

Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production

THOMAS M. GRUENENFELDER and DAVID B. PISONI
Indiana University, Bloomington, Indiana 47405

Several recent studies by Lehiste have reported that changes in fundamental frequency (F0) can serve as a cue to perceived vowel length and, furthermore, that the perceived lengthening of the vowel can influence perception of the voicing feature of stop consonants in syllable-final position. In Experiment 1, we replicated Lehiste's basic results for stop consonants in final position. Experiment 2 extended these results to postvocalic fricatives. The final consonant in syllables of intermediate vowel duration was more often perceived as voiced when F0 was falling than when F0 was monotone. In Experiment 3, we examined the F0 contours produced by eight talkers before postvocalic stop consonants and fricatives in natural speech for minimal pairs of words differing in voicing. The amount of change of F0 over the vowel was no greater before voiced than voiceless consonants, suggesting that the earlier perceptual effects cannot be explained by appealing to regularities observed in the production of F0 contours in vowels preceding postvocalic consonants.

Recently, several investigators have observed that perceived vowel duration can be influenced by the fundamental frequency (F0) contour of the vowel (e.g., Lehiste, 1976; Pisoni, Note 1; Wang, Lehiste, Chang, & Darnovsky, Note 2). In particular, vowels with a changing F0 were found to be perceived as longer than vowels with a monotone F0. The direction of change seems to have little influence on this effect—vowels with rising F0s and vowels with falling F0s were both perceived as longer than vowels with level F0s (Lehiste, 1976).

Variations in vowel duration are known to have at least two important effects on phoneme identification. First, vowel duration systematically influences the identification of the vowel itself (e.g., Stevens, 1959). Two vowels with identical formant frequencies but different durations are frequently identified as different vowels. This effect is particularly strong for the /ε/ (short vowel)-/æ/ (long vowel) distinction. Second, vowel duration has been shown to affect the perception of voicing of a following consonant (e.g., Denes, 1954). A consonant following a long vowel is more likely to be perceived as voiced

than an acoustically identical consonant following the same vowel of shorter duration.

Vowel duration, then, can serve as a phonemic cue. Because the F0 contour affects perceived vowel duration, it seems reasonable to suppose that the F0 contour over a syllable can also serve as a phonemic cue. In particular, placing a contour on the F0 of a synthesized vowel should affect its identification in a way similar to lengthening its duration. Furthermore, because of this perceived vowel-lengthening, consonants following a vowel with a changing F0 should more likely be perceived as voiced than consonants following a vowel with a monotone F0.

Lehiste (Note 3) has reported data that are consistent with the second hypothesis described above. She showed that F0 contour affects the perception of the voicing of a postvocalic consonant. Lehiste synthesized two stimulus continua, each of which ranged from the word "bat" to the word "bad." The stimuli in each series varied in vowel duration. In one series, the F0 of the vowel was kept constant at 80 Hz, whereas in the second, the F0 fell from 80 to 60 Hz. These test stimuli were presented to listeners who were required to identify them as either "bat" or "bad." The results indicated that perception changed from "bat" to "bad" at a shorter vowel duration for stimuli with a falling F0 than for those with a level F0. Similar results were obtained with a test continuum ranging from "beat" to "bead." That is, perception changed from "beat" to "bead" at a shorter vowel duration when F0 was falling than when F0 was level. Lehiste interpreted her results as

This paper is a substantially expanded version of a paper delivered by the first author at the 97th meeting of the Acoustical Society of America, June 12-17, 1979, Cambridge, Massachusetts. The research reported here was supported by NIMH Research Grant MH-24027 to Indiana University. Requests for reprints should be sent to either Thomas M. Gruenenfelder, who is now at Bell Laboratories, Holmdel, New Jersey 07733, or David B. Pisoni, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

support for the hypothesis that the changing F0 contour, through its mediating effect on perceived vowel duration, can serve as a cue to the voicing of a postvocalic consonant.

However, Rosen (Note 4) has failed to confirm the earlier findings that F0 affects the perceived duration of vowels. In Swedish, several pairs of vowels may be contrasted by duration alone. Using Swedish vowels and Swedish listeners, Rosen did not find systematic effects of F0 contour on vowel identification—a finding that is contrary to the hypothesis that F0 contour, through its mediating effect on vowel duration, can serve as a cue to vowel identification. Rosen's findings have to be interpreted carefully since he used a restricted range of durations in his test stimuli. Moreover, in addition to differences in duration in Swedish, there are also vowel quality differences that serve to differentiate minimal pairs. These differences were apparently not incorporated in Rosen's test stimuli.

The present paper reports the results of three experiments that examined the influence of F0 contour on segmental identification in greater detail. All the experiments were concerned with the use of F0 contours as a cue to postvocalic voicing. The first experiment was designed as a replication of Lehiste's earlier results for stop consonants. Because of Rosen's negative findings on vowel identification in Swedish, we thought a replication of Lehiste's work would be worthwhile. The second experiment generalized Lehiste's basic finding with postvocalic stops to postvocalic fricatives. Finally, the third experiment, a production study, examined the F0 contours produced before voiced and voiceless consonants in natural speech for several talkers. The production study was designed to determine whether the perceptual results could be accounted for by appealing to regularities observed in the production of F0 contours as a function of postvocalic consonantal voicing.

EXPERIMENT 1

Method

Subjects. Four Indiana University students served as subjects. All subjects were right-handed, native speakers of English and were paid \$3.00 for participating in a single experimental session that lasted about 75 min.

Stimuli. The stimuli for Experiment 1 were constructed using the Klatt (1980) software speech synthesizer as implemented on the PDP 11/05 computer in the Speech Perception Laboratory at Indiana University (Kewley-Port, 1978). Two 11-item test continua were synthesized, each ranging from the word "bat" to the word "bad." The parameters of the initial and final consonants were the same for all stimuli. To synthesize the initial /b/, F1 was increased linearly from 575 to 675 Hz, F2 from 1,325 to 1,425 Hz, and F3 from 2,370 to 2,470 Hz during the initial 35 msec of the stimulus. The final stop was also synthesized by means of formant transitions. F1 fell linearly from 675 to 280 Hz, F2 rose linearly from 1,425 to 1,600 Hz, and F3 rose linearly from 2,470 to 2,930 Hz

over the final 40 msec of the stimulus. The final stop was unreleased. The vowel was synthesized using steady-state formants of 675, 1,425, and 2,470 Hz. F4 and F5 were kept fixed at 3,300 and 3,700 Hz, respectively, throughout the duration of the stimulus.

In both test series, vowel duration was varied from 100 to 300 msec in 20-msec steps. The shortest vowel duration was intended to produce a good exemplar of the word "bat," and the longest, a good exemplar of the word "bad." For the monotone condition, F0 was held constant at 120 Hz throughout the duration of the stimulus. The falling F0 continuum was identical to the monotone series, except that the F0 was initially set at 150 Hz and fell linearly to 90 Hz during the second half of the vowel. All stimuli were low-pass filtered at 5 kHz, stored on a computer disk, and later output to subjects in real-time via a 12-bit D-A converter.

Procedure. The stimuli were presented over TDH-39 headphones at a comfortable listening level of about 80 dB SPL. The listeners identified the stimuli in the falling continuum and monotone continuum in separate blocks of trials. Each of the 11 stimuli in each continuum was presented once in each of 20 blocks. The order of stimulus presentation was randomized in each block. The listeners were required to identify each stimulus as either the word "bat" or "bad" by pressing one of two appropriately labeled buttons on a response box interfaced to the computer. Stimuli were separated by a 3-sec interstimulus interval.

Results and Discussion

Figure 1 shows the proportion of "bat" responses as a function of vowel duration for the monotone continuum (dotted line) and the falling continuum (solid line) averaged over the four subjects. The proportion of "bad" responses is simply 1 minus the proportion of "bat" responses. Each point is based on 20 observations by each of four subjects, for a total of 80 observations per point. The general pattern of results is clear. Perception changes from "bat" to "bad" at shorter vowel durations in the falling continuum than in the monotone continuum. This result replicates Lehiste's (Note 3) earlier finding that a changing F0 leads to a higher proportion of voiced responses for postvocalic stops. Three of the four subjects showed a pattern of results consistent

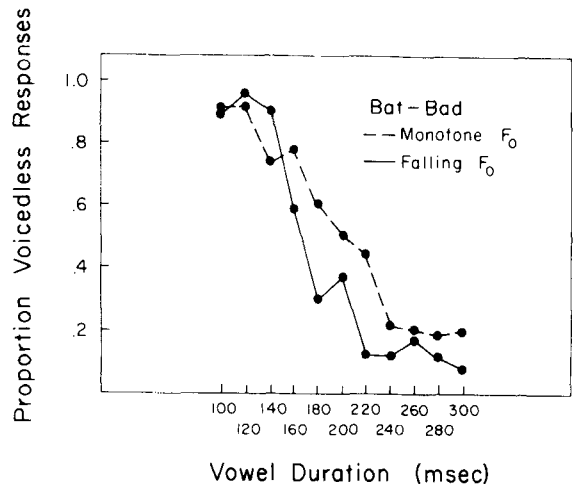


Figure 1. Proportion of "bat" responses as a function of vowel duration.

with that displayed in Figure 1, while the fourth showed little difference between the two sets of test stimuli.

For each subject, the category boundary for each test continuum was computed by linearly interpolating between the two stimuli that spanned the 50% point on the identification function. A *t* test for related measures on these category boundary values confirmed the earlier observation that perception changed from "bat" to "bad" at shorter vowel durations in the falling continuum than in the monotone continuum [$t(3) = 3.51$, $p < .05$]. Thus, the results of Experiment 1 successfully replicated Lehiste's earlier findings by showing that changes in F_0 contour can influence the perception of voicing in postvocalic stops. Experiment 2 was designed to extend these results to fricatives in postvocalic position.

EXPERIMENT 2

Method

Subjects. Twelve Indiana University students served as listeners in Experiment 2. All were right-handed, native speakers of English and were paid \$3.00 for their participation.

Stimuli. Two 10-item continua, each ranging from the word "cease" to the word "seize," were synthesized on the Klatt software synthesizer. For all stimuli, the initial and final fricatives were 155 and 170 msec in duration, respectively. The fricatives were synthesized using spectra containing peak center frequencies of 4,000 and 4,900 Hz and each with bandwidths of 1,000 Hz. In both continua, vowel duration varied from 50 to 350 msec in approximately equal logarithmic steps. The shortest duration was intended to produce a good exemplar of "cease" and longest, a good exemplar of "seize." The vowel was synthesized using formant frequencies of 280, 1,850, 2,900, 3,700, and 4,000 Hz. The two test continua were identical in all respects except for the F_0 contour. In the monotone condition, F_0 was held constant at 110 Hz throughout the vowel. In the falling condition, F_0 was initially set at 130 Hz and fell linearly to 90 Hz during the second half of the vowel.

Procedure. The procedure followed that of Experiment 1, except that the "cease-size" continua were used in place of the "bat-bad" continua. Each of the 10 stimuli in each continuum was presented once in each of 20 blocks. Five of the listeners heard the monotone continuum first and seven heard the falling continuum first.

Results and Discussion

The identification results are shown in Figure 2, which plots the proportion of "cease" responses as a function of vowel duration for the monotone continuum (dotted line) and the falling continuum (solid line). Each point is based on 20 observations by each of 12 subjects, for a total of 240 observations per point. The results show that perception changes from the voiceless "cease" to the voiced "seize" at shorter vowel durations in the falling continuum than in the monotone continuum, suggesting that F_0 contour affects perceived vowel duration. Of the 12 subjects, 9 showed a pattern of results consistent with the

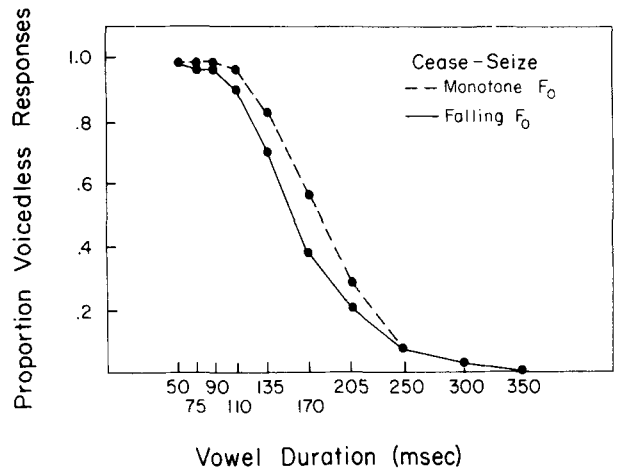


Figure 2. Proportion of "cease" responses as a function of vowel duration.

average data shown in Figure 2. Two subjects showed no difference between the two test continua, and only one showed a reversal of the pattern displayed in the figure. Category boundaries were computed as in Experiment 1. A *t* test confirmed that identification changed from "cease" to "seize" at shorter vowel durations in the falling F_0 continuum than in the monotone F_0 continuum [$t(11) = 5.08$, $p < .001$].

After Experiment 2 had been completed, a recent experiment by Derr and Massaro (1980) came to our attention. Derr and Massaro had subjects identify noun and verb forms of the word "use"—/jus/ and /juz/, respectively. These investigators compared monotone F_0 contours with both rising and falling contours and found that when F_0 was changing—either falling or rising—perception was biased toward /juz/. Taken together, the work of Derr and Massaro and our own findings provide strong support for the conclusion that a changing F_0 contour can serve as a voicing cue for postvocalic fricatives.

The results of Experiments 1 and 2, the work of Lehiste (Note 3), and the recent findings of Derr and Massaro (1980) all provide support for the hypothesis that F_0 contour can serve as a cue for postvocalic voicing, presumably through its mediating effect on the perceived duration of the preceding vowel. However, since both of the present experiments, as well as Lehiste's and Derr and Massaro's, used synthesized speech, they do not directly address the question of whether the F_0 cue is actually deployed systematically by listeners in producing these contrasts. Experiment 3 was therefore carried out to determine whether systematic differences in F_0 contour can be observed in the production of voicing contrasts in final position. The F_0 contours occurring before postvocalic consonants were analyzed for minimal pairs of words differing on the voicing of the final consonant.

EXPERIMENT 3

Method

Subjects. Eight laboratory personnel and psychology graduate students volunteered as subjects. All were male native speakers of English. Three had had some phonetic training. The speakers represented a cross section of American-English dialects.¹

Materials. The basic stimulus materials consisted of a set of 23 minimal pairs of words, which differed on the voicing of the final consonant. Sixteen of the pairs contained a final stop and seven pairs contained a final fricative. Several different lists were constructed by randomly intermixing these 46 words with 46 additional filler words. The 46 test words are shown in Table 1.

Procedure. Each speaker read through the word list once in citation form in a sound-attenuated room. Each pronunciation was recorded on audiotape with a high-quality microphone and then digitized via a 12-bit A-D converter for computer analysis. The F0 contours of Subjects J.S. and L.G. were analyzed using a computer algorithm, implemented at the Research Laboratory of Electronics at M.I.T., which is similar to the algorithm described by Gold and Rabiner (1969). In this algorithm, the speech signal is initially low-pass filtered to eliminate frequencies above the first formant region. The algorithm then determines the heights and positions of positive and negative peaks, as well as the difference in heights from peak to valley (negative peak), valley to peak, peak to previous peak, and valley to previous valley. Each of these measures is used to arrive at an independent measure of F0, and a complex decision rule is then used to determine a final "best" measure of F0 for any point in time. An autocorrelation routine, implemented in the Speech Perception Laboratory at Indiana University (Kewley-Port, 1979), was used to analyze the F0 contours of the remaining six speakers. An analysis of the production of J.S. using the autocorrelation routine in the Indiana lab showed no significant differences from the first analysis using the M.I.T. algorithm.

Table 1
Stimulus Items Used in Experiment 3

Voiced		Voiceless
Stops		
Bad		Bat
Bead		Beat
Bid		Bit
Bowed		Boat
Bud		But
Cued		Cute
Jog		Jock
Lube		Loop
Need		Neat
Pad		Pat
Pod		Pot
Rib		Rip
Rude		Root
Seed		Seat
Sued		Suit
Weed		Wheat
Fricatives		
Buzz		Bus
Eyes		Ice
Maize		Mace
Peas		Piece
Plays		Place
Rise		Rice
Seize		Cease

Results and Discussion

Table 2 shows the mean vowel duration (D), F0 at the start of the vowel (F0s), F0 at the end of the vowel (F0e), amount of change of F0 ($\Delta F0 = F0s - F0e$), and the rate of change of F0 (Hz/msec) for final fricatives and final stops for each of the eight speakers. The pair "bowed-boat" was not analyzed for speakers L.G., T.C., and T.F. since they pronounced "bowed" as /baud/ rather than /bod/. Mispronunciations, disfluencies, and hesitations also resulted in elimination of the pairs "seize-cess" for speaker T.H. and "jog-jock" for speaker D.B.

As expected, vowel duration was longer before voiced than voiceless consonants. This finding held true for all 23 minimal pairs for each speaker. The results of most interest concern the amount of change of F0 before voiced and voiceless final consonants. The F0 contour of speaker L.G. fell 9 Hz more before a final voiced stop than before a voiceless stop [$t(14) = 3.78, p < .01$]. Similarly, the F0 of speaker T.H. fell more before voiced than voiceless stops [$t(15) = 2.22, p < .05$]. Speaker A.W. showed a greater drop in F0 before voiced than voiceless fricatives [$t(6) = 3.76, p < .01$]. These results are consistent with the hypothesis that F0 contour is produced by talkers as a cue to final consonant voicing. However, the remainder of the production data are not consistent with this hypothesis. Speakers J.S., T.C., A.W., T.F., D.F., and D.B. showed no significant difference in the amount of change of F0 before voiced and voiceless stops. The same was true for the fricative data of speakers J.S., L.G., T.C., T.H., T.F., D.F., and D.B. These negative results are particularly striking given the consistent finding of much longer vowel durations before voiced than voiceless consonants, as shown in Table 2. Because of the longer durations of vowels occurring before voiced consonants, we might expect somewhat greater changes in F0 before voiced consonants, simply because of the longer duration available for those changes to occur. In short, the best conclusion that can be made from these analyses of the amount of change of F0 is that English speakers do not consistently produce greater amounts of change in F0 before voiced than voiceless consonants. While some subjects produce the differences reliably, the effects are not consistent across all subjects.

Two previous studies have examined F0 contours as a function of postvocalic consonantal voicing. Mohr (1971) used three speakers, none of whom were native speakers of English, and found, in agreement with the present results, that the F0 contour was not influenced by the voicing value of a postvocalic consonant. Similar results have been reported by Lea (1973), who used just two speakers and a corpus of test items that included many English nonwords.

Table 2
Vowel Duration (D), Starting F0 (F0s), Ending F0 (F0e), Change of F0 (Δ F0), and
Rate of Change of F0 (Hz/Msec) as a Function of Postvocalic Consonant Voicing

	D		F0s		F0e		Δ F0		Hz/Msec	
	Fricatives	Stops	Fricatives	Stops	Fricatives	Stops	Fricatives	Stops	Fricatives	Stops
Speaker: J.S.										
Voiced	256	200	95	96	76	79	19	17	.086	.084
Voiceless	142	115	98	98	78	80	20	18	.151	.167
Speaker: L.G.										
Voiced	189	200	120	124	87	92	33	32	.182	.166
Voiceless	91	89	120	119	89	96	31	23	.353	.287
Speaker: T.C.										
Voiced	278	230	132	156	114	132	18	24	.119	.110
Voiceless	143	113	137	160	116	140	21	20	.260	.192
Speaker: A.W.										
Voiced	237	208	117	115	86	88	31	27	.140	.135
Voiceless	119	114	122	122	100	100	22	22	.198	.200
Speaker: T.H.										
Voiced	325	240	149	148	122	123	27	25	.089	.103
Voiceless	167	120	146	149	122	130	24	19	.156	.182
Speaker: T.F.										
Voiced	236	202	108	109	91	95	17	14	.075	.073
Voiceless	143	114	104	108	92	94	12	14	.093	.129
Speaker: D.F.										
Voiced	314	219	107	106	70	74	37	32	.127	.151
Voiceless	155	109	108	111	76	77	32	34	.239	.332
Speaker: D.B.										
Voiced	344	246	134	133	92	94	42	39	.138	.168
Voiceless	140	107	125	138	93	102	32	36	.251	.358

Thus, it would appear that F0 is not systematically controlled in production as a function of the voicing of the final consonant.

GENERAL DISCUSSION

Taken together, the results of the present set of experiments produce a mixed picture of the relationship between F0 and voicing of postvocalic consonants. While the perceptual data from the present experiments clearly show a consistent effect of variations of F0 contour on perception of voicing for both stops and fricatives, the same consistent relationship was not observed in the analysis of F0 in speech production. Moreover, if we consider the combined effects of F0 and vowel duration, our results indicate that rate of change of F0 is slower before voiced than voiceless consonants, a result that is precisely the reverse of what had been observed in the earlier perceptual studies. Thus, while variation in the F0 contour may be capable of cuing the voicing of a postvocalic consonant, the generality of the specific effects as perceptual cues must be qualified since these differences are not reliably produced by all talkers in the same phonetic context.

Given that a particular speech cue has a perceptual effect, several hypotheses can be advanced concerning its underlying basis in speech production. First, its control in speech production may be mandated by mechanisms involved in speech articulation. For example, Haggard, Ambler, and Callow (1970) have shown that variations in the F0 contour of the initial portion of a consonant in a CV syllable is sufficient to cue the voicing of an initial consonant. An F0 contour that is high and falling during the early portion of the vowel indicates that the preceding consonant was voiceless, whereas a low and rising F0 indicates a voiced consonant. Studies of speech production have indicated that the differing F0 contours associated with voiced and voiceless consonants are due to physiological constraints inherent in articulation, particularly differences in subglottal pressure immediately after the release of the consonant (Mohr, 1971). That the perceptual system knows something about such physiologically determined cue is not surprising. The results of our production study and the earlier findings of Lea (1973) and Mohr (1971) all argue against the hypothesis that a similar phenomenon is responsible for the perceptual deployment of the F0 contour as a reliable cue to postvocalic

voicing in stops and fricatives. If inherent articulatory mechanisms dictated greater amounts of change in F0 before voiced than voiceless consonants, we would expect speakers to produce these larger changes, regardless of their native language or the specific corpus of test words examined. Since this effect does not occur strongly in our data, any account of the use of F0 contour as a cue to postvocalic consonantal voicing cannot have direct recourse to the systematic use of this distinction in speech production.

A second possibility is that the perceptual cue, while not mandated by physiological mechanisms inherent in articulation, has been used by the speakers of a particular language community, perhaps to serve as a redundant property for a particular phonetic contrast, or perhaps to replace an acoustic cue that is less salient perceptually or involves more articulatory effort. That is, the cue may be more language specific. In this case, we would expect the cue to be perceptually useful only to those speakers of languages that actually produce the cue and use it distinctively in their language. This second hypothesis also predicts that if the cue is used perceptually in a particular language, then native speakers of that language should also produce that cue reliably in order to distinguish minimal pairs of words. Since F0 contour is used by English listeners as a perceptual cue to postvocalic consonantal voicing, as evidenced by the results of Derr and Massaro (1980), Lehiste (Note 3), and Experiments 1 and 2, we would expect English speakers to produce this cue when speaking English. The production data of Lea (1973) and Mohr (1971) are not inconsistent with this prediction, since Lea's corpus included a large number of English nonwords and Mohr's study did not use native speakers of English. Experiment 3, which employed only native English speakers and a corpus consisting entirely of English words, failed to find evidence of a consistent effect of postvocalic consonantal voicing on speakers' productions of F0 contours and therefore argues against this hypothesis.

Finally, a third possibility is that a perceptual cue may be a manifestation of a more general psychophysical phenomenon. For example, at durations shorter than 200 msec, the auditory system can trade duration for intensity (e.g., Zwislocki, 1960). Thus, a "long" vowel could presumably be cued by increases in either duration or intensity or both. Similarly, Pisoni (1977) has recently suggested that the voice-onset-time cue to the voicing of word-initial stop consonants can be understood by reference to general auditory constraints on the ability to identify the relative onset times of two acoustic events. Thus, speakers of a particular language may or may not produce the cue, depending upon whether they are aware of its perceptual efficacy, whether their vocal tracts are capable of producing the cue, and the

amount of redundancy already present in the speech signal. In vision, the Oppel-Kundt illusion refers to the observation that filled extents appear longer than unfilled extents (Robinson, 1972). There is some evidence that a similar phenomenon may occur in audition. Temporal intervals filled with acoustic events are perceived as longer than empty temporal intervals (Ornstein, 1969). It must be pointed out that most of these studies used intervals ranging from thousands of milliseconds up to several minutes, and thus, their results are not directly generalizable to the much shorter durations of naturally produced vowels. Although it is difficult to specify exactly what an "event" is in a signal as complex as a speech waveform, it seems reasonable to suppose that a change in F0 is an event. If so, acoustic signals, including but not limited to speech signals, that change over time would be perceived as longer than signals that do not change over time. Changes in F0 of a vowel will therefore result in the perception of a longer vowel, which in turn cues a voiced, as opposed to a voiceless, postvocalic consonant.

The hypothesis that the perceived lengthening of vowel duration is a manifestation of a more general auditory phenomenon—in particular, that changing signals are perceived as longer than signals that are not changing—was first proposed and tested by Wang et al. (Note 2). In support of this hypothesis, Wang et al. found that a changing F0 increased the perceived duration of both speech and nonspeech sounds. Wang et al., however, also found that the monophthong /i/ was perceived to be equal in duration to the diphthong /ai/. This result appears to contradict the hypothesis. Because the diphthong is inherently a more complex vowel involving greater frequency change than the monophthong, the diphthong should be perceived as longer. A third result reported by Wang et al. might explain this contradictory result. These investigators observed that the vowel /i/ was perceived to be greater in duration than the vowel /a/. To explain this finding, Wang et al. noted that the intrinsic duration of /a/ is longer than the intrinsic duration of /i/. Thus, the listener expects /i/ to be shorter than /a/ and appears to compensate for this expectation. As a result, when /i/ and /a/ are of equal duration, the /i/ is actually perceived as longer. As Wang et al. note, the intrinsic duration of /ai/ is also greater than the intrinsic duration of /i/. Therefore, in the case of the /i/-/ai/ comparison, differences in intrinsic duration serve to increase the perceived duration of /i/, but differences in signal complexity serve to increase the perceived duration of /ai/. If these two effects are equal in magnitude, the net result is no difference in the perceived durations of /i/ and /ai/.

Most of the existing data on the influence of F0 contours on perceived duration are consistent with

the hypothesis that changing signals, in general, are perceived as longer than signals that are not changing. For example, Derr and Massaro (1980) found that both rising and falling contours biased the perception of postvocalic consonants toward the voiced alternative. Similarly, studies of perceived vowel duration have found that falling, rising, falling-rising, and rising-falling F0 contours all result in the perception of the vowel as longer than a comparable vowel with a monotone F0 (Lehiste, 1976; Pisoni, Note 1; Wang et al., Note 2). The important variable seems to be that F0 is changing; the direction of change and perhaps the amount of change seem unimportant (Lehiste, 1976).

In summary, changes in the F0 contour of a syllable appear to serve as an acoustic cue to the voicing of postvocalic stop and fricative consonants. However, this acoustic cue is not reliably produced by speakers, suggesting that its ability to serve as a perceptual cue is probably not tied to its regularity in speech production. Rather, the effect of the F0 contour to function as a perceptual cue may be the manifestation of a more general auditory phenomenon, important not only in speech perception, but also in other forms of auditory perception that involve complex multidimensional stimuli including speech.

REFERENCE NOTES

1. Pisoni, D. B. *Fundamental frequency and perceived vowel duration*. Paper presented at the 91st meeting of the Acoustical Society of America, Washington, D.C., April 1976.
2. Wang, W. S. -Y., Lehiste, I., Chang, C. -K., & Darnovsky, M. *Perception of vowel duration*. Paper presented at the 92nd meeting of the Acoustical Society of America, San Diego, November 1976.
3. Lehiste, I. *Contribution of pitch to the perception of segmental quality*. Paper presented at the 9th International Congress on Acoustics, Madrid, 1977.
4. Rosen, S. M. *Fundamental frequency patterns and the long-short vowel distinction in Swedish*. Speech Transmission Laboratory Quarterly Status and Progress Report 1/1977, 31-37.

REFERENCES

- DENES, P. Effects of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1954, 27, 761-764.

- DERR, M. A., & MASSARO, D. W. The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, 1980, 27, 51-59.
- GOLD, B., & RABINER, L. Parallel processing techniques for estimating pitch periods of speech in the time domain. *Journal of the Acoustical Society of America*, 1969, 46, 442-448.
- HAGGARD, M., AMBLER, S., & CALLOW, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 1970, 47, 613-617.
- KEWLEY-PORT, D. KLTEXC: Executive program to implement the KLATT software speech synthesizer. *Research on Speech Perception: Progress Report No. 4*. Department of Psychology, Indiana University, Bloomington, Ind., 1978.
- KEWLEY-PORT, D. SPECTRUM: A program for analyzing the spectral properties of speech. *Research on Speech Perception: Progress Report No. 5*. Department of Psychology, Indiana University, Bloomington, Ind., 1979.
- KLATT, D. H. Software for a cascade/parallel speech synthesizer. *Journal of the Acoustical Society of America*, 1980, 67, 971-995.
- LEA, W. A. Segmental and suprasegmental influences on fundamental frequency contours. In L. M. Hyman (Ed.), *Consonant types and tones*. Los Angeles: Linguistics Program, University of Southern California, 1973, 17-70.
- LEHISTE, I. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 1976, 4, 113-117.
- MOHR, B. Intrinsic variations in the speech signal. *Phonetica*, 1971, 23, 65-93.
- ORNSTEIN, R. E. *On the experience of time*. Baltimore, Md: Penguin Books, 1969.
- PISONI, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, 61, 1352-1361.
- ROBINSON, J. O. *The psychology of visual illusion*. London: Hutchinson, 1972.
- STEVENS, K. N. Effect of duration upon vowel identification. *Journal of the Acoustical Society of America*, 1959, 31, 109 (A).
- ZWISLOCKI, J. Theory of temporal auditory summation. *Journal of the Acoustical Society of America*, 1960, 32, 1046-1060.

NOTE

1. By American-English, we mean dialects of English spoken in the United States. Hereafter, English is meant to be interpreted as American-English.

(Received for publication April 10, 1980;
revision accepted August 27, 1980.)