite shallow and average labeling performance does not asymptote near 0% and 100%. Among several other studies of perceptual effects of vowel duration on final [voice] judgments, there is large variation in the extent to which identification functions exhibit the steep slopes characteristic of categorical perception. For example, the functions reported by Lehiste (1977), Lehiste & Shockey (1980), and Raphael (1972) are very steep in the region of the category boundary, whereas those reported by Denes (1955), Summers (1988), and Fisher & Ohde (1990) are much shallower, changing gradually throughout the range of vowel durations used. Even in the latter cases, however, the functions typically (although not invariably) asymptote near 0% and 100%. In Experiment 1, the shallowness of the average functions and inconsistent labeling of series-endpoint stimuli reflect the fact that a subset of our subjects tended to ignore syllable duration and to rely instead mainly on  $f_0$  cues in making [voice] judgments. Other subjects in our sample identified the syllable duration dimension much more categorically.

present study, an increased percentage of [+voice] responses was produced by either a lower steady-state  $f_0$  or a lower offset  $f_0$  when other variables were held constant. An analysis of variance of the percentage of [+voice] responses showed significant main effects of stimulus duration [ $F(6, 126) = 33.56, p < 0.001$ ], steady-state  $f_0$  [ $F(1, 21) = 28.84, p < 0.001$ ], and offset  $f_0$  [ $F(1, 21) = 9.59, p < 0.01$ ]. There were no significant interactions among the three variables.

An additional analysis of variance was performed in which the two separate  $f_0$  factors were combined into a single  $f_0$  contour factor with four levels: High-High, High-Low, Low-High, Low-Low. The effect of contour was, as expected, highly significant [ $F(3, 63) = 10.14, p < 0.001$ ]. The specific purpose of this analysis was to allow a planned comparison between the Low-Low series and the Low-High series. Our interest in this particular comparison is that it permits a test of the low frequency hypothesis against an alternative account of the role of  $f_0$  in syllable-final [voice] judgments offered by Lehiste (1977) (see General discussion). Although the Low-Low series yielded more [+voice] responses than the Low-High series, as predicted by the low frequency hypothesis, the difference was not significant ( $p = 0.10$ ).

The effects of independently varying steady-state  $f_0$  and offset  $f_0$  on final [voice] judgments patterned similarly to effects of varying steady-state  $F_1$  and offset  $F_1$ , discussed earlier (Castleman *et al.*, unpublished data). In both cases, lower frequency values during the vowel steady-state and offset yielded more [+voice] responses, and the effects of frequency variation in the two temporal regions were additive. The similar effects of  $F_1$  and  $f_0$  manipulation on final poststressed [voice] judgments extend the parallel, noted earlier, between the effects of  $F_1$  and  $f_0$  variation on initial prestressed [voice] judgments. All of these findings are thus consistent with the claim that a low  $f_0$  and a low  $F_1$  both contribute to a single integrated perceptual quality corresponding to the low frequency property (Stevens & Blumstein, 1981; Diehl & Kingston, 1991; Kingston & Diehl, 1994, 1995).

Results reported by Fischer & Ohde (1990) suggest that an important qualification of the low frequency hypothesis is necessary. Although Castleman *et al.* (unpublished data) found that lower values of  $F_1$  steady-state and  $F_1$  offset contributed additively to more final [+voice] identification responses, Fischer & Ohde obtained an interaction between these  $F_1$  variables. In particular, a lower  $F_1$  offset value had less effect on [voice] judgments in the case of a high vowel with a low  $F_1$  steady-state than in the case of a low vowel with a high  $F_1$  steady-state. Moreover, for a given  $F_1$  offset value, a high vowel/low  $F_1$  steady-state yielded fewer [+voice] responses than did a low vowel/high  $F_1$  steady-state, apparently contrary to the low frequency hypothesis.

The discrepancies between the results of Castleman *et al.* and those of Fischer & Ohde may be attributable to a difference between the two studies in the range of variation of the  $F_1$  steady-state values. Castleman *et al.* varied the  $F_1$  steady-state between 700 Hz and 800 Hz, so that all vowel tokens were well within the /a/ category for adult male talkers. Fischer & Ohde varied the  $F_1$  steady-state over a much larger range—between 350 Hz and 650 Hz—corresponding to differences among the adult male vowel categories /i/, /ɪ/, and /æ/. Leaving aside talker differences, it is reasonable to assume that vowel  $F_1$  is used by listeners primarily as a cue for vowel height (Delattre, Liberman, Cooper, & Gerstman, 1952; Miller, 1953) and secondarily as a cue for final consonant [voice] status (Summers, 1988). Only

when  $F_1$  variation is limited to within-vowel-category differences (and other factors such as age and sex of the implied talker are held constant) can a lower  $F_1$  steady-state be unambiguously interpreted as due to a [+voice] final consonant. Thus, a lower  $F_1$  steady-state contributes to the low frequency property associated with [+voice] consonants just to the extent that other factors influencing  $F_1$  have already been accounted for. Put another way, the phonetic values that make up the low frequency property, including both  $F_1$  and  $f_0$ , must be interpreted by listeners in relative rather than absolute terms. According to this qualification, the low frequency hypothesis predicts that the perceptual effects of varying  $F_1$  and  $f_0$  will pattern similarly when listeners can reasonably attribute each of those variations to the voicing status of the consonant rather than to other factors.

### 3. Experiment 2

In the study by Diehl & Molis (1995), cited earlier, the only part of the  $f_0$  contour that influenced [voice] judgments in utterance-medial position was the  $f_0$  value at voicing onset following the consonantal release. The similarity between this finding and the results for [voice] judgments in utterance-initial (hence, syllable-initial) prestressed position (Massaro & Cohen, 1976; Haggard *et al.*, 1981) suggests that listeners perceived the consonants in the VCV stimuli used by Diehl & Molis as prestressed and syllable-initial. Several aspects of the stimulus design are consistent with this interpretation. First, the initial vowel was approximately half the duration of the final CV, yielding an iambic metrical foot corresponding to an unstressed-stressed disyllable (Hayes, 1995). Second, across the stimulus set, closure duration was fixed, while VOT varied from 10 ms to 45 ms. Closure duration normally exhibits large variation between [+voice] and [-voice] categories in medial poststressed position but little variation in medial prestressed position (Lisker, 1957, 1972), while VOT shows just the opposite pattern of variation (Kahn, 1976). Thus, the particular [voice] cues in the Diehl & Molis stimuli were also consistent with an unstressed-stressed disyllable. Finally, according to most current phonological theories of syllabification (Kahn, 1976; Selkirk, 1982; Clements & Keyser, 1983), the consonant in an unstressed-stressed VCV disyllable is exclusively affiliated with the second syllable (i.e., the consonant is syllable-initial).

The aim of Experiment 2 was to study the effect of  $f_0$  on medial [voice] judgments in stressed-unstressed VCV disyllables. English consonants in this position are often analyzed as being affiliated with the first syllable, either exclusively (Hoard, 1971; Selkirk, 1982) or ambisyllabically (Kahn, 1976). In the latter case, the consonant is assumed also to be affiliated with the second syllable, so that it is both syllable-final and syllable-initial. A recent alternative to Kahn's analysis is that the consonant in stressed-unstressed VCV disyllables is a geminate (see, e.g., Burzio, 1994). If any of these analyses is correct, the effects of  $f_0$  variation on [voice] judgments in stressed-unstressed VCVs might be expected to pattern similarly to the effects observed in Experiment 1 for utterance-final poststressed consonants. That is, the domain of  $f_0$  influence should include at least the steady-state and offset  $f_0$  values of the first syllable. Moreover, if either the ambisyllabic or geminate account is correct, the domain of  $f_0$  influence should also include the  $f_0$  value at voicing onset in the second syllable.

Two versions of Experiment 2, differing in the design of the stimulus set, were

conducted in order to assess the specific domain of  $f_0$  influence. In both versions (2a and 2b), the VCV stimuli varied in closure duration with VOT fixed at a small positive value. As noted, this pattern of variation of [voice] cues is consistent with a stressed-unstressed disyllable. However, in Experiment 2a, two properties of the stimuli were atypical of English stressed-unstressed disyllables: the first and second syllables were equal in duration, and the unreduced vowel /a/ was used for both syllables. Experiment 2a permits a test of the specific domain of influence of  $f_0$  when the first and second syllables are nearly symmetrical in acoustic shape, and thus any differences in the role of  $f_0$  between the two syllables are unlikely to be attributable to acoustic differences per se. But an obvious disadvantage of this design is that the inconsistent stress cues may make syllabification difficult for the listener. In Experiment 2b, the VCV disyllables were designed to be acoustically more consistent with a stressed-unstressed pattern. In addition to the use of [voice] cues characteristic of stressed-unstressed disyllables, the first syllable was significantly longer than the second, and whereas the first syllable contained the unreduced vowel /a/, the second contained the vowel /ə/.

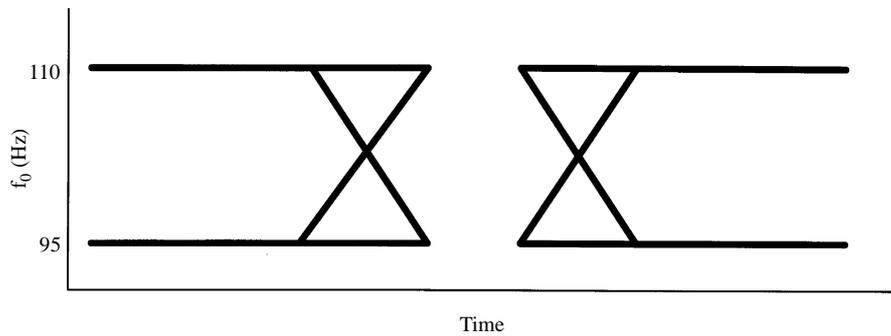
### 3.1. Method

#### 3.1.1. Stimuli

In both Experiments 2a and 2b, 16 three-formant stimulus series ranging from /VbV/ to /VpV/ were prepared using the Klatt88 synthesizer. All series varied in closure duration from 20 ms to 120 ms in 20 ms steps, with the closure interval entirely silent. Following the closure interval, there was a 10-ms VOT, simulated by attenuating the first formant and by exciting the higher formants by an aperiodic source. The 16 series differed according to the  $f_0$  value, either 110 Hz or 95 Hz, at each of four locations within the disyllable. From stimulus onset until 70 ms before closure onset,  $f_0$  was fixed at either the higher or lower value (*first syllable steady-state*). From the end of this steady-state interval to the onset of closure,  $f_0$  either remained at the onset value, or fell linearly from the higher to the lower value, or rose linearly from the lower to the higher value (*first syllable offset*). At voicing onset,  $f_0$  was set at either the higher or lower value (*second syllable onset*), and either remained constant or changed linearly to the opposing value (i.e., either lower or higher) over a 70 ms interval. At the end of this interval,  $f_0$  remained unchanged to the end of the stimulus (*second syllable steady-state*). Fig. 3 shows the resulting  $f_0$  contours used in each of Experiments 2a and 2b.

The stimuli in Experiment 2a consisted of /aba/-/apa/ disyllables, with preclosure and postclosure segments equal in duration. The initial and final steady-state formants were 100 ms in duration and had frequencies of 750 Hz ( $F_1$ ), 1150 Hz ( $F_2$ ), and 2400 Hz ( $F_3$ ). The linear CV and VC formant transitions were 35 ms in duration, and had offset/onset values of 150 Hz ( $F_1$ ), 750 Hz ( $F_2$ ), and 1700 Hz ( $F_3$ ).

The stimuli in Experiment 2b were the same except that the steady-state formant interval of the first syllable was lengthened to 150 ms, and the final vowel was changed to /ə/, with steady-state formant frequencies of 600 Hz ( $F_1$ ), 1300 Hz ( $F_2$ ), and 2450 Hz ( $F_3$ ). CV formant onset frequencies were identical to those in Experiment 2a.



**Figure 3.** Schematic diagram of the contours used in each of Experiments 2a and 2b. The 16 contours represent all possible combinations of the binary  $f_0$  values (110 Hz or 95 Hz) in four temporal regions of the disyllable: first syllable steady-state, first syllable offset, second syllable onset, and second syllable steady-state. The relative durations depicted are correct only for Experiment 2a; in Experiment 2b the second syllable (and associated steady-state  $f_0$  contour) was, as noted, shorter than the first.

### 3.1.2. Subjects and procedure

Twenty-one subjects served in Experiment 2a, and twelve served in Experiment 2b. All were drawn from the same population as the subjects in Experiment 1. In Experiment 2a subjects identified each the 96 stimuli (16 series  $\times$  6 closure durations) eight times in one hour long session. In Experiment 2b subjects identified each of the 96 stimuli five times in the session. Every other aspect of the procedure was the same as in Experiment 1.

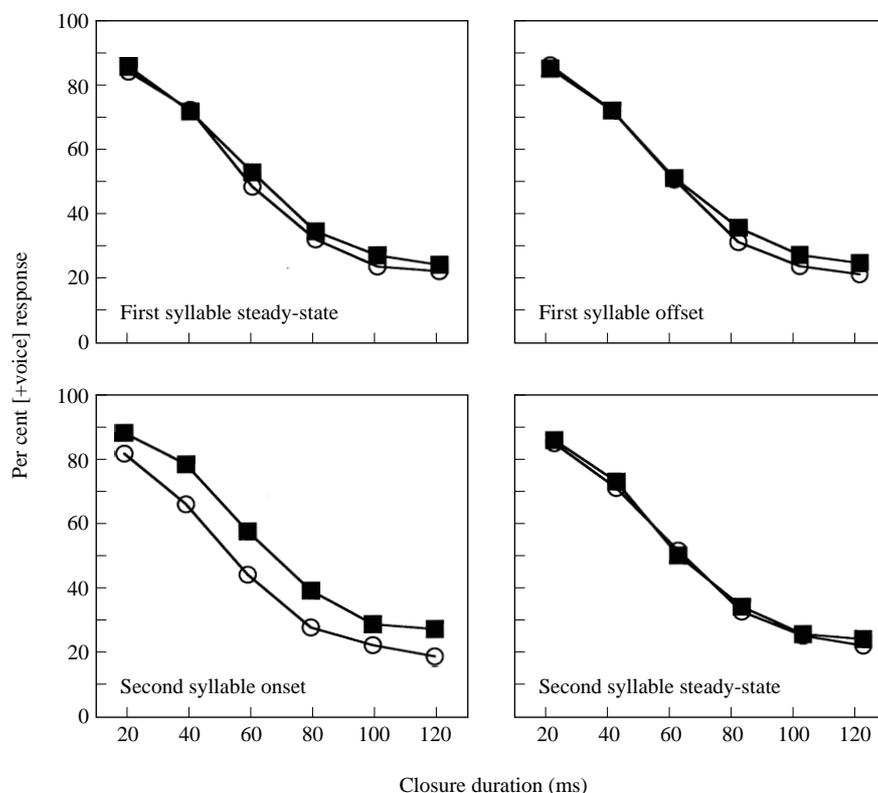
### 3.2. Results and discussion

The percentage of [+voice] responses for the various stimulus series as a function of closure duration is shown in Fig. 4 for Experiment 2a and in Fig. 5 for Experiment 2b. Consistent with earlier findings (e.g., Lisker, 1957), [+voice] judgments tended to decrease at longer closure durations.<sup>6</sup> More interesting was the effect of  $f_0$  variation: lower values of either the steady-state or offset  $f_0$  of the first syllable or the onset  $f_0$  of the second syllable, produced more [+voice] responses; only the steady-state  $f_0$  of the second syllable failed to influence labeling performance.

An analysis of variance of the percentage of [+voice] responses in Experiment 2a showed significant main effects of closure duration [ $F(5, 100) = 98.316, p < 0.001$ ], first syllable steady-state  $f_0$  [ $F(1, 20) = 4.578, p < 0.05$ ], and first syllable offset  $f_0$  [ $F(1, 20) = 5.152, p < 0.05$ ], and a main effect of second syllable onset  $f_0$  [ $F(1, 20) = 17.957, p < 0.001$ ], but no significant effect of second syllable steady-state  $f_0$  [ $F(1, 20) = 1.088, p > 0.3$ ]. There was also one significant two-way interaction between first syllable steady-state  $f_0$  and first syllable offset  $f_0$  [ $F(1, 20) = 28.276, p < 0.001$ ].

A corresponding analysis of variance for Experiment 2b yielded significant main

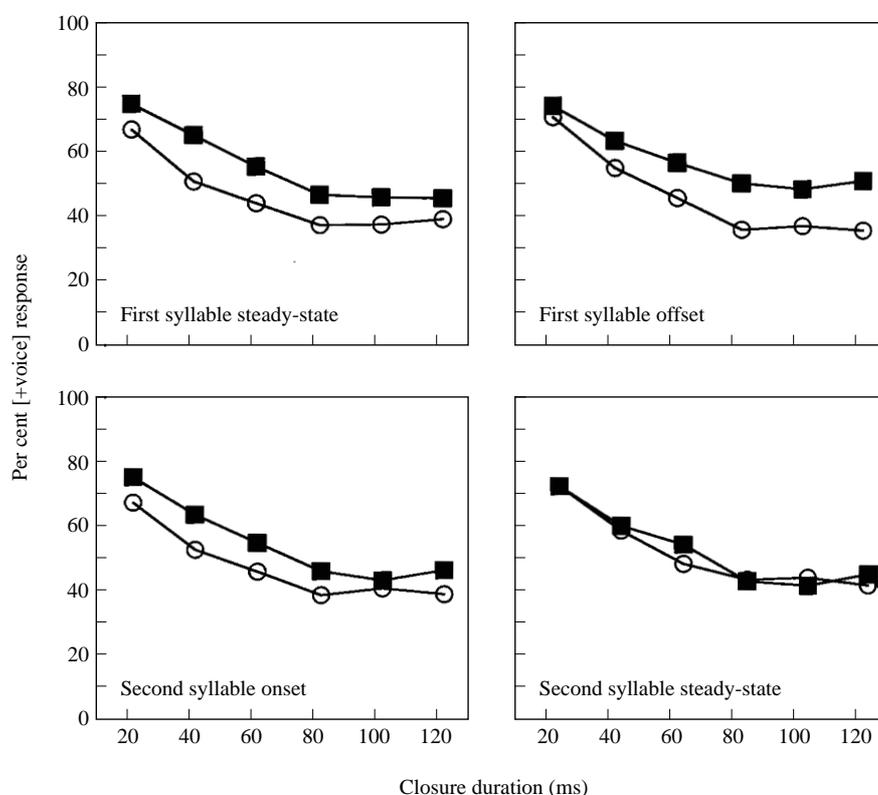
<sup>6</sup> As in Experiment 1, the identification functions (especially for Experiment 2b) were quite shallow, but largely for a different reason. Each function in Figs. 4 and 5 displays the results averaged over eight conditions (two  $f_0$  values in each of the other three temporal regions). To the extent that the  $f_0$  variable in any of these other regions has an effect on [voice] judgments, it contributes variance to the displayed effects of closure duration, and this helps to account for the relative shallowness of the identification functions.



**Figure 4.** Listener identification functions for Experiment 2a. Each panel displays the effect of  $f_0$  variation in one of the four temporal regions of the disyllable.  $\circ$ ,  $f_0 = 110$  Hz;  $\blacksquare$ ,  $f_0 = 95$  Hz.

effects of closure duration [ $F(5, 55) = 14.06$ ,  $p < 0.001$ ], first syllable steady-state  $f_0$  [ $F(1, 11) = 16.41$ ,  $p < 0.01$ ], first syllable offset  $f_0$  [ $F(1, 11) = 10.72$ ,  $p < 0.01$ ], second syllable onset  $f_0$  [ $F(1, 11) = 47.76$ ,  $p < 0.001$ ], but no significant effect of second syllable steady-state  $f_0$  [ $F(1, 11) = 1.64$ ,  $p > 0.2$ ]. Among  $f_0$  variables, there was one significant two-way interaction between the first syllable steady-state  $f_0$  and the second syllable onset  $f_0$  [ $F(1, 11) = 5.99$ ,  $p < 0.05$ ].

A second analysis of variance was performed on the data of Experiment 2b in which the separate  $f_0$  factors were combined into a single  $f_0$  contour factor, the effect of which was highly significant [ $F(3, 11) = 11.51$ ,  $p < 0.001$ ]. Analogous to Experiment 1, the specific purpose of this analysis was to permit a planned comparison between the four stimulus series whose first syllable had a Low-Low  $f_0$  contour and the four series whose first syllable had a Low-High  $f_0$  contour. (The results from Experiment 2b were considered more appropriate for this purpose than those of Experiment 2a because of the small size of the  $f_0$  effects in Experiment 2a.) Again, this comparison tests the low frequency hypothesis against an alternative hypothesis of Lehiste (1977) (see General Discussion). Consistent with the low frequency hypothesis, the four series with a Low-Low  $f_0$  contour in the first syllable yielded significantly more [+voice] responses than the four series with a Low-High  $f_0$  contour ( $p < 0.01$ ).



**Figure 5.** Listener identification functions for Experiment 2b. Each panel displays the effect of  $f_0$  variation in one of the four temporal regions of the disyllable.  $\circ$ ,  $f_0 = 110$  Hz;  $\blacksquare$ ,  $f_0 = 95$  Hz.

In both versions of Experiment 2, a lower value of either the first syllable steady-state or first-syllable offset  $f_0$  contributed to an increase in medial [+voice] labeling responses, although the effects were considerably smaller in Experiment 2a than in Experiment 2b. These  $f_0$  effects were in the same direction as those observed in Experiment 1 for [voice] judgments in utterance-final position. Also, in Experiment 2, a lower  $f_0$  in the second syllable increased [+voice] responses, but only in the region of voicing onset. This resembles the  $f_0$  effect reported for [voice] judgments in syllable-initial prestressed position both utterance-initially (Massaro & Cohen, 1976; Haggard *et al.*, 1981) and utterance-medially (Diehl & Molis, 1995).

A possible account of these parallel  $f_0$  effects across various studies rests on the assumption of differences in syllabification of VCV utterances dependent on stress location. Consider, first, the parallel  $f_0$  effects observed for the utterance-final (hence syllable-final) [voice] judgments in Experiment 1 and the utterance-medial [voice] judgments of Experiments 2a and 2b. If it is assumed that the consonant in the latter experiments was affiliated with the first syllable and was, therefore, syllable-final, then the parallel  $f_0$  effects might be ascribed to the common syllable position of the target consonant. As noted earlier, several theories of syllabification assign a single medial consonant of a stressed-unstressed disyllable to the first syllable exclusively (Hoard, 1971; Selkirk, 1982), ambisyllabically (Kahn, 1976), or as a geminate

(Burzio, 1994). The pattern of variation of [voice] cues used in Experiments 2a and 2b was consistent with stressed-unstressed disyllables, and in Experiment 2b vocalic and durational cues were consistent with that stress pattern as well. Accordingly, it is plausible that the consonant in Experiment 2 (especially 2b) was analyzed by listeners as syllable-final.

Next, consider the parallel between the second syllable  $f_0$  effect in Experiment 2 and the effects observed utterance-initially by Massaro & Cohen (1976) and Haggard *et al.* (1981) and utterance-medially by Diehl & Molis (1995). This parallel may perhaps be explained on the assumption that in all these cases the target consonant was syllable-initial. This assumption is obviously correct in utterance-initial case, and it is also likely to be correct in the medial case studied by Diehl & Molis, since the syllable durations and variation in [voice] cues in the latter case were consistent with an unstressed-stressed disyllable pattern. [As noted earlier, for this stress pattern the medial consonant is usually assumed to be exclusively affiliated with the second syllable (Kahn, 1976; Selkirk, 1982; Clements & Keyser, 1983).] For the disyllables of Experiment 2, the medial consonant would be affiliated with the second syllable under the assumption that the consonant is either ambisyllabic (Kahn, 1976) or a geminate (Burzio, 1994) in stressed-unstressed disyllables of this type. Although the issue remains controversial, articulatory evidence supporting the ambisyllabic/geminate assumption has recently been reported by Turk (1993). Therefore, the results of the present study and of earlier experiments suggest tentatively that the domain within which  $f_0$  variation affects [voice] judgments is determined by the syllable affiliation of the target consonant.

The above account is consistent with the considerably smaller effects of first syllable steady-state  $f_0$  and first syllable offset  $f_0$  in Experiment 2a than in Experiment 2b. Recall that for Experiment 2b, the [voice] cues, syllable durations, and vowel quality in the second syllable of the stimuli were all appropriate for a stressed-unstressed disyllable pattern. However, for Experiment 2a, only the [voice] cues of the stimuli were appropriate for that stress pattern. It appears likely, therefore, that the stimuli of Experiment 2a were perceived by listeners as somewhat ambiguous with respect to stress pattern. That is, those stimuli appear to have been located perceptually somewhere between the clear stressed-unstressed disyllable patterns of Experiment 2b and the clear unstressed-stressed disyllable patterns used by Diehl & Molis (1995), which yielded no first syllable  $f_0$  effects.

#### 4. General discussion

Experiment 1 tested a prediction of the low frequency hypothesis, viz., that effects of varying  $f_0$  on utterance-final [voice] judgments should pattern similarly to effects of varying  $F_1$  (Summers, 1988; Castleman *et al.*, unpublished data). This prediction was confirmed: analogous to the effects of  $F_1$  variation, a lower steady-state  $f_0$  and a lower offset  $f_0$  both increased [+voice] labeling responses, and their effects were additive. The similarity between the effects of  $f_0$  and  $F_1$  on utterance-final [voice] judgments extends the parallel, noted earlier, between the effects of these two variables on utterance-initial [voice] judgments (Lisker, 1975; Massaro & Cohen, 1976; Summerfield & Haggard, 1977; Haggard *et al.*, 1981; Kluender, 1991). These findings jointly support the claim that a low  $F_1$  and a low  $f_0$  both contribute to an integrated perceptual correlate of [+voice] consonants that we refer to as the low

frequency property (for a fuller discussion, see Stevens & Blumstein, 1981; Diehl & Kingston, 1991; Kingston & Diehl, 1994).

Experiment 2 examined the effective domain in which  $f_0$  variation affects medial [voice] judgments for VCV stimuli whose primary [voice] cues are consistent with a stressed-unstressed disyllabic pattern. A clear parallel was observed between the effects of  $f_0$  variation in the first syllable and the  $f_0$  effects observed in Experiment 1 for utterance-final [voice] judgments. There was also a parallel between the effects of  $f_0$  variation in the second syllable and  $f_0$  effects reported for utterance- and syllable-initial [voice] judgments (Massaro & Cohen, 1976; Haggard *et al.*, 1981; Diehl & Molis, 1995). Thus, [+voice] labeling responses increased as a function of a lower steady-state and offset  $f_0$  in the first syllable and as a function of a lower onset  $f_0$  in the second syllable, but the steady-state  $f_0$  value of the second syllable had no effect. The effects observed for the VCV disyllables of Experiment 2 were notably different from those reported by Diehl & Molis (1995) for unstressed-stressed VCV disyllables. In the latter case, only the onset  $f_0$  of the second syllable influenced [voice] responses. We tentatively attribute these varying outcomes to a difference in syllabification of the VCVs owing to a difference in stress pattern. According to this account, the stimuli in the Diehl & Molis study were syllabified as V.CV, whereas the stimuli of Experiment 2 (especially Experiment 2b) were syllabified (following Kahn, 1976, or Burzio, 1994) as VC.CV.

Before accepting the low frequency hypothesis as an explanation of the effect of  $f_0$  variation near the target consonant, it is necessary to consider two alternative proposals that have appeared in the literature. Lehiste (1976) found that the perceived duration of a vowel is greater for a changing than for a monotone  $f_0$ . On the basis of this result, Lehiste (1977, reviewed in Lehiste & Shockey, 1980) hypothesized that a changing  $f_0$  contour would make a following target consonant appear more [+voice], since actual vowel duration is typically greater before [+voice] than before [-voice] consonants. This prediction was, in fact, supported in Lehiste's (1977) study as well as in the studies by Derr & Massaro (1980) and Gruenenfelder & Pisoni (1980), cited earlier.

Both the studies by Lehiste (1977) and Gruenenfelder & Pisoni (1980) compared monotone stimuli to falling  $f_0$  stimuli with terminal  $f_0$  values lower than the monotone values. The results were thus consistent with either the low frequency hypothesis or Lehiste's "changing  $f_0$ " hypothesis. A critical test of the two hypotheses is to compare monotone stimuli and rising  $f_0$  stimuli having starting  $f_0$  values no lower than the monotone values. The low frequency hypothesis predicts that a consonant will be judged more often as [+voice] when preceded by the monotone  $f_0$ , whereas Lehiste's hypothesis predicts that more [+voice] responses will occur for the rising  $f_0$  stimuli. As noted earlier, the evidence on this issue is inconsistent: Derr & Massaro (1980) found that upward deflection of an otherwise monotone  $f_0$  yielded an increase in [+voice] judgments for the following consonant (consistent with the changing  $f_0$  hypothesis), whereas Kohler (1985) and Kohler & van Dommelen (1986) found that such an upward deflection produced more [-voice] judgments (consistent with the low frequency hypothesis). As for the present study, planned comparisons between the first-syllable Low-Low and Low-High conditions of Experiment 2b clearly favored the low frequency hypothesis: an upward movement of  $f_0$  before the target consonant uniformly increased [-voice] labeling responses. (In Experiment 1, the difference between the Low-Low

and Low-High conditions was in the same direction but not statistically significant.) Thus, on balance, the perceptual evidence appears to be more consistent with the low frequency hypothesis than with the changing  $f_0$  hypothesis.

Considerations of theoretical generality and parsimony also favor the low frequency hypothesis over the changing  $f_0$  hypothesis. The low frequency hypothesis offers a single unified account of the effects on [voice] judgments of (a) monotone *vs.* changing (rising and falling)  $f_0$  conditions in the present study as well as in most of the other studies reviewed in the preceding two paragraphs, (b) the  $f_0$  value of monotone stimuli (e.g., the Low-Low *vs.* High-High conditions in Experiment 1), and (c) the value of  $f_0$  onset following the consonant, and, additionally, the hypothesis accounts for (d) the various parallel effects of varying  $f_0$  and  $F_1$  for a given utterance position and stress pattern. In comparison, the changing  $f_0$  hypothesis accounts at most for a subset of the effects included under category (a) above.

A final argument against the changing  $f_0$  hypothesis was provided by Rosen (1977a, b). He measured effects of  $f_0$  contour on vowel duration discrimination and on category boundary location for the long-short vowel distinction in Swedish, and found that in both cases the effect of  $f_0$  on perceived duration was negligible. These results are at odds with the assumption that the effect of  $f_0$  contour on [voice] judgments is mediated by its effect of perceived duration.

A second alternative to the low frequency hypothesis has been proposed by Silverman (1986, 1987). According to Silverman's hypothesis, local perturbations in the  $f_0$  contour associated with the [voice] distinction are perceptually evaluated relative to the global intonation contour across the syllables near the target consonant. Relative to this global contour, a steeply falling  $f_0$  perturbation following the consonant release (i.e., an  $f_0$  movement starting from a frequency considerably higher than the interpolated value of the global intonation contour in that region) will signal a [-voice] consonant, while a slightly falling or level  $f_0$  will cue a [+voice] consonant. Silverman claims that if the global intonation contour is appropriately factored out, rising  $f_0$  perturbations will not occur for either [voice] category in natural utterances, and if rising perturbations are synthesized, they will be perceptually equivalent to a level or slightly falling  $f_0$  in cuing [+voice] consonants.

Silverman's hypothesis is similar to the low frequency hypothesis in regarding the value of  $f_0$  in the vicinity of the consonant as an important perceptual correlate of the [voice] distinction. The two hypotheses are also similar in assuming that  $f_0$  value near the consonant is perceptually interpreted in a relative manner. However, while in Silverman's view the interpretation is relative to the global intonation contour, the version of the low frequency hypothesis that we endorse assumes that  $f_0$  is interpreted relative to the speaker's overall average  $f_0$ . Diehl & Molis (1995) created various VCV stimulus series that were designed to test three versions of Silverman's hypothesis as well as three versions of the low frequency hypothesis (the latter differing according to the temporal domain over which  $f_0$  is assumed to affect [voice] judgments). For the unstressed-stressed disyllabic stimuli used in the study, the most effective version of the low frequency hypothesis for predicting listener identification performance was one that restricted the domain of  $f_0$  influence to the moment of voicing onset (see above discussion). Moreover, this version of the low frequency hypothesis turned out to be a far better predictor of identification performance than any of the three versions of Silverman's hypothesis.

We conclude that the low frequency hypothesis provides a simple and general

account of the influence of *f*<sub>0</sub> contour on [voice] judgments in various utterance and syllable positions, and that its predictive power is superior to that of the alternative proposals reviewed in the preceding paragraphs. Studies are now being planned to test more stringently the tentative claim that the domain of influence of the low frequency property on [voice] judgments is determined by the syllable affiliation of the target consonant.

This work was supported by research grant number 5 R01 DC00427-07, -08 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, to the second author. We thank Pam Beddor and two anonymous reviewers for their very helpful comments on an earlier version of this paper.

### References

- Burzio, L. (1994) *Principles of English stress*. Cambridge, U.K.: Cambridge University Press.
- Castleman, W. A., Hughes, D. L., & Diehl, R. L. (unpublished data) [The influence of steady-state and offset *F*<sub>1</sub> on final [voice] judgments.]
- Chistovich, L. A. (1969) Variations of the fundamental voice pitch as a discriminatory cue for consonants, *Soviet Physics-Acoustics*, **14**, 372–378.
- Clements, G. N., & Keyser, S. J. (1983) *CV phonology: a generative theory of the syllable*. Cambridge, MA: MIT Press.
- Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952) An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns, *Word*, **8**, 195–210.
- Denes, P. (1955) Effect of duration on the perception of voicing, *Journal of the Acoustical Society of America*, **27**, 761–764.
- Derr, M. A. & Massaro, D. W. (1980) The contribution of vowel duration, *f*<sub>0</sub> contour, and frication duration as cues to the /*juz*-/*jus*/ distinction, *Perception & Psychophysics*, **27**, 51–59.
- Diehl, R. L., Castleman, W. A., & Kingston, J. (1995) On the internal perceptual structure of phonological features: the [voice] distinction, *Journal of the Acoustical Society of America*, **97**, 3333 (Abstract).
- Diehl, R. L., & Kingston, J. (1991) Phonetic covariation as auditory enhancement: the case of the [+voice]/[-voice] distinction. *PERILUS: Papers from the conference on current phonetic research paradigms: implications for speech motor control, Stockholm University, August 13–16, 1991*. pp. 139–143.
- Diehl, R. L., & Molis, M. R. (1995) Effect of fundamental frequency on medial [voice] judgments, *Phonetica*, **52**, 188–195.
- Fischer, R. M., & Ohde, R. N. (1990) Spectral and duration properties of front vowels as cues to final stop-consonant voicing, *Journal of the Acoustical Society of America*, **88**, 1250–1259.
- Fujimura, O. (1971) Remarks on stop consonants: synthesis experiments and acoustic cues. In *Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen*, pp. 221–232. Copenhagen: Akademisk Forlag.
- Gruenenfelder, T. M., & Pisoni, D. B. (1980) Fundamental frequency as a cue to postvocalic consonantal voicing: some data from speech perception and production, *Perception & Psychophysics*, **28**, 514–520.
- Haggard, M., Ambler, S., & Callow, M. (1970) Pitch as a voicing cue, *Journal of the Acoustical Society of America*, **47**, 613–617.
- Haggard, M., Summerfield, Q., & Roberts, M. (1981) Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading *F*<sub>0</sub> cues in the voiced-voiceless distinction, *Journal of Phonetics*, **9**, 49–62.
- Hayes, B. (1995) *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995) Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America*, **97**, 3099–3111.
- Hoard, J. E. (1971) Aspiration, tenseness, and syllabication in English, *Language*, **47**, 133–140.
- Hombert, J.-M. (1978) Consonant types, vowel quality, and tone. In *Tone: a linguistic survey* (V. Fromkin, editor), pp. 77–111. New York: Academic Press.
- Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979) Phonetic explanations for the development of tones, *Language*, **55**, 37–58.
- House, A. S., & Fairbanks, G. (1953) The influence of consonant environment on the secondary acoustical characteristics of vowels, *Journal of the Acoustical Society of America*, **25**, 105–135.
- Kahn, D. (1976) *Syllable-based generalizations in English phonology*. Bloomington, IN: Indiana University Press.
- Kingston, J., & Diehl, R. L. (1994) Phonetic knowledge, *Language*, **70**, 419–454.

- Kingston, J., & Diehl, R. L. (1995) Intermediate properties in the perception of distinctive feature values. In *Phonology and phonetic evidence: papers in laboratory phonology IV* (B. Connell & A. Arvaniti, editors) pp. 7–27. Cambridge, U.K.: Cambridge University Press.
- Kluender, K. R. (1991) Effects of first formant onset properties on voicing judgments result from processes not specific to humans, *Journal of the Acoustical Society of America*, **90**, 83–96.
- Kohler, K. J. (1982)  $f_0$  in the production of lenis and fortis plosives, *Phonetica*, **39**, 199–218.
- Kohler, K. J. (1985)  $F_0$  in the perception of lenis and fortis plosives, *Journal of the Acoustical Society of America*, **78**, 21–32.
- Kohler, K. J., & van Dommelen, W. A. (1986) Prosodic effects on lenis/fortis perception: Preplosive  $F_0$  and LPC synthesis, *Phonetica*, **43**, 70–75.
- Lehiste, I. (1976) Influence of fundamental frequency pattern on the perception of duration, *Journal of Phonetics*, **4**, 113–117.
- Lehiste, I. (1977) Contribution of pitch to the perception of segmental quality. In *Proceedings of the 9th international congress on acoustics*, Madrid, July, p. 522.
- Lehiste, I., & Peterson, G. E. (1961) Some basic considerations in the analysis of intonation, *Journal of the Acoustical Society of America*, **33**, 419–425.
- Lehiste, I., & Shockey, L. (1980) Labeling, discrimination and repetition of stimuli with level and changing fundamental frequency, *Journal of Phonetics*, **8**, 469–474.
- Lisker, L. (1957) Closure duration and the intervocalic voiced-voiceless distinction in English, *Language*, **33**, 42–49.
- Lisker, L. (1972) Stop duration and voicing in English. In *Papers in linguistics and phonetics to the memory of Pierre Delattre*, (A. Valdman, ed.), pp. 339–343. The Hague: Mouton.
- Lisker, L. (1975) Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, **57**, 1547–1551.
- Massaro, D. W., & Cohen, M. M. (1976) The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction, *Journal of the Acoustical Society of America*, **60**, 704–717.
- Massaro, D. W., & Cohen, M. M. (1977) Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction, *Perception & Psychophysics*, **22**, 373–383.
- Miller, R. L. (1953) Auditory tests with synthetic vowels, *Journal of the Acoustical Society of America*, **25**, 114–121.
- Mohr, B. (1971) Intrinsic variations in the speech signal, *Phonetica*, **23**, 65–93.
- Ohde, R. N. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing, *Journal of the Acoustical Society of America*, **75**, 224–230.
- Peterson, G. E., & Barney, H. L. (1952) Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, **24**, 175–184.
- Petersen, N. R. (1983) The effect of consonant type of fundamental frequency and larynx height in Danish, *Annual Report of the Institute of Phonetics, University of Copenhagen*, **17**, 55–86.
- Raphael, L. F. (1972) Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English, *Journal of the Acoustical Society of America*, **51**, 1296–1303.
- Rosen, S. M. (1977a) The effect of fundamental frequency patterns on perceived duration, *Speech Transmission Laboratory—Quarterly Progress and Status Report*, 17–30.
- Rosen, S. M. (1977b) Fundamental frequency patterns and the long-short vowel distinction in Swedish, *STL-QPSR*, **1**, 31–37.
- Selkirk, E. O. (1982) The syllable. In *The structure of phonological representations (part II)* (H. van der Hulst & N. Smith, editors.), pp. 337–383. Dordrecht: Foris.
- Silverman, K. (1986)  $F_0$  segmental cues depend on intonation: the case of the rise after voiced stops, *Phonetica*, **43**, 76–91.
- Silverman, K. (1987) The structure and processing of fundamental frequency contours. Ph.D. Dissertation, University of Cambridge.
- Stevens, K. N., & Blumstein, S. E. (1981) The search for invariant acoustic correlates of phonetic features. In *Perspectives on the study* (P. D. Eimas & J. L. Miller, editors.), pp. 1–38. Hillsdale, NJ: Erlbaum.
- Summerfield, A. Q., & Haggard, M. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants, *Journal of the Acoustical Society of America*, **62**, 435–448.
- Summers, W. V. (1987) Effects of stress and final consonant voicing on vowel production: articulatory and acoustic analyses, *Journal of the Acoustic Society of America*, **82**, 847–863.
- Summers, W. V. (1988)  $F_1$  structure provides information for final-consonant voicing, *Journal of the Acoustical Society of America*, **84**, 485–492.
- Syrdal, A. K. (1985) Aspects of a model of the auditory representation of American English vowels, *Speech Communication*, **4**, 121–135.
- Turk, A. E. (1993) Effects of position-in-syllable and stress on consonant articulation. Ph.D. Dissertation, Cornell University.