Perception & Psychophysics
1980, Vol. 27 (1), 51-59

The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction

MARCIA A. DERR and DOMINIC W. MASSARO
University of Wisconsin, Madison, Wisconsin, 53706

A model was quantified to describe the integration of vowel duration, fricative duration, and fundamental frequency (F_0) contour as cues to final position fricatives differing in voicing. The basic assumptions are that perceived vowel duration and perceived frication duration are cues to the identity of final position fricatives and that both F_0 contour and vowel duration influence perceived vowel duration. Binary choice and rating responses to synthetic stimuli varying independently along the three dimensions were collected. The results were consistent with the assumption that F_0 contour operates by modifying perceived vowel duration, which is a direct cue. Unfortunately, the nature of the modification appears to be very similar in form to that which results from the integration of two independent cues in syllable identification. Therefore, the results do not allow a rejection of the idea that the perception of F_0 contour may directly cue the identity of final position fricatives.

Much of the research in speech perception has centered on the problem of determining the acoustic features that characterize certain classes of speech sounds. Given a short segment of speech, a vowel, consonant-vowel, or vowel-consonant syllable, for example, what are the acoustic cues which distinguish that sound from all other possibilities in the language? Determining the featural cues in speech, however, implicates several more basic questions. How are separate acoustic features evaluated and put together or integrated to form a single synthesized percept?

In the present study, the questions of features, their evaluation, and their integration will be addressed in terms of the voiced-voiceless distinction in word-final fricative consonants. There are many pairs of consonants in English which appear to differ only with respect to the distinctive feature of voicing (Jakobson, Fant, & Halle, 1961; Schane, 1973). These pairs occur within the classes of stops (/b,p/, /d,t/, /g,k/), fricatives (/z,s/, /v,f/, /ð,θ/, /ʒ,ʃ/), and affricates (/dʒ,tʃ/). The distinctive feature of voicing distinguishes minimal pairs of words such as the noun and verb meanings of *use*, *at* and *add*, and *tack* and *tag*. Close examination of these pairs reveals that a wide variety of acoustic characteristics is present and may serve as distinguishing features (Lisker, 1977; Massaro, 1975b; Massaro & Cohen, 1976, 1977; Wilder, 1975).

In studies measuring the durations of vowels within spoken words, vowels preceding voiced consonants are found to be somewhat longer than the same vowel spoken before the unvoiced cognates (Lehmann & Heffner, 1943; Peterson & Lehiste, 1960; Sharf, 1962; Umeda, 1975). For example, with the words (the) *use* and (to) *use* as test material, Denes (1955) found that vowel duration varied from 40 to 80 msec in voiceless /jus/ and from 120 to 200 msec in voiced /juz/. The data show that, in general, vowels preceding voiced consonants are one and a half times longer than vowels preceding voiceless consonants.

Within the class of fricatives, consonant duration is also found to vary between voiced and voiceless cognates. In addition to vowel duration differences between the noun and verb forms of *use*, Denes (1955) found that fricative duration varied from 200 to 380 msec for /jus/ but only 100 to 180 msec for /juz/. Later investigators have shown that the voiceless fricative /s/, in the initial as well as final position, is greater in duration than the voiced /z/ (Klatt, 1976; Umeda, 1977).

In order to investigate the respective roles of vowel duration and consonant duration in the perception of voicing in fricatives, Denes (1955), using a synthetic /ju-/ portion spliced to a spoken voiceless fricative portion, systematically varied vowel duration and fricative duration. The results indicated that both vowel duration and fricative duration influenced identification of the test syllables as /juz/ or /jus/. Although Denes did not provide a formal analysis,

This research was carried out as an undergraduate thesis by the first author under the direction of the second author. The research was supported in part by National Institute of Mental Health Grant MH19399 and a grant from the honors program at the University of Wisconsin. Michael Cohen provided technical assistance and Gregg Oden and Ilse-Lehiste contributed valuable comments on the research. Marcia Derr is now at Sperry-Univac, St. Paul, Minnesota. Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

0031-5117/80/010051-09$01.15/0

he interpreted the results to mean that the identification of the test syllable was a direct function of the duration of the vowel relative to the duration of the final consonant. He concluded that the voicing identifications decreased as the ratio of final consonant duration to vowel duration increased. If this conclusion is valid, it would imply that vowel duration and consonant duration are not evaluated independently. More recently, however, Massaro and Cohen (1977) demonstrated that a quantification of the realtive-duration interpretation does not provide an adequate description of Denes' results. The results are much more accurately described by assuming that vowel duration and frication duration are evaluated as totally independent cues.

Using synthetic speech stimuli, Raphael (1972) extended the work of Denes and others by testing the effectiveness of vowel duration as a cue to voicing in a variety of word-final stops, fricatives, and consonant clusters. He concluded that "regardless of the cues for voicing or voicelessness used in the synthesis of the final consonant or cluster, listeners perceived the final segments as voiceless when they were preceded by vowels of short duration and as voiced when they were preceded by vowels of long duration." However, Raphael acknowledged that vowel duration was not the only cue to voicing and that the presence of other cues during the final segment did affect perception. Generally, stimuli synthesized with cues indicating the presence of voicing during the final segment were identified as the voiced alternative following vowels of shorter duration than stimuli without these cues.

Since the duration of speech segments can cue differences between different speech sounds (Klatt, 1976), it is important to understand how duration is processed (Massaro & Idson, 1976, 1978; Idson & Massaro, 1977) and to investigate stimulus factors which influence perception of duration. Lehiste (1976) was concerned with what effect a changing fundamental frequency (F_0) pattern could have on perception of vowel duration. She found that the vowel /a/, when synthesized with a rising-falling or falling-rising F_0 contour, was judged to be longer than the same vowel of equal duration with a completely level contour. If a changing fundamental frequency causes an increase in perceived duration of a vowel, and if vowel duration is a cue to voicing in a word-final consonant, then it follows that the perception of voicing in a consonant could be affected by the fundamental frequency (F_0) contour of the preceding vowel. Following this reasoning, Lehiste (Note 1) demonstrated this exact result in a study varying vowel duration and F_0 contour in synthetic speech stimuli ranging between bead-beat or bad-bat (the final stop was unreleased). Vowel duration and F_0 contour of the vowel had a significant effect on the identification of the following stop. Identification responses showed that the shift from the voiced to voiceless alternatives occurred at shorter vowel durations for stimuli with a falling F_0 pattern than for stimuli synthesized with a monotone pattern.

Lehiste (Note 1) postulated a two-step process to account for her results: (1) The pitch change resulting from the F_0 change produces a longer perceived duration of the vowel, and (2) the longer perceived duration cues a voiced final consonant. In other words, perceived vowel duration depends on both actual vowel duration and fundamental frequency. The present paper formalizes and tests this idea within the framework of a general model of speech perception (Massaro, 1975a, 1975b; Oden & Massaro, 1978). Two properties of the general model should be stressed, since these properties differ from most contemporary views of speech perception. First, perceptual recognition of speech is mediated by vowel, consonant-vowel, or vowel-consonant syllable prototypes in LTM. Second, the feature detection and evaluation process provides information about the degree to which each feature is present in the speech sound.

The assumption of syllable prototypes contrasts with the commonly accepted notion of phonetic or phonemic prototypes in which phonetic or phonemic decisions mediate speech perception. Although it is only natural to say that a particular feature cues voicing, the perception of voicing qua voicing does not mediate syllable recognition in the present model. Acoustic features are evaluated and integrated to distinguish among the syllable prototypes in LTM. Accordingly, the conscious realization that two sounds both differ in voicing would occur only with specific instruction or with a postperceptual introspection (Paap, 1975).

The assumption of continuous rather than all-or-none featural information contrasts with the traditional view of binary features in linguistic theory (Jakobson, Fant, & Halle, 1961). More recently, Chomsky and Halle (1968) and Ladefoged (1975) have allowed a multivalued representation of featural information at a perceptual level. In our model, each feature is evaluated in terms of a fuzzy predicate that specifies the degree to which it is true that the sound has that particular characteristic (Oden & Massaro, 1978). Given the fuzzy information passed on by feature evaluation, the integration of the information from the various features is more complex than in traditional all-or-none classificatory schemes. The exact integration process will be specified exactly in the quantification of the model.

The model of independent features integration has been proposed and tested by Oden and Massaro (1978) with respect to the integration of voicing and place of articulation information for stop consonant identification, and by Massaro and Cohen (1976,

1977) in describing the integration of cues to voicing in initial position fricatives. This independent model can be applied to the current situation in which the listener is asked to indicate which of the two alternatives (to) *use* or (the) *use* occurs on each trial. Given that only the two prototypes /uz/ and /us/ in LTM are relevant to the identification response, it is sufficient to describe processing in terms of only these two alternatives. Each prototype is defined in LTM by an ideal set of acoustic features. It is assumed that the descriptions of /us/ and /uz/ contrast in terms of vowel duration and frication duration. The prototype /uz/ is defined by a long vowel duration and a short frication duration:

$$/uz/: \text{long vowel} \wedge \text{short frication}. \quad (1)$$

The prototype /us/ is defined by:

$$/us/: \text{short vowel} \wedge \text{long frication}. \quad (2)$$

It is possible to simplify the descriptions of the prototypes by allowing short and long to be complementary values. In this case, the value of short can be defined as the negation of long and the prototypes can be represented as:

$$/uz/: \text{long vowel} \wedge \text{NOT (long frication)}, \quad (3)$$

$$/us/: \text{NOT (long vowel)} \wedge \text{(long frication)}. \quad (4)$$

Given the definition of these prototypes, appropriate identification requires the evaluation of the two acoustic features: vowel duration and frication duration.

The model assumes that these features are detected independently and held in a preperceptual auditory storage. The primary recognition process evaluates each of the features relevant to the identification task by determining the degree to which each feature is present. Each feature is assigned a value between .0 and 1.0, representing the degree to which that feature is present in the speech sound. A feature value of .5 would indicate a feature that was perfectly ambiguous with respect to its presence in the sound. To simplify the notation, let V correspond to the value of the feature long vowel duration and F correspond to the value of the feature long frication duration. In the model, the negation of a feature is defined as 1 minus the value of that feature. Since V and F correspond to long vowel and long frication, respectively, $1 - V$ and $1 - F$ correspond to NOT(long vowel) and NOT(long frication), respectively. Therefore, the degree to which the two features match the alternative /uz/ is indexed by V and $1 - F$ and the degree to which the two features match the alternative /us/ is indexed by $1 - V$ and F.

The primary recognition process replaces each ideal feature in the prototypes with the feature values from feature evaluation. Accordingly, the feature values in the speech sound are entered in the prototypes

$$/uz/: V \wedge (1 - F), \quad (5)$$

$$/us/: (1 - V) \wedge F. \quad (6)$$

Given the multiple features, a key aspect of the model is the specification of how the features are combined or integrated together in order to determine how well the speech sound matches the alternative prototypes. The degree to which the speech sound matches the prototype is given by a multiplicative combination of the independent feature values. In this case, the degree to which the sound matches /uz/, P(uz) is given by

$$P(uz) = V \times (1 - F). \quad (7)$$

The model assumes that the degree to which a sound matches /uz/ is the degree to which the sound has a long vowel duration multiplied by the degree to which the sound has a short fricative duration. Analogously, P(us) is given by

$$P(us) = (1 - V) \times F. \quad (8)$$

The final part of the primary recognition process is the classification of the speech sound. The classification is assumed to be based on the relative degree to which the speech sound matches each of the alternative prototypes. The proportion of times a sound is judged to be /uz/, J(uz), should be equal to the degree to which the sound matches /uz/ relative to the sum of the degrees to which the sound matches /uz/ and /us/, respectively.

$$J(uz) = \frac{P(uz)}{P(uz) + P(us)} = \frac{[V \times (1 - F)]}{[V \times (1 - F)] + [(1 - V) \times F]}. \quad (9)$$

Analogously, the proportion of times the sound is judged to be /us/ should be

$$J(us) = \frac{P(us)}{P(uz) + P(us)} = 1 - J(uz). \quad (10)$$

The early study by Denes (1955) allows a quantitative test of the present model, since he independently varied vowel duration and frication duration as cues

to the two pronunciations of the homograph *use*. Twenty stimuli were generated by the orthogonal combination of four synthetic vowel durations and five durations of the frication portion. The frication portion was taken from natural speech and did not contain vocal-cord vibration. In order to fit the current model to the results, four parameter values of V and five values of F were estimated from the data by minimizing the squared deviations between the predicted and observed results using the iterative routine STEPIT (Chandler, 1969).

The observed results and the predictions given by Equation 10 are shown in Figure 1. The parameter values for V were .14, .24, .41, and .74 with vowel durations of 50, 100, 150, and 200 msec, respectively. The estimated parameters for F were .07, .22, .54, .78, and .82 for frication durations of 50, 100, 150, 200, and 250 msec, respectively. These parameter values reflect the important contribution of vowel duration and frication duration to the /uz/-/us/ distinction. The root mean squared deviation (RMSD) between the predicted and observed results was .033. This description is twice as good as one based on the assumption that the ratio of frication duration to vowel duration is the critical cue (Denes, 1955). A formalization of this idea gave a RMSD of .067 even though 15 parameters had to be estimated. These results support the idea that vowel duration and frication duration are independent cues to recognition of the two pronunciations of the homograph *use*. The goal of the current paper is to formalize and test one description of how F_0 contour during the vowel is combined with these cues in recognition.

It is assumed that the effect of F_0 contour on perceptual recognition is mediated by its influence on perceived vowel duration. As suggested by Lehiste (Note 1), (1) F_0 contour serves to modify the value of perceived vowel duration, and (2) this modified value of perceived vowel duration acts as an independent cue to voicing in the final position fricative. In the present formulation, perceived frication duration determines the cue value of frication duration and perceived vowel duration determines the cue value of vowel duration. Since two variables, actual vowel duration and F_0 contour, influence perceived vowel duration, it is necessary to specify how perceived vowel duration influences cue value. Derr and Massaro (Note 2) suggested that a direct proportional relationship was not reasonable. For example, at very short and very long vowel durations, the contour may greatly modify perceived duration but may have an extremely small effect on the vowel duration cue value for distinguishing the syllables /uz/ and /us/.

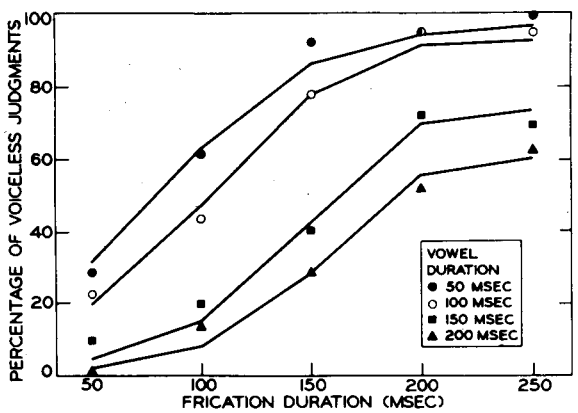

Figure 1. Predicted and observed voiceless identification as a function of vowel duration and frication duration (observed results from Denes, 1955).

The relationship between perceived vowel duration and cue value may be better described by an ogival function: at short perceived durations, cue value changes very slowly with changes in perceived duration; at medium perceived duration, cue value changes very rapidly; at long perceived durations, cue value changes slowly again as it reaches its asymptote. This relationship means that F_0 contour is much more critical for syllable identification when the cue value of vowel duration is relatively ambiguous at medium perceived vowel durations.

The ogival relationship between perceived vowel duration and cue value can be quantified by first assuming that both actual vowel duration and F_0 contour influence the perceived duration of the vowel. However, the cue value V of perceived vowel duration to syllable identification is *not* directly proportional to perceived duration. This follows from the idea that the cue value of vowel duration is more adequately represented as an ogival function of perceived vowel duration. Changes in perceived duration when it is very short or very long will have very little effect on the cue value of V, whereas equivalent changes when perceived vowel duration is intermediate will have a much larger effect on cue value V. Unfortunately, there is no direct estimate of perceived vowel duration in the present experiment, but only a direct index of its cue value V.

Perceived vowel duration P is assumed to be a linear function of vowel duration D with F_0 contour determining the rate parameter C

$$P = CD. \tag{11}$$

Finally, it is assumed that cue value V is an ogival function of perceived duration p,

$$V = \frac{\chi_p^y}{\chi_p^y + (1-\chi_p)^y}, \tag{12}$$

where $y \geq 1$ and $\chi_p = aP + b$, except that values of χ_p less than zero are set to zero and values greater than one are set to one. Equation 12 is ogival in form when the χ_p values fall between zero and one.

The y parameter determines the steepness of the middle section of the ogive. The values of χ_p are assumed to be a linear function of perceived duration P. The parameter a and b allow the ogive to shift linearly along the dimension of perceived duration. The cue values V are inserted in Equation 9 in order to describe the identification responses.

Two experiments were performed to examine vowel duration, frication duration, and fundamental frequency contour as cues to the /uz/-/us/ distinction. Replicating the study of Denes (1955), vowel duration and frication duration were varied in synthetic stimuli representing the /jus/-/juz/ pair. In addition the fundamental frequency contour was varied as in Lehiste (Note 1). In the first experiment, binary choice data were collected. The second experiment replicated the first experiment, except that a continuous rating response was used. Data from both experiments were described by the feature integration model. The knowledge gained from this investigation should contribute to our understanding of the process of feature integration in speech perception.

EXPERIMENT 1

Method

Subjects. The subjects were seven undergraduates who participated to fulfill an introductory psychology course requirement. All were native speakers of English and reported no hearing impairments. Each subject was tested on 2 consecutive days, the same time each day.

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 segment being specified by a list of synthesis parameter control vectors stored in the computer. Each list contained a target value, a transition time, and a transition type (linear, positive, or negative interpolation) for each changing parameter. All durations were specified as multiples of 5 msec. Parameter values were calculated in 5-msec increments and output to the synthesizer at the same rate.

The speech sounds were glide-vowel-fricative syllables varying along three dimensions (vowel duration, frication duration, and fundamental frequency contour) to form a voiced to voiceless continuum (/juz/ to /jus/). Figure 2 shows a schematic diagram of the parameters specifying one of the syllables used in this experiment. Spectrograms of two actual stimuli are shown in Figure 3. These two examples represent the extreme values of vowel duration and fricative duration. Differences in F_0 contour are not shown because they are not easily seen in wide-band spectrograms. All syllables had five formants and began with a three-formant voiced transition. The first formant (F_1) rose linearly (252-308 Hz) in 50 msec; the second formant (F_2) fell positively (2,397-951 Hz) in 80 msec; and the third formant (F_3) fell positively (3,020-2,468 Hz) in 120 msec. The fourth and fifth formants (F_4 and F_5) were fixed throughout the sound at 3,500 and 4,000 Hz, respectively. The fundamental frequency (F_0) began at one of three levels—168, 112, or 75 Hz. The next segment was a variable length steady state vowel with F_1 set at 308 Hz, F_2 at 951 Hz, and F_3 at 2,468 Hz. At the start of the steady state vowel, F_0 could either remain level or begin to rise or fall. The steady state vowel was followed by a 30-msec linear formant transition, with F_1 falling to 291 Hz and F_2 and F_3 rising to 1,131 and 2,614 Hz, respectively. The final stimulus segment was a variable-duration frication period, of which the first 50 msec maintained the vocalic segment with the amplitude of the buzz source being turned off linearly. The noise formants, K_1 and K_2, were set at neutral values equally appropriate for both /s/ and /z/, K_1 at 4,150 Hz and K_2 at 7,834 Hz.

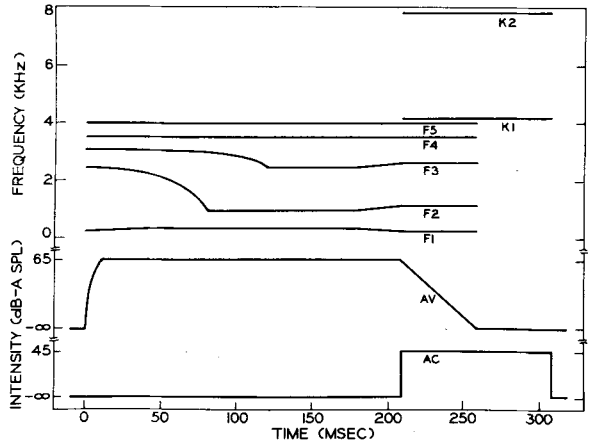

Figure 2. Schematic diagram of the properties of the speech stimuli.

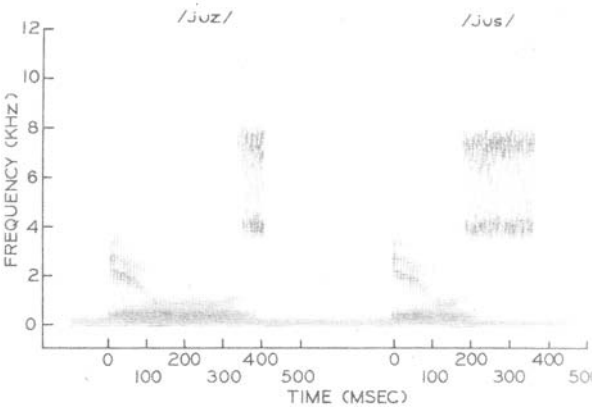

Figure 3. Spectrograms of two synthesized speech sounds representing the extreme values of vowel duration and frication duration.

For the benefit of readers' familiar with the OVE synthesizer, the amplitude values in terms of OVE specifications were as follows. The buzz source amplitude (AV) reached full volume of 28 dB 10 msec from the start of the syllable. It remained at this level until the onset of frication, at which point AV was turned off, decreasing linearly in 50 msec. The amplitude of the pseudo-random noise source simulating frication (AC) was 2 dB and had instantaneous onset and offset. The frication pole/zero ratio (AK) was set to 8 dB, i.e., the low-frequency spectrum level (below K_1) was set to 10 dB lower than AC.

Eighty different stimuli were synthesized by factorially combining four levels of vowel duration, four levels of frication duration and five fundamental frequency contours. The four possible vowel durations, including the 30-msec transition to frication, were 40, 90, 140, and 200 msec. The possible durations for the frication alone period which followed the 50-msec vocalic frication period were 20, 50, 90, and 140 msec. The F_0 contour was either level or changing. In order to control for the possible effects of starting or ending frequency, three level contours, high, medium, and low, and two changing contours, rising and falling, were included.

For the level contours, F_0 was set at 168, 112, or 75 Hz; the rising F_0 contour and the falling F_0 contour both started at 112 Hz. For both moving F_0 contours, the frequency changed linearly about half an octave in 250 msec. For the rising contour, F_0 frequency increased linearly from 112 to 168 Hz. For the falling contour, F_0 frequency decreased linearly from 112 to 75 Hz. In terms of rate of change, F_0 rose at the rate of 224 Hz/sec and fell at the rate of 140 Hz/sec. For the rising and falling conditions, F_0 remained at 112 Hz throughout the initial 120-msec glide and began to rise or fall at the onset of the steady state vowel period and continued throughout the remainder of the vocalic period. Since the rate of rising or falling was constant across all levels of vowel duration, the F_0 values at the end of the vowel varied with each vowel duration.

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 headphones (Koss Pro/4AA). Listening sound pressure, measured by a Bruel Kjaer Type 2203 sound-level meter and Type 4153 artificial ear, ranged from 62 to 65 dBA SPL for vocalic portions of the stimuli and registered at 45 dBA SPL during frication. Three or four subjects were tested simultaneously in sound-attenuated rooms.

Each trial began with the presentation of a stimulus selected randomly without replacement in blocks of 80 trials. The subjects were instructed to listen to each stimulus and decide whether it sounded more like /jus/, as in the noun "the use," or more like /juz/, as in the verb "to use." The subjects were told that there would be no ordered pattern to the presentation of the different stimuli; rather, the order was random. They were encouraged to make the best possible guesses for ambiguous stimuli. Following a stimulus presentation, each subject responded by pressing a button labeled "the use" or a button labeled "to use." The subject had 2 sec to respond, after which three asterisks appeared on a visual display to signal the end of the response period. A new trial began 1.25 sec after the end of the response period.

On the 1st day, a practice session of 80 unscored trials was given to familiarize the subjects with the response apparatus and the stimulus range. The subjects then participated in two sessions on each of 2 days. Each session lasted approximately 15 min. Between sessions, the subjects were given a 5-min break. During each session, the subjects were presented with three blocks of the 80 stimuli—240 trials per session. Ten randomly chosen practice trials preceded each session. The subjects were not informed that these trials were not scored. Over the 2 days, each subject contributed 12 observations for each of the 80 stimuli.

Results

The data were pooled over sessions and the proportions of voiced or /juz/ responses in each condition were calculated. A three-way (Vowel Duration by Frication Duration by Fundamental Frequency Contour) analysis of variance was carried out on the proportion data.

The mean proportion of voiced responses was .522. Voiced identifications increased from .202 to .752, on the average, with increases in vowel duration [$F(3,18) = 60$, $p < .001$]. As frication duration decreased, voiced judgments shifted from .320 to .752 [$F(3,18 = 23$, $p < .001$]. There was a main effect of F_0 contour [$F(4,24) = 9.7$, $p < .001$]. With level contours, voiced responses averaged .452, .412, and .462 for high, medium, and low F_0 contours, respectively. Voiced judgments were .694 for the falling contour and .590 for the rising contour. All the two-way interactions and the three-way interactions were also statistically significant.

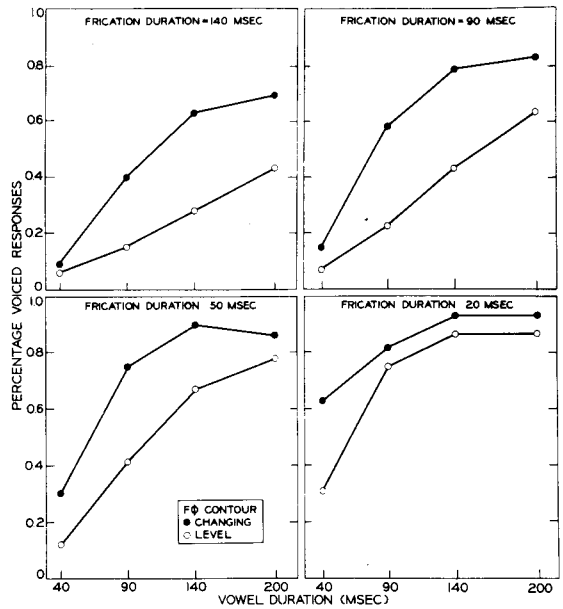

Figure 4. Percentage of voiced (juz) responses as a function of vowel duration, frication duration, and F_0 contour.

In summarizing the data, the three level contours were pooled together, as were the two changing contours. These data are shown in Figure 4, with proportion of /juz/ identifications plotted as a function of vowel duration, F_0 contour, and frication duration. In agreement with past studies, Figure 4 shows that the perception of voicing in final-position fricatives was found to be dependent on the contributions of three cues. As found by Denes (1955), voiced responses increased with increases in vowel duration and decreases in frication duration. The results also replicated Lehiste's (Note 1) finding that a changing F_0 contour during the vowel increased the proportion of voiced judgments of the following consonant. The stimuli captured almost the entire response range, as evidenced by a low voiced proportion (.06) for stimuli synthesized with a 40-msec vowel duration, 140-msec fricative duration, and a level F_0 contour, and a high proportion (.94) for stimuli synthesized with a 200-msec vowel duration, a 20-msec fricative duration, and a changing F_0 contour.

EXPERIMENT 2

Method

Subjects. Ten undergraduate students participated in the experiment to fulfill an introductory psychology course requirement. All were native English speakers and reported no hearing impairments. Each subject was tested on 2 consecutive days, the same time each day.

Stimuli. The stimuli were identical to those in Experiment 1.

Procedure. The procedure was identical to those in Experiment 1, except that the subjects were required to perform a rating task instead of a binary choice task. Following the stimulus presentation, each subject made a response by setting the pointer of a linear potentiometer, 5.5 cm long. The left end of the scale was

labeled "the use," the right end was labeled "to use." When satisfied with the position of the pointer, the subject caused the position to be recorded by pressing a button to the right of the scale on the response box. The potentiometer acted as a voltage divider, and the resulting voltage was measured by a multiplexed A-D converter and scaled so that the resulting score varied from 0 to 1.0 in steps of .02. The next trial began 1 sec after the last of the subjects made a response.

The subjects were asked to rate the stimuli on a scale from /jus/, as in the noun "the use" to /juz/, as in the verb "to use." They were told that the left end of the scale represented a good /jus/ while the right end was for a good /juz/. Having heard all the stimuli in a practice session, they were told to adjust their responses so as to try to cover the entire scale.

Results

The rating responses were pooled over sessions and averaged, and a three-way analysis of variance (Vowel Duration by Frication Duration by F_0 Contour) was performed.

As vowel duration increased from 40 to 200 msec, the average ratings became more voiced from .179 to .727 [$F(3,27) = 90$, $p < .001$]. As frication duration was decreased from 140 to 20 msec, the ratings shifted from predominantly voiceless (.295) to predominantly voiced (.700) [$F(3,27) = 53$, $p < .001$]. The overall effect of contour was also significant [$F(4,36) = 10.5$, $p < .01$], with falling and rising contours being rated .560 and .528, respectively. The high-, medium-, and low-level contours were rated .401, .468, and .474, respectively, less voiced than in both the rising and falling contour conditions. All two-way interactions and the three-way interaction were also statistically significant.

Figure 5 provides a summary of the data. Again, level and changing contours are plotted separately and average rating response is plotted as a function of vowel duration and fundamental contour at each level of frication duration. These results replicate the general pattern of the results of Experiment 1. The ratings became more voiced with increases in vowel duration, decreases in frication duration, and the presence of a changing F_0 contour. Response ratings ranged from .087 for stimuli with the shortest vowel duration, longest frication duration, and a level F_0 contour to .848 for stimuli with the longest vowel duration, shortest frication duration, and a changing F_0 contour.

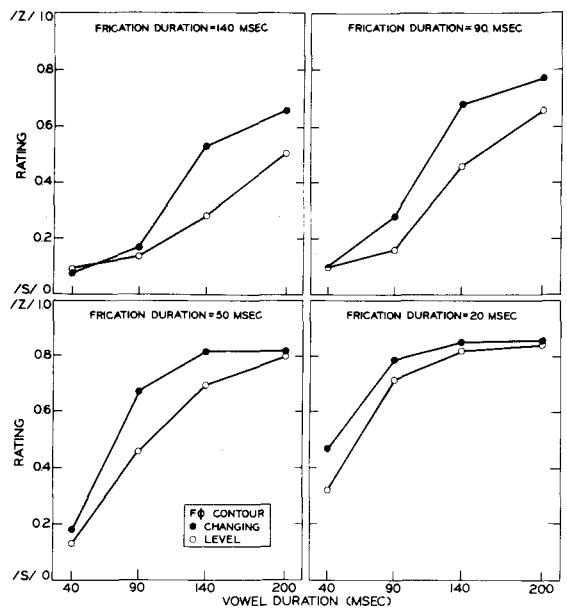

Figure 5. Average rating response as a function of vowel duration, frication duration, and F_0 contour.

DISCUSSION

The data from both experiments were described by the model as formalized in Equations 9 through 12. The model was fit to the group response means of Experiment 1. The parameter values were estimated by minimizing the squared deviations between the predicted and observed responses using the iterative minimization routine STEPIT (Chandler, 1969). To simplify the estimation procedure, the fitting of the model involved three steps. First, Equation 9 was fit to the 80 unique experimental conditions by estimating 4 F values for the 4 frication durations and 20 V values for the 20 combinations of the 5 F_0 contours and the 4 vowel durations. Second, the 20 estimated V values were fit by Equations 11 and 12 by estimating 5 rate parameters C corresponding to the 5 F_0 contours and the parameter values of a, b, and y. Finally, the 4 F parameters from the first fit and the 8 parameters from the second fit were used to predict the 80 unique experimental conditions. The predictions of the 80 independent conditions are, therefore, based on 12 free parameters. The root mean squared deviation (RMSD) was .077.

Table 1 gives the 5 rate parameters C corresponding to the 5 F_0 contours and the parameter values of a, b, and y. Table 2 gives the predicted cue values for the 20 combinations of F_0 contour and vowel duration; the 4 F values corresponding to the 4 frication durations are also given in the table. The higher values of the rate parameter C for rising and falling F_0 contours than for the steady-state contours reflect the longer perceived vowel durations for changing F_0 contours. The cue values correspond to the form of the results; the sounds were heard as more voiced with longer vowel durations, changing F_0 contours, and shorter frication durations.

The same analyses were carried out on the rating results from Experiment 2 by assuming that the rating response was a direct index of J(uz). The parameter values and the cue values are presented in Tables 3 and 4. The RMSD value was .069. The parameters are similar to those obtained from Experiment 1. The only large difference is that the rate parameter for a falling contour is smaller in Experiment 2 than

Table 1
The Five Rate Parameters C Corresponding to the Five F_0 Contours and the Parameters a, b, and y for Experiment 1

F_0 Contour (C)				
High	Rise	Medium	Fall	Low
.44	.64	.40	.79	.44

Note–a = .0009, b = .45, c = 11.67.

Table 2
The Cue Values Corresponding to the V and F Values in Equation 9 as a Function of Vowel Duration, F_0 Contour, and Frication Duration for the Predicted Results Given by Equations 9-12 for the Results of Experiment 1

Vowel Duration (msec)	F_0 Contour					Frication Duration (msec)	
	High	Rise	Med	Fall	Low		
40	.16	.22	.15	.29	.16	140	.75
90	.32	.52	.29	.68	.32	90	.60
140	.55	.81	.48	.92	.54	50	.40
200	.75	.94	.68	.98	.75	20	.16

Table 3
The Five Rate Parameters C Corresponding to the Five F_0 Contours and the Parameters a, b, and y for Experiment 2

F_0 Contour				
High	Rise	Medium	Fall	Low
.42	.56	.48	.59	.47

Note–a = .0009, b = .44, y = 12.06.

Table 4
The Cue Values Corresponding to the V and F Values in Equation 9 as a Function of Vowel Duration, F_0 Contour, and Frication Duration for the Predicted Results Given by Equations 9-12 for the Results in Experiment 2

Vowel Duration (msec)	F_0 Contour					Frication Duration (msec)	
	High	Rise	Med	Fall	Low		
40	.14	.19	.16	.20	.16	140	.76
90	.29	.45	.36	.47	.34	90	.65
140	.50	.74	.61	.76	.59	50	.40
200	.71	.91	.82	.92	.80	20	.21

in Experiment 1. The similarity in the model's description of ratings and binary choice responses supports the idea that they are equally good measures of speech perception.

In summary, the present model gives a reasonably good description of both binary choice judgments and rating responses. Although the results are adequately described by the assumption that F_0 contour modifies perceived vowel duration, the possibility remains that F_0 contour is an independent cue to syllable identity. Derr and Massaro (Note 2) formalized two models in order to distinguish between whether the perceived pitch of F_0 contour is a direct cue to voicing in final-position consonants or whether the effect of F_0 contour is indirect, resulting from its influence on perceived vowel duration. The results of the experiments reported here could not discriminate between the models and, therefore, the data are uninformative with respect to the issue of direct or indirect effects of F_0 contour. Given that these two possibilities are difficult to distinguish in terms of their quantitative predictions, it will be necessary to devise a qualitative test between them.

REFERENCE NOTES

1. Lehiste, I. *Contribution of pitch to the perception of segmental quality.* Paper presented at the 9th International Congress on Acoustics, 1977.

2. Derr, M. A., & Massaro, D. W. *The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction.* WHIPP Report No. 8, Wisconsin Human Information Processing Program, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706, September 1978.

REFERENCES

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

CHOMSKY, N., & HALLE, M. *The sound pattern of English.* New York: Harper and Row, 1968.

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

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

IDSON, W. L., & MASSARO, D. W. Perceptual processing and experience of auditory duration. *Sensory Processes*, 1977, **1**, 316-337.

JAKOBSON, R., FANT, C. G. M., & HALLE, M. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, Mass: MIT Press, 1961.

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.

LADEFOGED, P. *A course in phonetics*. New York: Harcourt Brace Jovanovich, 1975.

LEHISTE, I. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 1976, **4**, 113-117.

LEHMANN, W. P., & HEFFNER, R. M. S. Notes on the length of vowels. *American Speech*, 1943, **18**, 208-215.

LISKER, L. Rapid versus rabid: A catalogue of acoustic features that may cue the distinction. *Journal of the Acoustical Society of America*, 1977, **62**(S1), S77(A).

MASSARO, D. W. *Experimental psychology and information processing*. Chicago: Rand-McNally, 1975. (a)

MASSARO, D. W. (Ed.) *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975. (b)

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.

MASSARO, D. W., & COHEN, M. M. Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics*, 1977, **22**, 373-382.

MASSARO, D. W., & IDSON, W. L. Temporal course of perceived auditory duration. *Perception & Psychophysics*, 1976, **20**, 331-352.

MASSARO, D. W., & IDSON, W. L. The temporal course of perceived vowel duration. *Journal of Speech and Hearing Research*, 1978, **21**, 37-55.

ODEN, G. C., & MASSARO, D. W. Integration of featural information in speech. *Psychological Review*, 1978, **85**, 172-191.

PAAP, K. R. Theories of speech perception. In D. W. Massaro (Ed.), *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

PETERSON, G. E., & LEHISTE, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 1960, **32**, 693-703.

RAPHAEL, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, **51**, 1296-1303.

SCHANE, S. A. *Generative phonology*. Englewood Cliffs, N.J: Prentice-Hall, 1973.

SHARF, D. J. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech*, 1962, **5**, 26-30.

UMEDA, N. Vowel duration in American English. *Journal of the Acoustical Society of America*, 1975, **58**, 434-445.

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

WILDER, L. Articulatory and acoustic characteristics of speech sounds. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

(Received for publication September 19, 1978;
revision accepted October 29, 1979.)