*Perception & Psychophysics*
*1977, Vol. 22 (4), 373-382*

# Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction

DOMINIC W. MASSARO and MICHAEL M. COHEN
*University of Wisconsin, Madison, Wisconsin 53706*

The present series of experiments used factorial designs to evaluate which acoustic features are primarily responsible for the voicing distinction in the syllables /zi/ and /si/. Increases in frication duration tend to make the syllable more voiceless only if vocal cord vibration is absent or at a very low level during the frication period. Increasing the period between the onset of frication and the onset of vocal cord vibration changes the syllable from a predominantly voiced to a predominantly voiceless sound. This period, called voice onset time, can account for the change in perception regardless of simultaneous changes in the total frication duration or the relative duration of the frication period that contains vocal cord vibration. Changes in fundamental frequency had a large influence on the voicing judgments. With low fundamental frequencies, the judgments were predominantly voiced, whereas with high fundamental frequencies, voiceless judgments were predominant. The quantitative judgments of individual observers were described by a ratio-rule model that assumes a multiplicative combination of the independent cues, voice onset time and fundamental frequency. The model also provided a good description of previous studies of the acoustic cues used in the perception of voicing of fricatives.

The distinctive feature of voicing distinguishes between the respective members of a number of cognate pairs in English. The consonants /z/ and /s/ have the same place and manner of articulation but contrast in voicing: /z/ is voiced and /s/ is voiceless. Scully (1971) found that /s/ and /z/ were similarly articulated in the supraglottal vocal tract in terms of both the amount and timing of the tongue movements. The only obvious articulatory difference was that the vocal-cord vibration is not present during /s/, but occurs through nearly all of /z/. This paper continues previous work by Massaro and Cohen (1976) utilizing functional measurement techniques (Anderson, 1970, 1975) to evaluate what acoustic features are responsible for the voicing distinction in the syllables /zi/ and /si/.

Figure 1 presents spectrograms of the sounds /zi/ and /si/. Both sounds can be described by a period of frication followed by a transition to the steady state vowel. The primary difference between the two sounds appears to be the presence of vocal cord vibration in /zi/ but not in /si/. Vocal-cord vibration is seen in the spectrogram as the dark voicing bar at the lowest frequencies during the frication period. A second potential cue is the duration of the frication period. Although it is not shown in Figure 1, the average duration of the voiceless fricative /s/ is about 40% greater than the average duration of the voiced /z/ (Klatt, 1974, 1976; Umeda, 1977). In the first experiment, the intensity of vocal cord vibration is independently varied with the duration of the frication period in order to assess how these variables contribute to the voicing distinction between /si/ and /zi/.

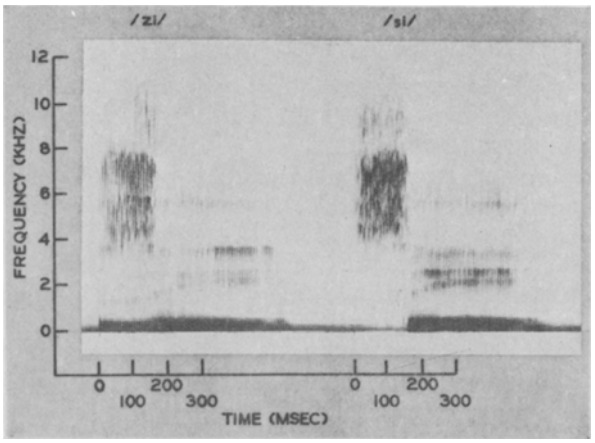


**Figure 1. Spectrograms of the syllables /zi/ and /si/ as spoken by an English talker.**

Another way of defining the cues of vocal-cord vibration and frication duration is in terms of voice onset time (VOT), the time between the onset of the syllable and the onset of vocal-cord vibration. With this description, the voiced fricative /z/ would have a short VOT and the voiceless fricative /s/ would have a long VOT. Experiment 2 tests the hypothesis that VOT can account for changes in both vocal-cord vibration and frication duration. In Experiment 3, this measure of VOT will be contrasted with mea-

This research was supported by Public Health Service Grant MH-19399. Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

suring VOT from the offset of the frication period. Finally, we have also observed that the fundamental frequency ($F_0$) of vocal-cord vibration is lower in /zi/ than in /si/. In Experiments 2 and 3, several values of $F_0$ were independently varied to evaluate its contribution to the voicing distinction between /zi/ and /si/.

Denes (1955) carried out an experiment to evaluate the contribution of vowel duration and frication duration in the perception of voicing in word-final position. The test alternatives were the two pronunciations of the homograph "use" as in the noun "the use" and the verb "to use." Four durations of the synthetically produced vowel were orthogonally varied with five durations of frication taken from real speech. No vocal-cord vibration was present during the frication period. The results from this experiment are presented in Figure 2. The results show that the proportion of voiceless responses decreased with increases in vowel duration and increased with increases in frication duration. Denes (1955) concluded that the ratio of the consonant-to-vowel duration was the critical determinant of perceived voicing. The final fricative should appear voiceless to the extent that this ratio is large. Inspection of the points in Figure 2, however, reveals that this was not the case. Although a one-to-one ratio gives 29% /s/ responses when the durations are 50 msec, the proportion of /s/ responses is 52% when the durations are 200 msec. Similarly, a two-to-one consonant vowel duration ratio of 100/50 gives 62% /s/ responses, whereas 96% /s/ responses were given at a 200/100 ratio. The results are more accurately described by a ratio-rule model that assumes a multiplicative combination of the two independent cues, frication duration and vowel duration. The model and the description of these results are presented in the General Discussion section.

Cole and Cooper (1975) varied the duration of frication without vocal-cord vibration in the fricative vowel syllables /sa/ and /fa/ and the affricate /cha/ with the duration of the vowel. The frication duration was varied by splicing out small segments of the frication before the vowel and then closing the gap. The duration of the vowel was shortened by simply removing the final part of the vowel. Frication duration had a large effect on the voicing judgments, whereas vowel duration had a negligible effect. In all cases, the syllable tended to be identified as the voiced cognate with decreases in frication duration. On the basis of these experiments, frication duration appears to cue the voicing of both word-initial and word-final fricatives, whereas vowel duration appears to cue voicing only when the fricative is in word (or syllable) final position.

Although frication duration functions as a cue to voicing, no direct evaluation has been made of its cue value as a function of vocal-cord vibration during the frication period. From informal observations, Denes argued that "it does not much matter whether the spectrum of the final consonant is harmonic or inharmonic" (Denes, 1955, p. 762). Accordingly, increases in the duration of frication should increase the likelihood of a voiceless judgment regardless of whether or not vocal-cord vibration occurs during the frication period. On the other hand, Massaro and Cohen (1976) showed that judgments of initial fricatives were more likely to be voiced to the extent that vocal-cord vibration occurred during the frication period. Given that Denes varied only the fricative duration without vocal-cord vibration and that Massaro and Cohen varied only the proportion of a fixed frication duration that contained vocal-cord vibration, it is necessary to vary both frication duration and intensity of vocal-cord vibration in the same experiment. In the first experiment, five levels of frication duration were orthogonally varied with five levels of the intensity of vocal-cord vibration during the frication period.

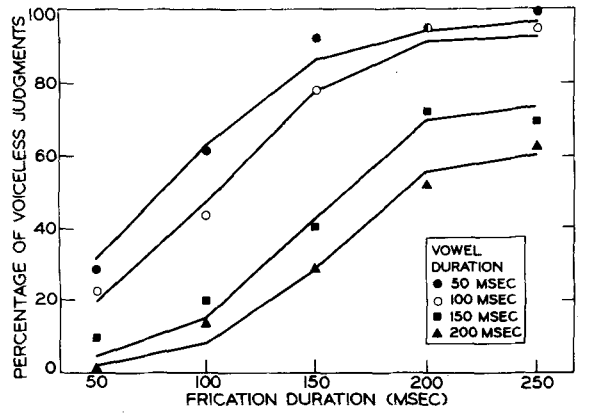


**Figure 2. Percentage of voiceless judgments of the final fricative in the word "use" as a function of vowel duration and the duration of frication. The observed values (points) are taken from Denes (1955). The lines give the predictions of a model assuming the multiplicative combinations of the independent cues, vowel duration, and frication duration, before a ratio rule is applied.**

## EXPERIMENT 1

### Method

**Subjects.** The subjects were seven undergraduates who were tested on 2 consecutive days. The subjects participated to fulfill an introductory psychology course requirement.

**Stimuli.** All stimuli were produced during the experiment proper by a formant series resonator speech synthesizer (FONEMA OVE-IIId) under the control of a PDP-8/L computer (Cohen & Massaro, 1976). Each stimulus was specified by a temporally ordered set of lists of synthesizer parameter control vectors. Within a list, each parameter vector specified a target value, a transition (arrival) time, and a transition type. Each list also specified how much time until the next list. All transitions were linear except the final fall in vowel amplitude, which was negatively decelerated. Time values were specified and parameters calculated in 10-msec increments. The speech synthesis was carried out in real-time during the experiment.

The speech sounds were fricative-vowel syllables. The syllables

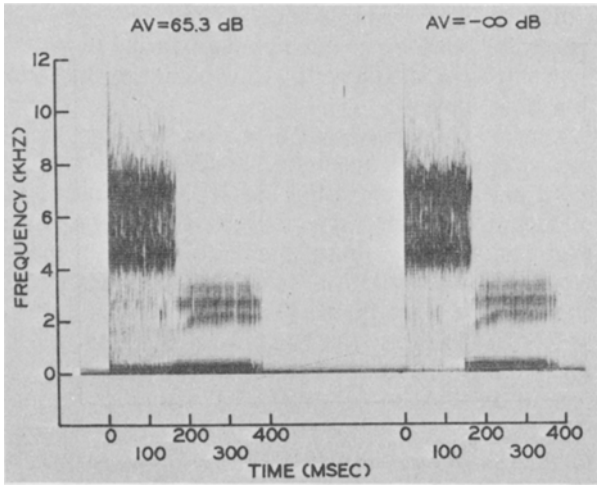


**Figure 3. Spectrograms of two of the synthesized syllables used in Experiment 1.**

can be represented by a frication period followed by a transition to the steady state vowel. For all of the syllables, the fricative formant frequencies were fixed at values that were appropriate for /z/ and /s/ and the vowel was /i/. Figure 3 presents spectrograms of two of the synthesized sounds differing during frication in the intensity level of the buzz source simulating vocal cord vibration. These sounds can be compared to natural sounds given in Figure 1. Figure 4 gives a schematic diagram of the syllables used in the experiment. The values AC and AV give the amplitude values for the noise and buzz sources, respectively. The K1 and K2 refer to the fricative formant frequencies. (Although it is not included in Figure 3, the amplitude of the fricative antiformant was fixed at 4 dB.) All syllables used in the experiments had this basic form.

The 25 test stimuli were generated by factorially combining five levels of frication duration with five levels of the amplitude of the buzz source (AV) during the frication period. The duration of the frication period was 60, 90, 120, 150, or 180 msec. The amplitude level of the buzz source, AV, was $-\infty$, 55.3, 58.6, 61.9, or 65.3 dB SPL (B) measured without any frication present. The $F_0$ value was always 163 Hz.

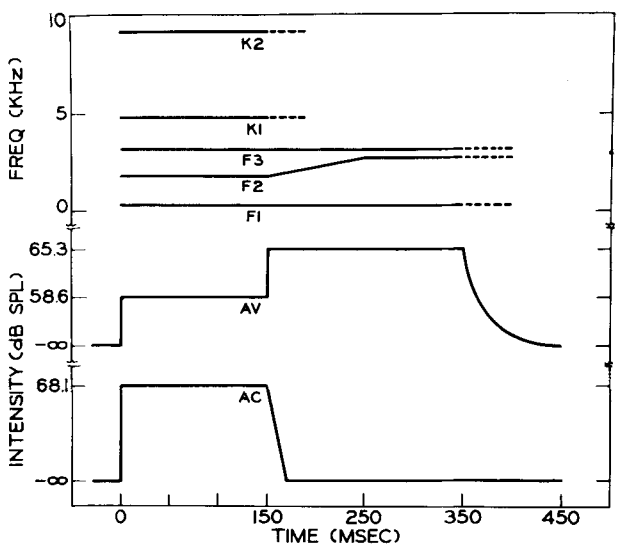


**Figure 4. Schematic diagram of the fricative formants (K1, K2), vowel formants (F1, F2, F3), and the buzz (AV) and noise (AC) sources for one of the syllables used in Experiment 1.**

**Procedure.** All experimental events were controlled by a PDP-8/L computer. The output of the speech synthesizer was amplified (McIntosh Model MC-50) and presented over Koss PRO-4AA headphones at a level of 68.1 to 69.5 dB SPL (B) during the period of frication depending on the amplitude of AV, and at 65.3 dB SPL (B) during the vowel /i/. The spectrograms in Figure 3 give some measure of the relative intensities of the noise and voicing harmonics in the stimuli. Four subjects could be tested simultaneously in individual sound-attenuated rooms.

Each trial began with the presentation of a syllable, selected randomly without replacement in blocks of 25 trials. Following presentation of the stimulus, each observer made a response by setting the point of a linear potentiometer, 5.5 cm long, the left end of which was labeled Z and the right end, C. When the subject was satisfied with the position of the pointer, he recorded the position by pressing a small button to the right of the scale. The potentiometer acted as a voltage divider; the resulting voltage was measured by a multiplexed A-D converter and scaled so that the resulting score varied on an interval from 0 to 49. When analyzing the data, this interval was normalized so that it varied between 0 and 1. The next trial began 1.0 sec after the last of the subjects made his or her response.

The subjects were asked to listen to the stimuli and to indicate where on a scale from C to Z each stimulus fell. The subjects were told that the leftmost scale setting represented a good prototype /zi/ whereas the rightmost setting, C, represented a good prototype /si/. The observers were warned not to try to discern any order to the pattern of presentation since the order of stimuli was strictly random. The subjects participated in two sessions on each of 2 consecutive days. In each session, 10 blocks of 25 trials each were presented. Each of the 25 stimuli was randomly presented once in a block of 25 trials. Unknown to the subject, the first two blocks of each session were not recorded. This gives a total of 32 observations per subject for each of the 25 stimuli. The mean response values for each of the 25 stimuli were calculated by the computer at the end of each experimental day. Before the first session of the first day, the subjects responded to 40 unscored practice trials.

## Results

Figure 5 plots the average ratings as a function of frication duration and the amplitude of the buzz source simulating vocal-cord vibration during the frication period. The syllables were judged as more /si/-like with increases in the frication duration, $F(4,24) = 38.95$, $p < .001$, and with decreases in the amplitude of buzz source during the frication period, $F(4,24) = 19.24$, $p < .001$. Figure 5 and the significant interaction between these two variables, $F(16,96) = 12.79$, $p < .001$, show that the increases in frication duration provide a strong cue to hearing the sound as voiceless only when there is a low level of vocal-cord vibration during the frication period. Increasing the amplitude of the buzz source to 65.3 dB completely neutralizes the cue value of frication duration.

## Discussion

The results of Experiment 1 show that the cue value of frication duration is critically dependent on the degree of vocal-cord vibration during the frication period. The absence of vocal-cord vibration during the frication period was probably a necessary condition for the finding of a large cue value for frication duration in the experiments of Cole and Cooper (1975) and Denes (1955). Increasing the frication

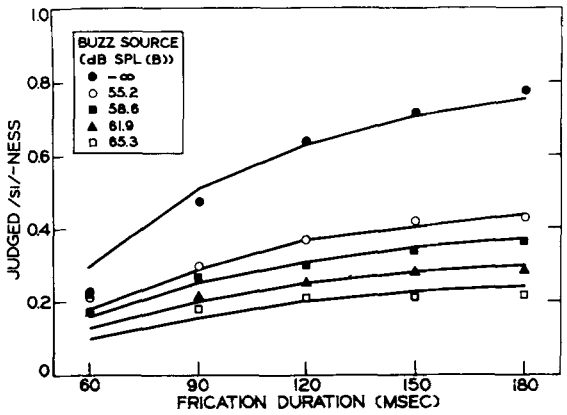


**Figure 5. Mean /si/-ness judgments (points) as a function of the duration of the frication period and the intensity of the buzz source simulating vocal cord vibration during the frication period. The lines give the predictions of Equation 4.**

period without vocal cord vibration increases the /si/-ness of the fricative, whereas increasing the frication duration with vocal-cord vibration has no effect. Accordingly, a more appropriate definition of this cue might be the time between the onset of the syllable and the onset of vocal-cord vibration. This dimension, called voice onset time (VOT), has been shown to cue voicing in the syllables /zi/ and /si/. Holding frication duration constant, Massaro and Cohen (1976) found that increasing VOT increased the likelihood of a /si/ judgment. The question that remains to be answered is to what extent voice onset time is a sufficient description of the cue value of vocal-cord vibration and frication duration. The next experiment therefore independently varied voice onset time and frication duration in order to determine whether voice onset time alone can account for the judgments or whether frication duration must also be incorporated into the judgmental process.

In addition to voice onset time and frication duration, the next experiment also manipulated the fundamental frequency ($F_0$) of vocal-cord vibration. The $F_0$ at the onset of vocal-cord vibration has a large potential cue value to the voicing of initial consonant sounds (Lehiste & Peterson, 1961). Lea (1972) made a series of measurements of segmental and suprasegmental influences on $F_0$ values. The effects of voicing on $F_0$ were evaluated for consonants (C) and vowels (V) in həCVĆ utterances spoken by two trained English talkers. The initial $F_0$ values in the stressed vowels were found to be about 20% higher when the preceding consonant was voiceless than when it was voiced. The $F_0$ contour during the first 100 msec of the vowel also differed as a function of the voicing of the preceding consonant. The contour rose following a voiced consonant and fell following a voiceless consonant. Both the initial $F_0$ values and the $F_0$ contours were reliable cues to the voicing of the preceding consonant. The $F_0$ values during the stressed vowel were *not* affected by the voicing of the following consonant.

Lea (1972) carried out a second study to include measurements of unstressed syllables. Two-syllable word pairs with contrasting stress patterns and similar phonemic sequences were spoken by the same talkers. For example, the noun and verb forms of *permit* would allow an assessment of the $F_0$ values in both stressed and unstressed syllables. In contrast to the həCVĆ utterances, the $F_0$ contour of the vowel did not reliably reflect the voicing of the preceding consonant in these word pairs. A rising $F_0$ contour indicated that the preceding consonant was voiced only 55% of the time. A falling $F_0$ contour was not reliable evidence of a preceding voiceless consonant, since this contour could occur in an unstressed syllable regardless of the voicing of the preceding consonant. Lea concluded that the $F_0$ contours were not direct cues to either stress or voicing of the preceding consonant. If the stress of a syllable could be directly determined independently of the $F_0$ contour, the contour could provide some information about the voicing of the preceding consonant. If a falling contour is observed in a stressed syllable, the preceding consonant must be voiceless, whereas a rising contour in an unstressed syllable is a reliable index that the preceding consonant is voiced.

Massaro and Cohen (1976) evaluated $F_0$ as a cue to the voicing distinction between /si/ and /zi/. The frequency of $F_0$ at the onset of voicing influenced the voicing judgments, whereas the $F_0$ frequency contour had very little effect beyond that accounted for by the frequency of $F_0$ at the onset of vocal-cord vibration. Observers tended to perceive the sounds as voiceless with increases in the frequency of $F_0$ at the onset of vocal-cord vibration. A control condition showed that the cue value of $F_0$ frequency was not due to the possibility of less energy at $F_1$ with increases in $F_0$ frequency.

Massaro and Cohen (1976) independently varied voice onset time and the frequency of $F_0$ at the onset of vocal-cord vibration in order to assess the relative cue value of both of these variables. Both cues were critical for the voicing judgment, and the results showed that the judgment of a particular stimulus was primarily a function of the least ambiguous cue. A formal model based on this assumption provided a good description of the quantitative voicing judgments. The next experiment extends the previous studies by varying frication duration, voice onset time, and $F_0$ frequency. The experiment also allows a direct evaluation of the relative importance of each of these cues in determining the voicing judgment.

## EXPERIMENT 2

### Method

**Subjects.** Six volunteers from an introductory psychology class participated on 2 consecutive days.

**Stimuli.** The stimuli had the same basic form as those in Experiment 1. Four levels of frication duration (FD), four levels of voice onset time (VOT), and four levels of the frequency of $F_0$ were factorially combined, giving a total of 64 stimuli. The FD values were 120, 150, 180, and 210 msec. The VOT values were 30, 60, 90, and 120 msec. The frequency values of $F_0$ were 163, 183, 206, and 224 Hz. The buzz source was off during the VOT interval and was set to about 64 dB SPL (B) at the onset of vocal-cord vibration. In each session, five blocks of 64 trials were sampled randomly without replacement. Unknown to the subject, the first block of each of the four sessions was not recorded, giving a total of 16 observations per subject for each of the 64 stimuli. All other procedural details were the same as in Experiment 1.

### Results

Although the mean /si/-ness values decreased steadily from .53 to .45 with increases in overall frication duration, this effect was not statistically significant, $F(3,15) = 2.18$, $p > .10$. Figure 6 presents the mean /si/-ness values as a function of VOT measured from the onset of the syllable and the $F_0$ frequency value of vocal-cord vibration. The sounds were judged as more /si/-like with increases in VOT, $F(3,15) = 17.08$, $p < .001$, and with increases in $F_0$ frequency, $F(3,15) = 91.39$, $p < .001$. The mean /si/-ness values increased from .40 to .59 with increases in VOT and from .30 to .69 with increases in $F_0$. Although the interaction of VOT and $F_0$ was not significant, $F(9,45) = 1.65$, $p > .1$, $F_0$ frequency had a larger effect at the middle VOT values; similarly, VOT had a larger effect at the middle $F_0$ frequency values.

### Discussion

In the range of the values tested here, changes in VOT provide larger changes in the perception of voicing than do changes in frication duration. In fact, the small changes in voicing perception with increases in frication duration were actually in the opposite direction from what one would expect if increases in frication duration were a cue to an unvoiced sound. This result agrees with the finding in Experiment 1 that frication duration was not an effective cue if vocal-cord vibration was present during the frication period. For a given VOT in Experiment 2, increases in frication duration resulted in increases in the period of frication with vocal-cord vibration. Accordingly, this interval was not critical for the judgments. The VOT interval, the period of frication without vocal-cord vibration, is an effective cue for the perception of voicing. Finally, the results replicated the previous findings that the frequency of $F_0$ at the onset of vocal cord vibration is another important cue for the perception of voicing in these fricatives.

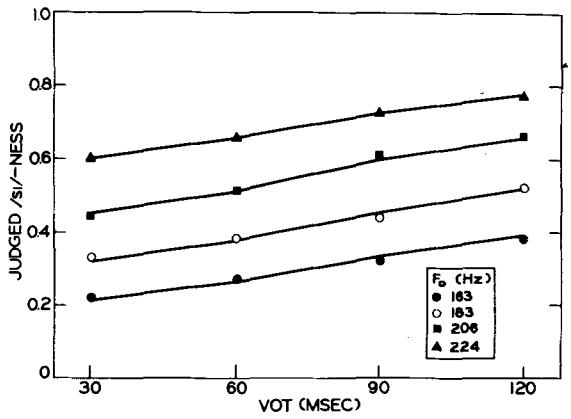


**Figure 6. Mean /si/-ness judgments (points) as a function of VOT (measured from the onset of the syllable) and the $F_0$ of vocal-cord vibration. The lines give the predictions of Equation 1.**

## EXPERIMENT 3

Our analysis indicates that frication duration without vocal-cord vibration is one critical cue to the perception of voicing of initial fricatives. Given that the VOT dimension measures this period, it can account for the changes in the perception of voicing independent of changes in total frication duration. The results support the usefulness of the measure of VOT taken from the onset of the frication period. Another measure that might seem plausible, is to define VOT from the offset of the frication period. In this case, VOT measures the duration of the frication period with vocal-cord vibration. Given that Experiments 1 and 2 have shown that the frication period with vocal-cord vibration is not an effective cue to voicing, we would expect that it would not be a sufficient description of changes in the perception of voicing in these syllables. To test this idea, we replicated Experiment 2 exactly, but now we measured VOT from the offset rather than the onset of the frication period. We predict that frication duration will have a large effect on the voicing judgments, since, for a given VOT interval, increases in frication duration are equivalent to increases in the frication period without vocal-cord vibration. The effect of VOT measured from frication offset should still be significant, however, since this period is confounded with VOT measured from frication onset. The goal of this experiment is to show that while VOT measured from the onset of the frication period accounts for the major changes in voicing judgments regardless of total frication duration, this is not the case when VOT is measured from the offset of the frication period.

### Method

**Subjects.** Six volunteers from an introductory psychology course participated for 2 days.

**Stimuli and Procedure.** The stimuli were identical to those in

Experiment 2 except that VOT was measured from the offset of the frication period rather than from its onset. The procedure was identical to that of Experiment 2.

## Results

Figure 7 plots the mean /si/-ness values as a function of frication duration, $F_0$, and the VOT measured from the offset of the frication period. The syllables were judged as more /si/-like with increases in frication duration, $F(3,15) = 21.45$, $p < .001$. The judgments increased from .37 to .60 with increases in frication duration from 120 to 210 msec. The judgments of the syllables were more /si/-like with decreases in VOT, $F(3,15) = 28.40$, $p < .001$. The mean response decreased from .80 with a VOT of 30 msec to .36 with a VOT of 120 msec. The syllables were judged to be more /si/-like with increases in $F_0$, $F(3,15) = 8.81$, $p < .005$. The average response was .39 with a $F_0$ value of 163 Hz and .59 with a $F_0$ value of 224 Hz. The higher order interactions primarily reflected the fact that the magnitude of the effect of one variable was largest at the intermediate levels of the other variables, those levels that gave the most ambiguous judgments.

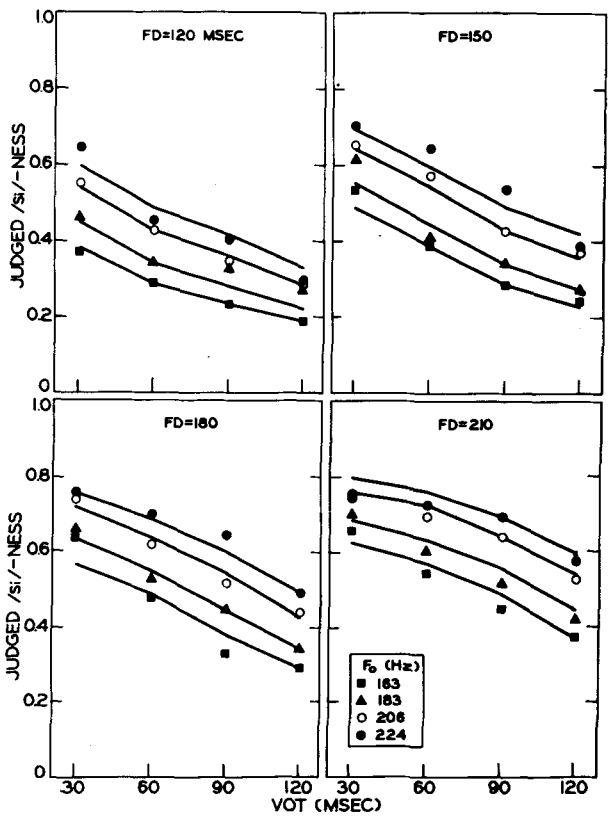


**Figure 7. Mean /si/-ness judgments (points) as a function of frication duration, VOT (measured from the offset of the frication period), and the $F_0$ of vocal-cord vibration. The lines give the predictions of Equation 1 with VOT measured from the onset of vocal-cord vibration.**

## Discussion

The results confirm our predictions that VOT measured from the offset of the frication period is not a sufficient description of the changes in the voicing judgments with changes in the total frication duration and the duration of the frication period that is voiced. Accordingly, the frication period without vocal-cord vibration provides the most complete and parsimonious description of changes in the voicing judgments with independent changes in total frication duration and the duration of the frication period that is voiced.

# GENERAL DISCUSSION

Massaro and Cohen (1976) described the voicing judgments of these fricatives in the framework of a general model of speech perception. The model assumes that the sound-wave pattern corresponding to a syllable is transduced by the auditory receptor system and feature detection process into a set of acoustic features in preperceptual auditory storage (PAS). These features are combined into a percept by the primary recognition process. The percept corresponds to the phenomenological experience of hearing a particular sound at some location in space. In Massaro and Cohen's study, the fundamental frequency ($F_0$) was orthogonally varied with voice onset time (VOT) measured from the onset of the frication period. It was assumed that the acoustic correlates of $F_0$ frequency and VOT are detected as acoustic features and stored in PAS. The Detection process gives for each feature a value between 0 and 1 which indexes the degree to which that feature matches the ideal feature for each alternative.

Given two mutually exclusive alternatives such as /si/ and /zi/, it is assumed that the outcome of a feature comparison on one alternative is the complement of the same feature comparison on the other alternative. If V is the outcome of a VOT comparison, then a perfect match on the VOT comparison with /si/ would give $V = 1$, and the comparison of the same feature with /zi/ would give $V = 0$. If F is the outcome of a $F_0$ comparison and equals .5 for one alternative, it will also equal .5 for the other. In the general case, it is assumed that V and F index the degree to which the respective features of VOT and $F_0$ match with /si/, whereas $(1 - V)$ and $(1 - F)$ index the degree to which they match with /zi/.

The perceived /si/-ness of a particular syllable is assumed to be equal to a multiplicative combination of the /si/-ness of each of the component features. Given the features of VOT and $F_0$, the /si/-ness of a particular sound would be VF, whereas $(1 - V)(1 - F)$ would represent the degree of /zi/-ness of that same sound. The judged /si/-ness, J(C), of the same sound is based on the perceived degree of /si/-ness relative to the sum of the perceived degree of /si/-ness and

**Table 1**
**The Parameter Values Given by Equation 2 for the Four Levels of $F_O$, VOT*, and FD for Each of the Six Subjects in Experiment 2**

| | $F_O$ (Hz) | | | | VOT (msec) | | | | FD (msec) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | 163 | 183 | 206 | 224 | 30 | 60 | 90 | 120 | 120 | 150 | 180 | 210 |
| 1 | .33 | .40 | .58 | .69 | .42 | .45 | .52 | .61 | .48 | .57 | .53 | .53 |
| 2 | .27 | .43 | .61 | .73 | .38 | .45 | .57 | .60 | .54 | .56 | .52 | .48 |
| 3 | .24 | .44 | .56 | .69 | .45 | .47 | .50 | .56 | .44 | .39 | .40 | .38 |
| 4 | .34 | .45 | .56 | .64 | .29 | .44 | .57 | .67 | .51 | .46 | .43 | .37 |
| 5 | .31 | .43 | .55 | .72 | .35 | .42 | .59 | .64 | .71 | .59 | .51 | .33 |
| 6 | .28 | .39 | .55 | .80 | .46 | .50 | .50 | .55 | .51 | .48 | .53 | .53 |
| Average | .30 | .42 | .57 | .71 | .39 | .46 | .54 | .61 | .53 | .51 | .49 | .44 |

**VOT is measured from the onset of the frication period.*

the perceived degree of /zi/-ness. This judgment process follows Luce's (1959) ratio-rule model. The judged /si/-ness, J(C), of the sound is equal to the ratio of perceived /si/-ness to perceived /si/-ness plus perceived /zi/-ness.

$$J(C) = \frac{VF}{VF + (1-V)(1-F)}. \quad (1)$$

The primary feature of this model is that the final judgment gives more weight to less ambiguous featured in the syllable. For example, if V = .5 and F = .8, J(C) would be equal to .86. A simple adding of the acoustic features before the ratio rule is applied would give J(C) = .7 (cf. Massaro & Cohen, 1976). The quantitative description of the multiplicative model was about a factor of five times better than the description given by the comparable adding model.

The multiplicative model will be employed here to assess the degree to which VOT measured from the onset of the frication period is a sufficient description of the total frication period and the duration of the period that is voiced. The parameter estimates from the model can be used to index the relative importance of each of the cues in the voicing judgments. A feature is important in the judgment process to the extent that the parameter value is lower or higher than .5. In Experiment 2, VOT, $F_0$, and FD were independently varied in a factorial design, producing a total of 64 syllables. Assuming that each of these dimensions corresponds to an acoustic feature, the predictions of the model follow Equation 2,

$$J(C) = \frac{VFD}{VFD + (1-V)(1-F)(1-D)}, \quad (2)$$

where V, F, and D represent the /si/-ness values of VOT, $F_0$, and FD, respectively.

This model was fit to the individual judgments of each of the six subjects in Experiment 2. For each subject, the parameter values for V, F, and D were estimated by minimizing the squared deviations between the predicted and observed judgments using the iterative minimization routine STEPIT (Chandler, 1969). Table 1 presents the parameter values for the description given by Equation 2. The parameter values make apparent the large contribution of $F_0$ and VOT to the voicing judgments and the small contribution of FD. The parameter values indexing the degree of /si/-ness increased from .30 to .71 with increases in $F_0$ and from .39 to .61 with increases in VOT, whereas increases in FD decreased the degree of /si/-ness from .53 to .44. These values show that FD contributed very little to the judgment process relative to the contributions of $F_0$ and VOT.

In order to test whether eliminating the contribution of FD in the model would significantly reduce the good description of the judgments, these same data were described by Equation 1. Table 2 contrasts the descriptions of the two models in terms of the goodness-of-fit criterion, the root mean square. The root mean square is the square root of the average of the sum of the squared deviations between the predicted and observed results. Although the predictions given by Equation 2 provide a somewhat better fit, it does not seem to be significantly better considering the fact that an additional four parameters were estimated in Equation 2 relative to Equation 1. The reasonably good description of the 64 data points with just eight parameters offers strong support for the model defined by Equation 1. Figure 6 gives the average predictions of Equation 1 for the average results of the six subjects.

**Table 2**
**The Root Mean Square Deviations Between Observed and Predicted Values Given by Equations 1 and 2 for Each of the Six Subjects in Experiment 2**

| Subject | Equation 1 | Equation 2 |
|---|---|---|
| 1 | .075 | .070 |
| 2 | .058 | .055 |
| 3 | .060 | .056 |
| 4 | .076 | .062 |
| 5 | .147 | .085 |
| 6 | .081 | .079 |
| Average | .088 | .069 |

Table 3
Parameter Values Given by Equation 2 for the Four Levels of $F_O$, VOT*, and FD for Each of the Six Subjects in Experiment 3

| | $F_O$ (Hz) | | | | VOT (msec) | | | | FD (msec) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | 163 | 183 | 206 | 224 | 30 | 60 | 90 | 120 | 120 | 150 | 180 | 210 |
| 1 | .44 | .51 | .52 | .53 | .59 | .48 | .46 | .46 | .36 | .39 | .44 | .47 |
| 2 | .24 | .37 | .61 | .74 | .60 | .52 | .49 | .39 | .34 | .43 | .51 | .58 |
| 3 | .46 | .46 | .52 | .56 | .73 | .54 | .42 | .28 | .21 | .39 | .52 | .63 |
| 4 | .38 | .46 | .55 | .61 | .72 | .58 | .44 | .30 | .35 | .52 | .63 | .75 |
| 5 | .32 | .45 | .57 | .65 | .73 | .54 | .42 | .31 | .29 | .45 | .55 | .63 |
| 6 | .46 | .47 | .53 | .54 | .63 | .57 | .44 | .38 | .57 | .58 | .61 | .65 |
| Average | .38 | .45 | .55 | .61 | .67 | .54 | .45 | .35 | .35 | .46 | .54 | .62 |

**VOT is measured from the offset of the frication period.*

Equations 1 and 2 were also applied to the individual judgments of Experiment 3. The significant difference from the preceding analysis is that VOT is measured from the offset of the frication period. Table 3 presents the parameter values given by Equation 2 for $F_0$, VOT, and FD for each of the six subjects. In contrast to Experiment 2, the parameter values for all three dimensions show large changes with changes in the dimension values. Increases in $F_0$ and FD increased the parameter values indexing the degree of /si/-ness from .38 to .60 and from .35 to .62, respectively. Increases in VOT decreased the /si/-ness values from .67 to .35. In agreement with our earlier analysis, the large change in the parameter values of FD shows that VOT measured from the offset of $F_0$ is not sufficient to account for the influence of total frication duration. This analysis is substantiated by the very poor description of the data by Equation 1 when VOT is measured from the offset of the frication period (cf. Table 4). Including the FD dimension in the description of the judgments in Equation 2 provides an adequate account of the data. Table 4 shows that the description given by Equation 2 decreases by one-half the root mean square deviation of the description by Equation 1.

We predict that an adequate description of the judgments in Experiment 2 can be given by Equation 1, however, if VOT is measured from the onset rather than the offset of the frication period. When

Table 4
The Root Mean Square Deviations Between Observed and Predicted Values Given by Equations 1 and 2 for VOT Measured from Offset of Frication and by Equation 1 for VOT Measured from Onset of Frication for Each of the Six Subjects in Experiment 3

| Subject | Equation 1 VOT from Offset | Equation 2 VOT from Offset | Equation 1 VOT from Onset |
|---|---|---|---|
| 1 | .077 | .064 | .067 |
| 2 | .088 | .045 | .046 |
| 3 | .148 | .040 | .041 |
| 4 | .141 | .056 | .052 |
| 5 | .125 | .053 | .059 |
| 6 | .073 | .066 | .077 |
| Average | .113 | .055 | .058 |

VOT is measured from the onset of the frication period, seven levels of VOT occurred in the experiment. The goodness of fit of this model is also presented in Table 4, and Table 5 gives the parameter values for $F_0$ and VOT. The predictions of Equation 1 for VOT measured from frication onset are shown in Figure 7. The results show that VOT measured from the onset of the frication period provides a good description of the judgments, without the necessity of incorporating the contribution of FD. This analysis supports our prediction that VOT measured from the onset of the frication period accounts for both the total frication duration and the duration of the frication period that is voiced.

Table 5
The Parameter Values Given by Equation 1 for the Seven Levels of VOT* and the Four Levels of $F_O$ for Each of the Six Subjects in Experiment 3

| | $F_O$ (Hz) | | | | VOT (msec) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | 163 | 183 | 206 | 224 | 0 | 30 | 60 | 90 | 120 | 150 | 180 |
| 1 | .44 | .51 | .52 | .53 | .37 | .34 | .34 | .41 | .44 | .51 | .54 |
| 2 | .24 | .37 | .61 | .74 | .23 | .31 | .39 | .48 | .54 | .60 | .63 |
| 3 | .46 | .46 | .52 | .56 | .09 | .22 | .27 | .41 | .61 | .71 | .79 |
| 4 | .38 | .46 | .55 | .61 | .21 | .31 | .41 | .56 | .75 | .80 | .85 |
| 5 | .33 | .45 | .58 | .65 | .12 | .26 | .32 | .50 | .61 | .70 | .80 |
| 6 | .46 | .47 | .53 | .54 | .46 | .45 | .57 | .60 | .63 | .75 | .74 |
| Average | .39 | .45 | .55 | .61 | .25 | .32 | .38 | .49 | .60 | .68 | .73 |

**VOT is measured from the onset of the frication period.*

Can the current model be applied to the results of Experiment 1 which independently varied frication duration (FD) and the amplitude of vocal-cord vibration (AV) during the frication period? At first glance, it could be assumed that FD and AV are independent acoustic features and that these are combined multiplicatively and then evaluated by the ratio rule. However, we believe that FD and AV are *not* independent acoustic features. The contribution of FD is critically dependent on the amplitude of voicing AV. With low amplitudes of AV, the frication is essentially unvoiced and increasing its duration is comparable to increasing VOT. Therefore, increases in FD should increase the /si/-ness of the judgments. With high amplitudes of AV, the frication is essentially voiced and increases in its duration should have very little effect on the /si/-ness judgments (cf. Experiment 2). Given that the degree of voicing during frication will be a direct function of the amplitude of AV, it can be assumed that AV serves as a multiplicative weight for the cue value of FD. With low values of AV, FD is given a large weight and increases in FD increase the /si/-ness of the judgments. With high amplitudes of AV, however, FD is given a very small weight so that increases in FD make very little difference in the amount of /si/-ness. Following this logic, $D_A$, the cue value of FD at a particular value of AV, can be described by

$$D_A = AD, \tag{3}$$

where A and D are the parameter values for AV and FD, respectively. Applying the ratio rule gives the following predictions for the judgments:

$$J(C) = \frac{D_A}{D_A + (1 - D_A)} = D_A, \tag{4}$$

so that the judgments are given directly by Equation 3. The judgments of the individual observers were fit by Equation 4, and the parameter values and goodness-of-fit values are given in Tables 6 and 7, respectively. For comparison, the same judgments were fit with the assumption that FD and AV are independent acoustic features.

**Table 6**
**The Parameter Values Given by Equation 4 for the Five Values of FD (Milliseconds) and the Five Values of AV [Decibels SPL (B)] in Experiment 1**

| | Subject | | | | | | | Aver- |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | age |
| FD | | | | | | | | |
| 60 | .38 | .31 | .48 | .26 | .26 | .32 | .32 | .33 |
| 90 | .47 | .39 | .84 | .42 | .37 | .45 | .88 | .55 |
| 120 | .59 | .56 | .97 | .56 | .60 | .56 | .97 | .69 |
| 150 | .63 | .68 | 1.00 | .66 | .78 | .68 | .98 | .77 |
| 180 | .70 | .75 | .96 | .77 | .86 | .79 | .97 | .83 |
| AV | | | | | | | | |
| $-\infty$ | .91 | .83 | .92 | .81 | .95 | .95 | 1.00 | .91 |
| 55.2 | .47 | .45 | .80 | .49 | .59 | .61 | .31 | .53 |
| 58.6 | .45 | .42 | .76 | .41 | .45 | .45 | .21 | .45 |
| 61.9 | .41 | .39 | .48 | .44 | .32 | .29 | .25 | .37 |
| 65.3 | .41 | .40 | .32 | .36 | .19 | .16 | .25 | .30 |

**Table 7**
**The Root Mean Square Deviations Between Observed and Predicted Values Given by Equations 4 and 5 for Each of the Seven Subjects in Experiment 1**

| Subject | Equation 4 | Equation 5 |
|---|---|---|
| 1 | .054 | .064 |
| 2 | .052 | .063 |
| 3 | .024 | .042 |
| 4 | .032 | .040 |
| 5 | .050 | .072 |
| 6 | .023 | .014 |
| 7 | .049 | .071 |
| Average | .042 | .056 |

Table 8 gives the parameter values, and the goodness of fit values are shown in Table 7. Contrasting the two models, Equation 3 allows a better description of the judgments than the description given by Equation 5.

The present model can also be adapted to describe the contribution of vowel duration and frication duration in the perception of word-final consonants. In the Denes (1955) study, subjects judged whether the final fricative in "use" was voiced as in the verb "to use" or voiceless as in the noun "the use." Given that the frication duration did not contain vocal-cord vibration, it corresponds to our measure of voice onset time (VOT). We believe that it is this measure of VOT rather than frication duration itself that is the critical dimension influencing the voicing judgments. Given that the order of the consonant and vowel are reversed in final relative to initial fricatives, measuring VOT from the offset of the frication period in final fricatives is completely consistent with our VOT measure taken from the onset of the frication period in initial fricatives. We assume that vowel duration (D) and VOT are perceived independently and combined multiplicatively before the ratio rule is applied. In this case, the proportion of /si/ responses, P(C), would equal

$$P(C) = \frac{VD}{VD + (1 - V)(1 - D)}, \tag{6}$$

where V and D are the /si/-ness values given the VOT and vowel duration, respectively. The predictions of this equation are given by the lines in Figure 1. The estimated parameter values for V were .07, .22,

**Table 8**
**The Parameter Values Given by Equation 5 for the Five Values of FD (Milliseconds) and the Five Values of AV [Decibels SPL (B)] for Each of the Seven Subjects in Experiment 1**

| | Subject | | | | | | | Aver- |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | age |
| FD | | | | | | | | |
| 60 | .20 | .16 | .26 | .12 | .11 | .14 | .16 | .16 |
| 90 | .25 | .21 | .54 | .20 | .18 | .21 | .37 | .28 |
| 120 | .31 | .29 | .65 | .27 | .31 | .27 | .41 | .36 |
| 150 | .33 | .34 | .68 | .33 | .42 | .35 | .45 | .41 |
| 180 | .37 | .38 | .64 | .37 | .46 | .43 | .44 | .44 |
| AV | | | | | | | | |
| $-\infty$ | .70 | .67 | .75 | .69 | .75 | .77 | .96 | .76 |
| 55.2 | .46 | .46 | .64 | .50 | .55 | .57 | .36 | .51 |
| 58.6 | .45 | .44 | .61 | .44 | .44 | .46 | .25 | .44 |
| 61.9 | .42 | .41 | .34 | .46 | .32 | .32 | .30 | .37 |
| 65.3 | .42 | .42 | .20 | .40 | .20 | .19 | .29 | .30 |

.54, .78, and .82 for VOTs of 50, 100, 150, 200, and 250 msec, respectively. The parameter values for D were .86, .76, .39, and .26 for vowel durations of 50, 100, 150, and 200 msec, respectively. The parameter values confirm the observation that increasing the duration of the vowel tends to make the sound more voiced, whereas increasing the VOT tends to make the sound more voiceless. The good fit of the 20 observations with nine parameter values supports the assumption of the multiplicative combination of these two cues. In contrast to this model, quantification of an analogous model based on the ratio of consonant-to-vowel durations gives an unacceptable description, even though 15 parameters were estimated from the data. The root mean square (rms) was .033 for the multiplicative model and .067 for the model based on the ratio of consonant-to-vowel durations. The data were also described by an additive combination of cues before the ratio rule was applied. The rms for the additive model was .080, a description that is about 2½ times poorer than the multiplicative model.

## CONCLUSION

The current experiments and theoretical analyses have illuminated the manner in which acoustic features are combined in speech perception. The acoustic features of fundamental frequency and voice onset time measured from the onset of frication were the critical determinants of the perception of voicing difference between /si/ and /zi/. Frication duration has very little effect beyond that accounted for by voice onset time. The quantitative results of individual judgments were described by a model that assumes that the features are perceived independently and combined multiplicatively in order to obtain the overall voicing quality of the sound. The judgment is determined by a ratio rule based on perceived degree of /si/-ness relative to the sum of the perceived degree of /si/-ness and the perceived degree of /zi/-ness. The experimental and theoretical procedures offer a high potential for substantive analyses of how acoustic features are utilized in speech perception.

## REFERENCES


ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153-170.

ANDERSON, N. H. On the role of context effects in psychophysical judgment. *Psychological Review*, 1975, **82**, 462-482.

CHANDLER, J. P. Subroutine STEPIT finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, **14**, 81-82.

COHEN, M. M., & MASSARO, D. W. Real-time speech synthesis. *Behavior Research Methods & Instrumentation*, 1976, **8**, 189-196.

COLE, R. A., & COOPER, W. E. Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 1975, **58**, 1280-1287.

DENES, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, **27**, 761-764.

KLATT, D. H. The duration of [s] in English words. *Journal of Speech and Hearing Research*, 1974, **17**, 51-63.

KLATT, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, **59**, 1208-1221.

LEA, W. A. *Intonational cues to the constituent structure and phonemics of spoken English*. Unpublished PhD dissertation, Purdue University, 1972.

LEHISTE, I., & PETERSON, G. E. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, 1961, **33**, 419-425.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

MASSARO, D. W., & COHEN, M. M. The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America*, 1976, **60**, 704-717.

SCULLY, C. A comparison of /s/ and /z/ for an English speaker. *Language and Speech*, 1971, **14**, 187-200.

UMEDA, N. Consonant duration in American English. *Journal of the Acoustical Society of America*, 1977, **61**, 846-858.


(Received for publication March 16, 1977;
revision accepted June 24, 1977.)