Psychological Review
1978, Vol. 85, No. 3, 172-191

Integration of Featural Information in Speech Perception

Gregg C. Oden and Dominic W. Massaro
Univeristy of Wisconsin—Madison

A model for the identification of speech sounds is proposed that assumes that (a) the acoustic cues are perceived independently, (b) feature evaluation provides information about the degree to which each quality is present in the speech sound, (c) each speech sound is defined by a propositional prototype in long-term memory that determines how the featural information is integrated, and (d) the speech sound is identified on the basis of the relative degree to which it matches the various alternative prototypes. The model was supported by the results of an experiment in which subjects identified stop-consonant–vowel syllables that were factorially generated by independently varying acoustic cues for voicing and for place of articulation. This experiment also replicated previous findings of changes in the identification boundary of one acoustic dimension as a function of the level of another dimension. These results have previously been interpreted as evidence for the interaction of the perceptions of the acoustic features themselves. In contrast, the present model provides a good description of the data, including these boundary changes, while still maintaining complete noninteraction at the feature evaluation stage of processing.

Although considerable progress has been made in the field of speech perception in recent years, there is still much that is unknown about the details of how speech sounds are perceived and discriminated. In particular, while there has been considerable success in isolating the dimensions of acoustic information that are important in perceiving and identifying speech sounds, very little is known about how the information from the various acoustic dimensions is put together in order to actually accomplish identification. The present article proposes and tests a model of these fundamental integration processes that take place during speech perception.

Much of the study of features in speech has focused on the stop consonants of English. The stop consonants are a set of speech sounds that share the same manner of articulation: Their production begins with a buildup of pressure behind some point in the vocal tract, following which is a sudden release of that pressure. In terms of their production, the six stops in English can be classified using the two featural dimensions of place of articulation and voicing. *Place of articulation* refers to the point in the oral cavity at which the air flow is blocked or occluded. *Voicing* refers to whether or not vocal-cord vibration occurs during the period of occlusion and release. The six stops of English consist of three cognate pairs that share place of articulation but differ in voicing: The consonants /p/ and /b/ are labial, /t/ and /d/ are alveolar, and /k/ and /g/ are velar. The first member of each pair is voiceless and the second is voiced.

The above classification based on speech *production* follows from the idea that place of articulation can be described independently of voicing. Analogously, much of the research on speech *perception* has operated on the corresponding idea that the perception of place of articulation can occur independently of the perception of voicing. Data supporting this premise were accumulated in research using the pattern playback synthesizer (Delattre, Liberman, & Cooper, 1955; Liberman, Delat-

This research was supported in part by National Institute of Mental Health Grant MH 19399 and grants from the Wisconsin Alumni Research Foundation. James Bryant, Michael Cohen, and David Warner provided assistance in performing the experiment and the members of the Wisconsin Human Information Processing Program (WHIPP), especially Lola Lopes, provided useful comments on this research.

Requests for reprints should be sent to Gregg C. Oden or Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

tre, & Cooper, 1958). This research revealed that perception of place of articulation was primarily a function of one set of acoustic cues, whereas perception of voicing was primarily a function of another set of cues. Perception of voicing was shown to be influenced by the voice onset time (VOT), the time between the onset of the release burst and the onset of vocal-cord vibration, and also by the degree of aspiration during the VOT period (Liberman et al., 1958). On the other hand, perception of place of articulation was shown to be primarily a function of the second and third formant (F_2 and F_3) transitions and of the burst frequency (Delattre et al., 1955; Harris et al., 1958; Hoffman, 1958). Another type of data that was taken to support the independent processing of speech features was the pattern of confusion errors obtained when subjects listened to speech sounds presented against various levels of noise (Miller & Nicely, 1955).

Although these early experiments supported the perceptual independence of place and voicing, more recent research appears to indicate that there is some dependence in the perception of place and voicing information (Abramson & Lisker, 1973; Haggard, 1970; Lisker & Abramson, 1970; Smith, 1973). Lisker and Abramson (1970), for example, using synthesized speech sounds, showed that voicing judgments in English were critically dependent on place of articulation. The boundary between voiced and voiceless sounds, measured in terms of VOT, was about 23 msec for labials, 37 msec for alveolars, and 42 msec for velars. These perceptual results agreed rather well with the range of VOT values derived from acoustical measurements of natural speech (Klatt, 1975; Lisker & Abramson, 1964). As Miller (1977) points out, however, the duration of the formant transitions in Lisker and Abramson's synthetic stimuli differed for the different places of articulation. These duration differences may, therefore, be directly responsible for the differences in VOT boundaries. Eliminating this problem, Miller still found significant differences, although the change in the VOT boundary was now only about one fourth as large (4.75 msec) as that reported by Lisker and Abramson.

The change in the VOT boundary with place of articulation might seem to be evidence against the perceptual independence of place and voicing in stop consonants. However, whether or not such changes in the voicing boundary constitute evidence against perceptual independence depends on the underlying model of speech perception that is assumed. It is possible that the acoustic features of place and voicing may be perceived independently and that changes in the voicing boundary may simply result from the way in which features are evaluated, combined, and matched against memorial representations of the alternative consonants.

The model proposed in the present article provides a detailed description of the processes that may be involved in using featural information to identify speech sounds. This model will be tested directly by using the procedures of information integration theory (Anderson, 1974). In the present case, these procedures involve the formulation of a model consisting of a set of algebraic rules to describe the integration processes. This model is then tested with identification data for synthetic speech stimuli that have been factorially generated by independently varying acoustic cues for voicing and for place of articulation.

The proposed integration model, which will be described in detail in the next section, can be articulated within the framework of a more general auditory information-processing model (Massaro, 1975a, 1975b). Figure 1 presents a schematic diagram of the auditory recognition process in Massaro's model. According to this model, the auditory stimulus is transduced by the auditory receptor system and acoustic features are detected and stored in preperceptual auditory storage (PAS). The features stored in PAS are a direct consequence of the properties of the auditory stimulus and the auditory receptor system. It is assumed that the feature detection process cannot be modified by learning or by the listener's knowledge or expectations. The features are assumed to be independent; the value of one feature does not influence the value of another at this stage of processing.

The primary recognition process evaluates each of the acoustic features in PAS and compares or matches these features to those that define perceptual units in long-term memory

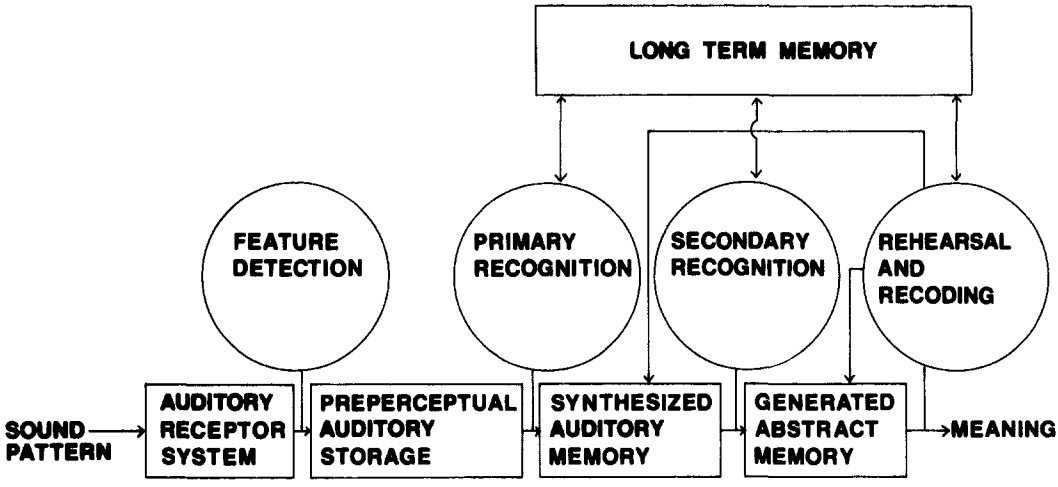

Figure 1. Schematic diagram of the general auditory information processing model.

(LTM). Every perceptual unit has a representation in LTM, which is called a *sign* or *prototype.* The prototype of a perceptual unit is specified in terms of the acoustic features that define the ideal acoustic information as it would be represented in PAS. The recognition process operates to find the prototype in LTM that best matches the acoustic features in PAS. It should be stressed that the primary recognition operation is not simply a pure template matching of features. In speech perception, there is good evidence for a normalization process that adjusts for variations in the voice quality of different speakers, speaking rate, and so on (see Massaro, 1975b, pp. 88–92, for a review of the evidence concerning voice quality). In our view, the adjustment operation does not have a direct influence on the evaluation of the acoustic features but occurs at the later prototype matching stage (see below). For the present, the allowance of this top–down influence should comfort those who are justifiably skeptical of the sufficiency of only bottom–up processes. In addition, as will be described below, the prototypes are not simply loose conglomerations of features but rather are propositions that may be, in principle, arbitrarily rich in logical structure.

The perceptual outcome of primary recognition is held in synthesized auditory memory (SAM). In contrast to feature evaluation, the outcome of the primary recognition process is influenced by the listener's knowledge and expectations and can be modified by learning experience. The secondary recognition process translates the perceptual code in SAM into an abstract code in generated abstract memory (GAM). The critical difference between SAM and GAM is in terms of the properties of the stored information. The synthesized percept of a friend's voice can be thought of as the actual sound experience, whereas the abstract encoding might be in terms of defining characteristics such as low and harsh. This model has previously been evaluated primarily with experiments on the dynamics of auditory information processing and speech perception. Within this context, the goal of the present work is to extend and quantify the model to describe how the listener integrates the various acoustic features in the identification of a speech sound.

Fuzzy Logical Model of Phoneme Identification

According to the proposed integration model, there are three conceptually distinct operations involved in phoneme identification: (a) The *feature evaluation* operation determines the degree to which each feature is present in PAS, (b) the *prototype matching* operation determines how well each candidate phoneme provides an absolute match to the speech sound, and (c) the *pattern classification* operation determines which phoneme provides the

best match to the speech sound relative to the other phonemes under consideration.

Feature Evaluation

The feature evaluation process provides information about the degree to which each feature is present in the speech sound. Rather than assuming that the listener simply detects presence or absence, we assume that the feature is perceptually more or less present. This assumption is supported by the results of recent studies that, in contrast to the earlier work on categorical perception, have shown that acoustic features are perceived continuously rather than in an all-or-none fashion. Barclay (1972) required subjects to identify consonants as either /b/ or /g/ even though phonetically all were instances of /d/. Under these conditions, subjects more often identified a sound as /b/ the closer it was to the labial end of the place dimension. Pisoni and Tash (1974) found that the latency for deciding whether two sounds are the same phoneme is dependent on their *degree* of similarity with respect to VOT. Using a discrimination task that minimized the sensory interference from successive stimuli, Pisoni and Lazarus (1974) and Carney, Widin, and Viemeister (1977) demonstrated that subjects can reliably discriminate speech sounds that are acoustically different but phonetically the same. McNabb's (Note 1) subjects were more confident in their phonetic classifications for stimuli that were more extreme on the acoustic dimension. All of these results are consistent with the assumption that listeners can hear the degree to which acoustic features are present in speech sounds.

The assumption of continuous acoustic features contrasts with the traditional description of binary all-or-none distinctive features (Jakobson, Fant, & Halle, 1961) but corresponds to the more recent treatment of distinctive features provided by Chomsky and Halle (1968). They distinguish between the classificatory and phonetic function of distinctive features. The features are envisioned as binary (+ or −) only in their classificatory function. In their phonetic or descriptive function, they are multivalued features that relate to aspects of the speech sounds and the perceptual representation (Chomsky & Halle, 1968, p. 298). Similarly, Ladefoged (1975) distinguishes between the phonetic and phonemic level of description. A feature describing the phonetic quality of a sound has a value along a continuous scale, whereas a feature classifying the phonemic oppositions is given a discrete value. In terms of our model, the representation of acoustic features in PAS would be comparable to the continuous values of their phonetic features. Even though place and voicing would be expressed as continuous rather than discrete, the phonemic judgment may still be discrete. That is to say, the listener can hear the degree of voicing but the listener's judgment in a forced-choice classification task with the six stops as alternatives will be either voiced or voiceless. Analogously, the degree of alveolarity of a stop consonant can be perceived, but the classification will be labial, alveolar, or velar.

Since acoustic features vary continuously from one speech sound to another, they can be represented as predicates that may be more or less true rather than only absolutely true or false (Goguen, 1969; Zadeh, 1975). These so-called *fuzzy predicates* represent the feature evaluation process: Each predicate is applied to the speech sound and specifies the degree to which it is true that the sound has the relevant acoustic characteristic. For example, if we use the notation t(A) to signify the truth value of Proposition A, then

$$\mathrm{t}[\textsc{voiced}(S_{ij})] = .65 \qquad (1)$$

represents the fact that it is .65 true that a given speech sound (S_{ij}), from the ith row and jth column of the factorial stimulus design, is perceived to be voiced. Similarly,

$$\mathrm{t}[\textsc{alveolar}(S_{ij})] = .30 \qquad (2)$$

signifies that it is .30 true that the speech sound is perceived to be alveolar. To simplify the notation, let

$$A_i = \mathrm{t}[\textsc{alveolar}(S_{ij})] \qquad (3)$$

and

$$V_j = \mathrm{t}[\textsc{voiced}(S_{ij})], \qquad (4)$$

so that A_i and V_j are subjective values that specify the degree to which the speech sound is perceived to be alveolar and voiced, respectively.

Prototype Matching

Each phoneme is defined by a prototype in long-term memory corresponding to a proposition such as

/b/: (LABIAL) AND (VOICED), (5)

/p/: (LABIAL) AND [NOT (VOICED)], (6)

/d/: (ALVEOLAR) AND (VOICED), (7)

and

/t/: (ALVEOLAR) AND [NOT (VOICED)]. (8)

Proposition 5 says simply that /b/ is labial *and* voiced, Proposition 6 specifies that /p/ is labial *and* not voiced, and so on. We actually assume that the relevant prototypes in LTM correspond to consonant–vowel syllables rather than stop consonants. The acoustic cues to stop-consonant phonemes depend critically on vowel context, and this lack of invariance disqualifies the stop-consonant phoneme as a perceptual unit prototype in long-term memory (Massaro, 1975b). However, for ease of exposition and because the vowel is constant, we will refer to the classification of these sounds as *phoneme identification.*

These simple propositions are themselves not fuzzy and are identical to the traditional, discrete featural definitions of these phonemes. However, in the fuzzy logical model, these prototypes are translated directly into fuzzy propositions that are the matching functions that specify the degree to which a given speech sound matches the LTM prototype of each of the associated phonemes. The translation from prototype to matching function involves two steps. First, the features in the prototypes must be replaced with the fuzzy featural predicates from the feature evaluation stage. Second, conjunction and negation must be defined for the fuzzy case. On the basis of previous work in speech perception (Massaro & Cohen, 1976, 1977; Oden, in press) and also in other cognitive domains (Oden, 1977), we assume that conjunction and negation follow Equations 9 and 10, respectively:

$$t(A \wedge B) = t(A) * t(B) \tag{9}$$

and

$$t(\neg A) = 1 - t(A), \tag{10}$$

where A and B are arbitrary propositions.

Consequently, the four matching functions corresponding to the prototypes given above are

$$B(S_{ij}) = L_i V_j, \tag{11}$$

$$P(S_{ij}) = L_i(1 - V_j), \tag{12}$$

$$D(S_{ij}) = A_i V_j, \tag{13}$$

and

$$T(S_{ij}) = A_i(1 - V_j). \tag{14}$$

For example, the degree to which a perceived speech sound will match the prototype of /b/ is specified by the matching function $B(S_{ij})$. According to Equation 11, this matching function for /b/ is equal to the degree to which the sound is labial multiplied by the degree to which the sound is voiced. Equations 12–14 define matching functions that specify the degree to which the speech sound matches the prototypes for /p/, /d/, and /t/, respectively.

Pattern Classification

In the final operation, the speech sound is classified on the basis of the relative degree to which it matches the various alternative phoneme prototypes as specified by the matching functions. It is assumed that the person classifies the sound as being an instance of whichever phoneme provides the best match. However, since perception is a noisy process in which a given physical stimulus will be perceived differently at different times, phoneme classification is necessarily a probabilistic process. Probabilistic choice processes of this sort may be modeled in a number of theoretically different ways that are formally similar (e.g., Luce, 1959; Thurstone, 1927). For the purposes of the present article, it will be assumed that the choice process follows Luce's model. Thus, for example, in the present experiment in which listeners were asked to identify the initial consonant of the speech sounds as /b/, /p/, /d/, or /t/, the probability that a given speech sound is identified as /b/ rather than /p/, /d/, or /t/ should be

$$p(b|S_{ij}) = \frac{B(S_{ij})}{B(S_{ij}) + P(S_{ij}) + D(S_{ij}) + T(S_{ij})}. \tag{15}$$

In general, the probability of identifying a sound to be a particular phoneme should be

equal to the goodness of the match of the sound to that phoneme relative to the sum of the goodness-of-match values for all of the phonemes being considered.

If we expand Equation 15 by inserting the equations for the various matching functions, the result is

$$p(\mathrm{b}|S_{ij}) = \frac{L_iV_j}{L_iV_j + L_i(1 - V_j) + A_iV_j + A_i(1 - V_j)}. \quad (16)$$

In this case, however, the denominator is simply equal to $L_i + A_i$, and this will, of course, therefore be the case for the other three phonemes as well. Thus, with the fuzzy logical model, the probabilities that a given speech sound will be identified to be /b/, /p/, /d/, or /t/ are given by the following equations:

$$p(\mathrm{b}|S_{ij}) = L_iV_j/(L_i + A_i), \quad (17)$$

$$p(\mathrm{p}|S_{ij}) = L_i(1 - V_j)/(L_i + A_i), \quad (18)$$

$$p(\mathrm{d}|S_{ij}) = A_iV_j/(L_i + A_i), \quad (19)$$

and

$$p(\mathrm{t}|S_{ij}) = A_i(1 - V_j)/(L_i + A_i). \quad (20)$$

Test of the Model

Unfortunately, very little of the previous experimental work on speech perception can be used to test the model. Most of these studies do not address the integration problem, since only a single acoustic dimension was used in a given experiment. There are a number of reasons that may explain why this procedure has been used almost exclusively. First, most formal linguistic representations of a given dimension are discrete rather than continuous (Jakobson et al., 1961). For example, a stop consonant is considered to be either completely voiced or completely voiceless rather than, say, .7 voiced and .3 voiceless. With such a binary representation, the integration of information from the place and voicing dimensions would simply be logical conjunction: The consonant /b/ is represented as voiced *and* labial, /t/ is voiceless *and* alveolar, and so on. Within this framework, the identification of sounds involving a number of dimensions would be expected to follow directly from the results of the relevant single-dimension experiments. For example, if the discrete feature hypothesis were correct, the results when a subject is asked to make voicing judgments within a particular place of articulation could be expected to generalize to the more natural situation in which the subject must integrate information across both voicing and place of articulation.

A second possible reason why only single-dimension experiments were carried out is that it is traditional in psychophysical research to vary a single dimension while holding all other dimensions constant. It is only recently that data reduction techniques such as analysis of variance have been used in this work. With the few factorial speech perception experiments that were done earlier, the data analyses were effectively reduced to single-dimensional analyses, since no analyses of interactions were performed (Harris et al., 1958; Hoffman, 1958). Thus, the critical information that might have shed some light on the integration problem was essentially left unused.

Massaro and Cohen (1976, 1977) were concerned with integration processes that take place prior to the integration of voicing and place of articulation information. Specifically, they addressed the question of how the various acoustic cues are integrated to arrive at a *single* acoustic phonetic distinction, such as the difference between voiced and voiceless sounds. Previous research has shown that voicing of initial stop consonants can be cued by a variety of acoustic features. These features include VOT; the presence versus absence of aspiration during VOT; the fundamental frequency (F_0) at the onset of vocal-cord vibration; the presence or absence of significant F_1 transitions at the onset of vocal-cord vibration; the frequency of F_1 at the onset of vocal-cord vibration; and the frequency, intensity, and duration of the aperiodic information in the release burst at the onset of the stop consonant.

Massaro and Cohen (1976) utilized a similar framework to the one presented here to study how two acoustic dimensions are evaluated and integrated in the perception of the single feature of voicing. Rather than varying just a single dimension, they simultaneously varied two or more dimensions through several values

in a factorial design. Listeners were asked to rate the degree to which the speech sound was heard as /si/ relative to /zi/. In one experiment, the stimuli were generated by crossing several levels of VOT with several levels of F_0. The stimuli were heard as more /zi/-like with decreases in VOT and decreases in F_0. The quantitative results were used to test the predictions of the model presented here. The assumption that acoustic dimensions were combined multiplicatively as in Equations 11–14 provided a significantly better description than the assumption of an additive combination. This experiment and others (Massaro & Cohen, 1977) provide solid support for the model in the domain in which multiple acoustic cues contribute to one phonetic distinction.

A study by Sawusch and Pisoni (1974) was one of the first to systematically vary voicing and place of articulation in order to examine the nature of the featural integration process. This experiment consisted of four parts that were run separately. In Part 1, all of the stimuli were voiced consonant syllables, but the acoustic cues to place of articulation were varied and the subjects identified the syllables as either /ba/ or /da/. In Part 2, all of the speech sounds were labial, but VOT was varied, and the subjects identified these syllables either as /ba/ or /pa/. In Parts 3 and 4, voicing and place of articulation were covaried from labial–voiced to alveolar–voiceless. With this technique, the sound was made more alveolar as it was made more voiced. In Part 3, these syllables had to be classified by the subjects as either /ba/ or /ta/; whereas in Part 4, the subjects were allowed to classify the syllables as /ba/, /da/, /pa/, or /ta/. On the basis of the results of this experiment, Sawusch and Pisoni rejected a simple additive feature model and proposed a more complex model including a cross-product term. This latter model provided a better account for the data.

Recently, Oden (in press) has shown that the fuzzy logical model provides an even better account for the data of Sawusch and Pisoni (1974). Of particular interest is the series of sounds for which both place and voicing were covaried and which was presented to the subjects twice, once to make a forced choice to identify each sound as either /ba/ or /ta/ and the other time to identify the sounds as either /ba/, /pa/, /da/, or /ta/. According to the fuzzy logical model, the feature evaluation and prototype matching operations should not change under these two conditions. All that should change, for example, for the probability of identifying a sound to be /ba/, is which terms are included in the denominator of the equation for the pattern classification operation. The fuzzy logical model was successful in describing the data of Sawusch and Pisoni's experiment, including the data for these different response conditions. However, while this experiment did vary both voicing and place of articulation, it does not provide a thorough test of the fuzzy logical model, since these dimensions were not independently varied.

Experiment

In order to adequately test the proposed fuzzy logical model, it is necessary to have subjects identify phonemes for which the degree of voicing and the degree of place of articulation are varied independently. In the present experiment, subjects identified the initial phoneme of synthesized consonant–vowel syllables as being either /b/, /p/, /d/, or /t/. Acoustic cues to voicing and to place of articulation were independently varied through several values.

Method

Stimuli. Each stimulus was a syllable of 320-msec duration consisting of a stop consonant followed by the vowel /ae/ as in "bat." The acoustic cues to the voicing and the place of articulation of the consonant part of the syllables were varied independently in a 5×7 factorial design. The five different degrees of voicing were produced by varying VOT in 10-msec steps from 0 to 40 msec. The seven different levels on the place-of-articulation dimension were produced by varying the frequencies at which the second and third formants (F_2 and F_3) began. The actual values for the various levels of this factor are listed in Table 1, and Figure 2 presents spectrograms of two of the stimuli.

For each syllable, the fundamental frequency (F_0) was 126 Hz and remained at this value throughout the syllable until a linear decrease to 112 Hz during the last 120 msec of the syllable. The first formant (F_1) started at 200 Hz and increased to 734 Hz in a negatively accelerated manner over the first 30 msec of the sound. During the first 70 msec, F_2 and F_3 increased or decreased (in a negatively accelerating manner), respectively, to reach the frequencies of F_2 (1,600 Hz)

and F_3 (2,851 Hz) of the vowel. The fourth and fifth formants (F_4 and F_5) were constant at 3,500 and 4,000 Hz, respectively.

The energy source for the initial part of the stop-consonant transition period depended on the VOT value. For a VOT value of 0 msec, the voicing source was turned on at the onset of the syllable and increased linearly to full amplitude in 20 msec. For VOT values greater than 0, the syllable began with the onset of aspiration, which served as the energy source for the F_1 through F_5 formants. The aspiration reached full amplitude instantaneously and remained on during the VOT interval. At the end of the VOT interval, the aspiration source was turned off with a linear fall time of 20 msec. Although VOT of a synthetic speech sound is defined to be the interval between the onset of the speech sound and the onset of the buzz source, the immediate rise time and the 20-msec fall time of the amplitude of aspiration mean that the perceived VOT interval was probably somewhat longer than the nominal value. Figure 2 shows the resulting transitions for 0- and 40-msec VOTs, respectively.

Procedure. On each trial a syllable was randomly selected without replacement from the 5 × 7 stimulus design. The stimulus was presented to the subject and followed by a 2-sec response interval, which ended with the onset of a 250-msec visual signal to the subject. The subject identified the stimulus as /b/, /p/, /d/, or /t/ and indicated his response by pushing an appropriately labeled button. The presentation of the next test stimulus followed the end of the previous response interval by 1 sec. On the first day, the subject was read the instructions and was given a practice session of a block of 35 trials. The subject was then asked if he had any questions, and his responses were checked to insure that he had responded on each practice trial. This was followed by two experimental sessions with roughly 10-minute breaks between sessions. On the second day, the subject was run through two more experimental sessions. Each session consisted of 10 blocks of the full set of 35 speech sounds in the stimulus design, for a total of 40 responses per subject to each of the 35 sounds. The experiment took about 1 hour each day, and each subject was tested on 2 consecutive days.

Table 1
Starting Frequencies (Hz) of F_2 *and* F_3 *Formants of the Seven Different Speech Sounds*

Stimulus	F_2	F_3
1	1,270	2,263
2	1,345	2,397
3	1,425	2,614
4	1,510	2,770
5	1,600	2,934
6	1,695	3,020
7	1,796	3,200

Note. Steady-state values for F_2 and F_3 are 1,600 and 2,851 Hz, respectively.

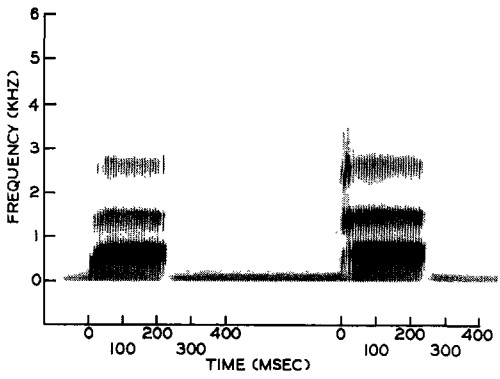

Figure 2. Spectrograms of two stimuli (a /ba/ and a /pa/).

Apparatus and subjects. All stimuli were produced on-line during the experiment by a formant series resonator speech synthesizer (FONEMA OVE-IIId) controlled by a PDP-8/L computer (Cohen & Massaro, 1976). The stimuli were specified as concatenations of steady-state and transition segments. Synthesizer control parameters (e.g., F_0) indicated at each segment boundary the parameter values that were to be changed for a given interval. Segment durations were always multiples of 10 msec. Intermediate values within a segment were computed with linear or nonlinear interpolation as appropriate and were output to the synthesizer every 10 msec. The output of the speech synthesizer was amplified with a McIntosh MC-50 amplifier and presented over headphones (Koss Model 4AA) at a comfortable listening intensity (about 76 dB SPL). Four subjects could be tested simultaneously in separate sound-attenuated rooms.

Sixteen subjects served in the experiment. The subjects were solicited from the University of Wisconsin community and were paid $4 for their participation.

Results and Discussion

Figure 3 presents the data from this experiment. Each panel gives the data for a given level of VOT; and within each panel, the four curves give the identification probabilities for the four phonemes. As can be seen in Figure 3, the shapes of these curves change markedly but in a systematic fashion from panel to panel. The total pattern of data presented in all five panels provides the important information about how place and voicing information is integrated. However, this manner of presenting the data makes it difficult to determine how well the model fits the data.

Figure 4 plots the same data with separate panels for each of the four phonemes. This

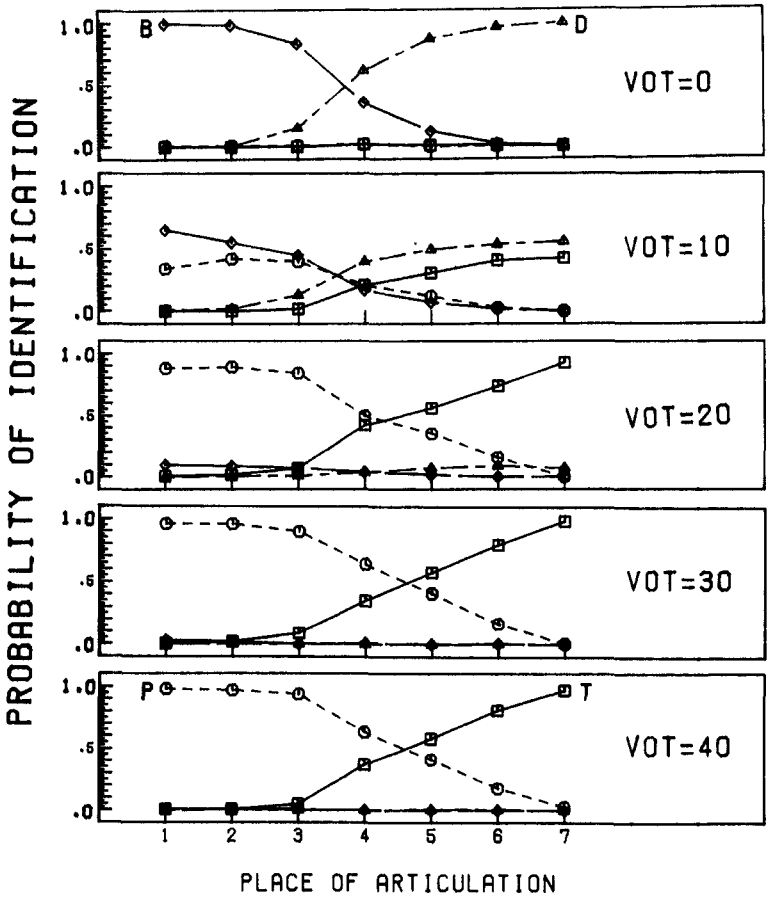

Figure 3. Identification probabilities for each response alternative for each speech sound. (Each panel presents the data for a given level of voice onset time [VOT]. Diamonds, triangles, circles, and squares represent the data for /b/, /d/, /p/, and /t/, respectively.)

figure also gives the predictions of the fuzzy logical model when fitted to the data. In this and the following figures of this type, the data are the points and the predictions are represented by the curves. In this experiment, the predictions are obtained by fitting the model separately for each individual subject and then averaging these predictions over subjects. Thus, the data for each subject are fitted individually, and both the data and the predictions in this figure are averaged over all 16 subjects. In each panel of the graph, the spacing of the levels along the abscissa is proportional to the spacing of the marginal means across the seven levels of the formant transitions. This spacing was computed separately for each of the four response types and then averaged over response types, so that the spacing along the abscissa is the same for all four panels. Spacing the levels of the abscissa in this way allows the pattern of the predictions of the model to be more easily seen.

To fit the model to the data of each subject, the computer subroutine STEPIT (Chandler, 1969) was used. This subroutine iteratively adjusts the values of the parameters until it finds that set of values which results in predictions of the model that come closest to fitting the data. *Closeness of fit* was defined in terms of the sum of the squared deviations of the data from the model. Fitting the model required 12 parameters: 5 to specify the degree of voicing for each level of VOT and 7 to specify the degree of labiality for each level of the place-of-articulation factor.

As can be seen in Figure 4, the model pro-

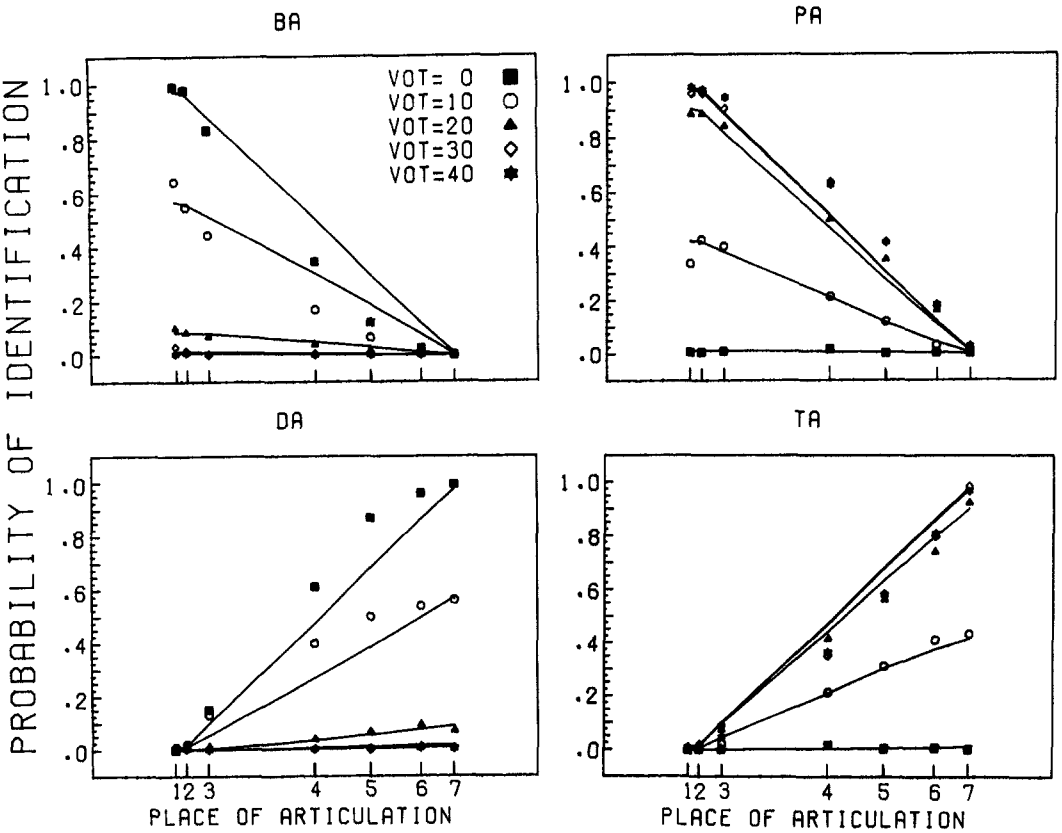

Figure 4. Identification probabilities and predictions of the simple fuzzy logical model. (Each panel presents the data for a given response alternative. Note that the spacing along the abscissa is proportional to the spacing of the subjective place values. VOT = voice onset time.)

vides a general account of the data but deviates systematically from the data. Not many of the deviations are very large and the grand root mean squared deviation of the model from the data is .092, which is fairly small considering that 105 independent data points are fitted for each subject using only 12 free parameters. Of the 140 data points, only 105 are independent because the four response probabilities must sum to one for each of the 35 separate stimuli. In this experiment, for the range of stimuli used, it was assumed that a sound was perceived to be alveolar to the degree that it was not perceived to be labial, that is, $L_i = 1 - A_i$. (Relaxing this assumption had virtually no effect on the fit of the model.) Accordingly, it was only necessary to estimate the seven values of A_i to account for the place features.

Because of the large number of degrees of freedom (93) for this test of the model, the relatively good fit can be taken as general support of the fuzzy logical model. However, it is possible that other models with an equivalent number of parameters might do just as well. Accordingly, an alternative model based on additive rather than multiplicative feature combinations was fitted to the data using the identical procedures. In this model, the matching functions given in Equations 11–14 are replaced by the following:

$$B(S_{ij}) = L_i + V_j, \tag{21}$$

$$P(S_{ij}) = L_i + (1 - V_j), \tag{22}$$

$$D(S_{ij}) = A_i + V_j, \tag{23}$$

and

$$T(S_{ij}) = A_i + (1 - V_j). \tag{24}$$

The rest of the model remains the same. Thus, this alternative "additive feature integration"

model uses exactly the same number of parameters as with multiplicative feature integration and, therefore, is equivalent in simplicity and power. Nevertheless, it was unable to provide a satisfactory account of the data as is reflected in its root mean squared deviation of .247.

The performance of the additive feature model indicates that the substantially better fit of the comparable fuzzy logical model should be taken as support for this latter model. However, despite this success, the fact that of the deviations of the data from the fuzzy logical model those which are of any size are clearly systematic, both within and between panels of Figure 4, indicates that there are important effects that are left unaccounted for by the model as formulated so far. Therefore, a more complex version of the model was developed in an attempt to provide a more complete account of the data.

Featural modification in the phoneme prototypes. The simple version of the fuzzy logical model assumes that each feature is treated the same for each of the alternative phonemes that have that particular feature. This is the simplest case and is what is implicitly assumed in discrete feature theories. However, there are reasons to suppose that, in fact, some of the features may be expected on the average to take on more extreme values for some phonemes than for others. For example, the typical voice onset time of voiceless stop consonants produced under natural conditions is longer for alveolar than for labial stop consonants (Lisker & Abramson, 1970). Consequently, it may be that the actual subjective prototype in long-term memory incorporates information about the necessary extremity of the various features for the idealized phoneme. For example, the prototypes for /b/ and /d/ might more accurately be defined as

/b/: (LABIAL) AND [VERY (VOICED)] (25)

and

/d/: (ALVEOLAR) AND
[MODERATELY (VOICED)]. (26)

These modifiers do not mean that a *perfect* /b/ is now considered to be any more voiced than is a *perfect* /d/. What the modifiers do signify is that extremity on these features is more important for /b/ than for /d/; that is, that with Equation 25, the goodness of the match to the ideal /b/ falls off more rapidly as the speech sound becomes less voiced.

If such modifiers are psychologically real, then they must somehow be allowed for within the model. The problem is how the modifiers are manifested in the matching functions of the prototypes. Zadeh (1972, 1975) has proposed that modifiers of this sort should be represented as power functions. For example, Zadeh suggests that "very" may be defined as

$$t[\text{VERY}(A)] = t(A)^2, \tag{27}$$

where A is any proposition. In the general case, modifiers (mod) of this sort will be represented by exponents:

$$t[\text{mod}(A)] = t(A)^q, \tag{28}$$

where q may take on any value depending on whether the actual subjective modifier corresponds to "very," "extremely," "moderately," "somewhat more than moderately," or perhaps some modifier for which we have no common phrase.

Thus, in the general form, the matching function for /b/ may be represented as

$$B(S_{ij}) = (L_i)^q(V_j)^r, \tag{29}$$

where the subscripts on L_i and V_j indicate that they vary as the speech sound (S_{ij}) varies; whereas, in contrast, the exponents q and r have no subscripts, since they are constant for a given phoneme and over all speech sounds.

The basic nature of the pattern classification operation of the model remains unchanged with the addition of these modifiers. The probability of identifying a syllable to be /b/ will still follow from Equation 15. Of course, when Equation 15 and the equations for the other phonemes are expanded by inserting the equations for the matching functions, the resulting formulae will be much more complex than Equations 17 through 20. Despite this greater mathematical complexity, however, the complex fuzzy logical model is conceptually very nearly the same as the simple fuzzy logical model. The only substantive change is the addition of the modifiers to the prototypes. All of the other changes in the equations are superficial and follow directly given the structure of the psychological model.

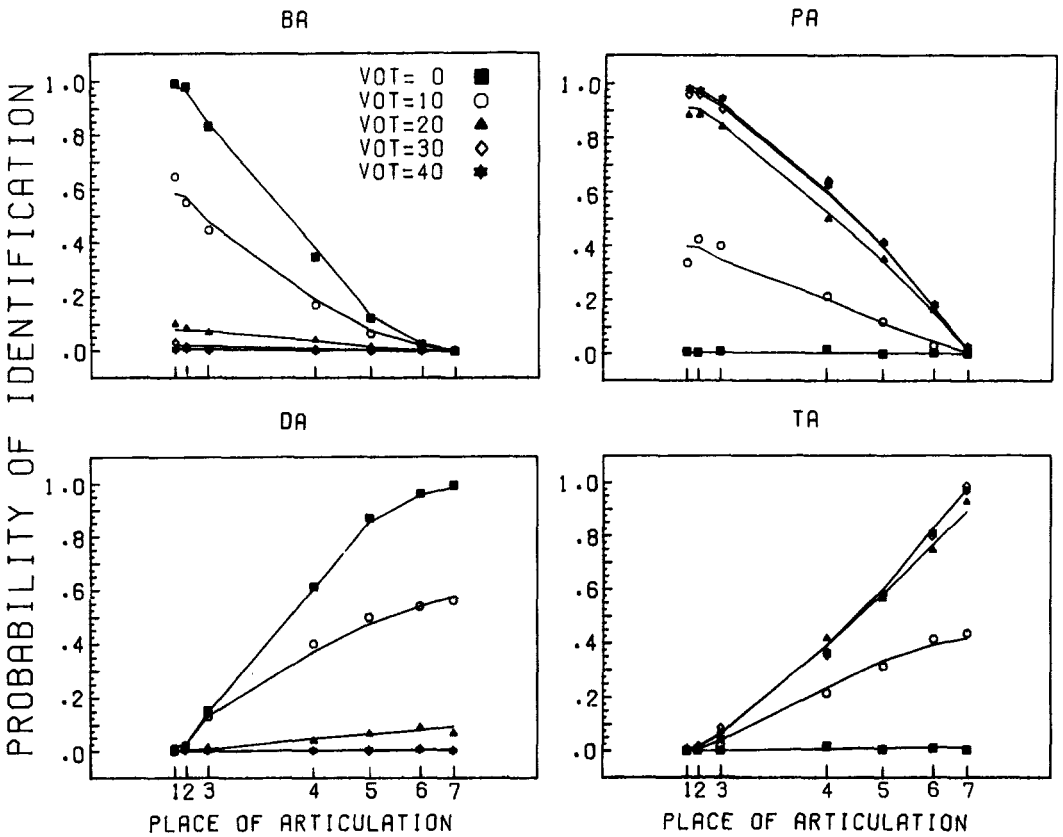

Figure 5. Identification probabilities and predictions of the complex fuzzy logical model. (Note that the spacing along the abscissa is proportional to the spacing of the subjective place values. VOT = voice onset time.)

To determine whether the observed results could be better described by the addition of modifiers to the prototype definitions, the complex version of the fuzzy logical model was also fitted to the data of this experiment. The complex version of the model required an additional 8 parameters for the modifier exponents, for a total of 20. Figure 5 presents the data along with the predictions of the complex fuzzy logical model. As is clear from this figure, the complex version of the model provides a very close fit to the data. This conclusion is also evident from the grand root mean squared deviation of .039 for this model. Thus, the data provide very strong support for the fuzzy logical model of phoneme identification and also for the necessity of including modifiers in the definitions of the phoneme prototypes.

Again, it is useful to compare the performance of the complex fuzzy logical model with that of other models with the same number of parameters. For example, featural weighting can also be included in the additive featural model. In the case of additive combination rules, it is most natural to make the weights multiply their respective features, such as

$$B(S_{ij}) = qL_i + rV_j. \tag{30}$$

The addition of weights improved the fit of the additive feature model, resulting in a root mean squared deviation of .161. This is, of course, not nearly as good a fit as the complex fuzzy logical model, even though it uses the same number of parameters. In fact, this weighted additive features model still does not account for the data as well as the simple fuzzy logical model.

It might be thought that it is the exponents per se that allow the complex fuzzy logical model to provide such a good fit to the data.

Table 2
Parameter Estimates for the Complex Fuzzy Logical Model: Average Values of Phoneme Prototype Modifiers

Phoneme	Dimension	
	Voicing	Place of articulation
/b/	2.44	2.79
/p/	2.37	1.04
/d/	2.04	1.69
/t/	1.95	2.21

However, when exponential weights are incorporated into the additive feature model to produce matching functions such as

$$B(S_{ij}) = (L_i)^q + (V_j)^r, \quad (31)$$

the root mean squared deviation is .230, which is not even as good a fit as with the multiplicatively weighted additive feature model.

One more model was fitted to the data. This was a version of the complex fuzzy logical model, but instead of using 12 parameters to represent the five degrees of voicing and the seven degrees of labiality, these values were obtained as simple functions of the levels of the corresponding factor of the design. Thus, for example, the degrees of labiality were given by the following equation:

$$L_i = \frac{x_i^c}{x_i^c + (1 - x_i)^c}, \quad \text{where} \quad c \geq 1, \quad (32)$$

which is ogival in form when the x_i parameters are constrained to fall between zero and one. The larger the c parameter, the steeper is the middle section of the ogive. The x_i values were obtained by a simple linear function of the level of the factor, that is,

$$x_i = ai + b, \quad (33)$$

except that values less than zero were set to zero and values greater than one were set to one. Taken together, Equations 32 and 33 specify an ogival type of curve in three parameters: a, b, and c. The first two allow the ogive to shift linearly along the place factor to position the boundary at the proper place and the third determines the sharpness of the boundary. A similar three-parameter ogival function was used to obtain the degree of voicing values from the levels of the voicing factor. Thus, altogether there were three place parameters, three voicing parameters, and eight exponential weight parameters for a total of 14 and a saving of six from the original complex fuzzy logical model.

Despite this considerable decrease in the number of parameters, this version of the model fit the data nearly as well as the full 20-parameter version (root mean squared deviation = .045). It is especially interesting to note that although this version of the model requires only two more parameters than the simple fuzzy logical model, the fit is still more than twice as good in terms of the root mean square criterion. Thus, the point is reemphasized that the success of the model is not simply due to the number of parameters used, but rather to the fact that the parameters are combined in a way that captures the structure of the underlying psychological processes.

Details of the effects of featural modification. It is interesting to note in Figure 5 the effect of including the exponents corresponding to the modifiers in the prototypes. In contrast to the curves representing the model predictions in Figure 4, those in Figure 5 are not diverging fans of straight lines. Rather, these curves display a marked bowed shape, whose direction depends on the actual modifiers of a given phoneme prototype relative to those for the other phonemes. The values obtained for the exponents corresponding to these modifiers are given in Table 2. The parameters for voicing and degree of alveolarity are shown in Table 3. These values are averages of the values obtained for each subject that were used in fitting the model to the subject's data.

Table 3
Parameter Estimates for the Complex Fuzzy Logical Model: Average Degrees of Voicing and Alveolarity for Levels of Stimulus Design

Level of voicing factor	Degree of voicing	Level of place factor	Degree of alveolarity
1	0	1	.01
2	0	2	.06
3	.01	3	.18
4	.57	4	.52
5	.93	5	.76
		6	.99
		7	1.00

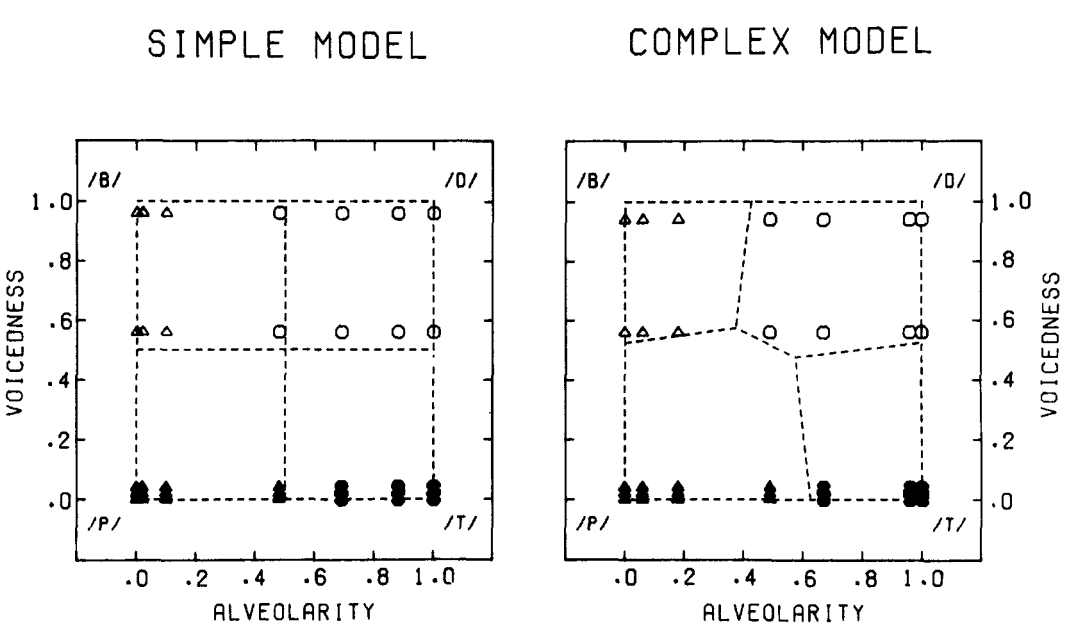

Figure 6. Psychological parameter spaces for the simple and complex fuzzy logical models. (The dashed lines partition the parameter spaces into regions in which a given response alternative is predicted by the respective model to be most likely. Each point represents the subjective position of a stimulus and specifies what phoneme it was most often identified to be, with open triangles, closed triangles, open circles, and closed circles standing for /b/, /p/, /d/, and /t/, respectively.)

Another useful way to represent the effect of the exponents in the complex fuzzy logical model is given in Figure 6. The two panels of this figure each represent the *psychological* parameter space of the stimuli in terms of the two critical acoustic dimensions. The four corners of each square represent the ideal points for the four phonemes. It should be stressed that these diagrams represent the psychological, not the physical, parameter space. For example, the vertical dimension represents the subjective degree of voicing not some arbitrary physical measure such as VOT.

In each square of Figure 6, the dashed lines indicate the boundaries separating those stimuli most likely to be identified as one phoneme from those most likely to be identified as another phoneme. The left panel of the figure gives the partitioning of the parameter space as predicted by the simple version of the model. In this case, if the value for alveolarity is greater than .5, the stimulus will be identified more often as alveolar no matter what the value of voicing and so on. The right panel gives the partitioning as predicted by the complex version of the model. Here, the effects of the modifiers are clear: The large exponents associated with extreme modifiers such as "very" or "quite" cause a restriction of the region in which the stimuli are identified to be instances of that phoneme. In the description of the present results, the alternatives /b/ and /t/ now include less of the total parameter space.

The panels of Figure 6 also show in another fashion that the complex version of the model provides a superior account of the data. The stimuli used in the experiment are shown in each panel of the figure as open and filled circles and triangles. These points are positioned in their respective positions in the parameter space as determined from the parameter values used in fitting the respective version of the model to the data.[1] The points classify each stimulus according to the phoneme it was

[1] Note that the parameter values for place and for voicing are not the same for the two versions of the model (see Figure 6). This is because with the simple model, these parameters are influenced by the effects in the data that are accounted for in the complex model by the exponents.

most often identified to be. Of particular interest is the fourth level of alveolarity, counting from the left of the figure. For this level, the two most voiced stimuli were most often identified as /d/, the *alveolar voiced* phoneme; whereas the other three stimuli were most often identified as /p/, the *labial unvoiced* phoneme. This change in the boundary along the place dimension as a function of voicedness may also be seen in Figure 3.

Miller (1977) and Repp (Note 2) also found large differences in the place boundary for voiced and voiceless labial and alveolar stop consonants. The changeover from labial to alveolar responses shifts toward the alveolar end as the speech sounds are made more voiceless. This kind of boundary change clearly cannot be accounted for by either noninteractive, discrete feature theories or by the simple fuzzy logical model. However, as the right-hand panel of Figure 6 shows, this effect is nicely accounted for by the complex fuzzy logical model. It is simply a natural manifestation of the modifiers in the phoneme prototypes. Thus, it is additional evidence in support of this model that it is able to provide a good account for this "phonetic boundary shift" and that it is able to do so without having to resort to explanations about complex feature interactions.

The three-dimensional graphs in Figure 7 illustrate this same partitioning of the parameter space in a more complete and less abstract fashion. These are graphs of the degree to which each of the four phoneme prototypes matches each possible stimulus specified in terms of subjective values. To make the graphs more legible, only the upper part of each matching function, where it provides a better match to the stimuli than do the alternative matching functions, is shown. That is, the matching function of a given phoneme is shown only for the region of the parameter space where that phoneme is predicted to be preferred. The left-hand three-dimensional graph illustrates the characteristics of the simple model: The matching functions are identical in shape for each of the phonemes and increase to the optimal point at each corner in a relatively gradual manner. In contrast, the right-hand graph shows that with the complex model, some of the functions increase much more steeply than others and, in general, more steeply than with the simple

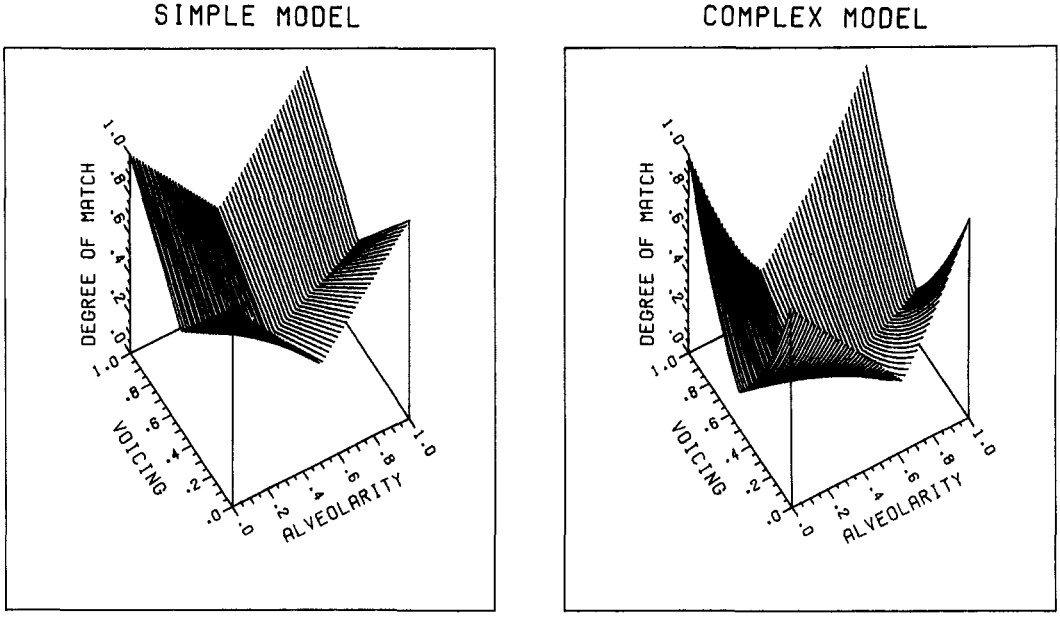

Figure 7. Hypothetical matching functions for the simple and complex fuzzy logical models. (The matching function for each phoneme specifies the degree to which the prototype for that phoneme matches each possible speech sound. Each matching function is shown only for that part of the parameter space where the corresponding phoneme is predicted to be the most frequent response.)

model. Note also that the places where the functions intersect, that is, the bottom of the "valleys," are the boundaries that were presented in Figure 6.

General Discussion

Related Approaches to Pattern Identification

One of the most important conceptual forebears of the fuzzy logical model is Selfridge's (1959) Pandemonium model. Selfridge proposed four levels of processing, the lowest of which, that consisting of "data demons," was simply the level of reception and sensory storage. The other three levels correspond directly to the three stages of the fuzzy logical model: The "computational demons" perform feature evaluation, the "cognitive demons" do kinds of featural integration that are analogous to the present prototype matching, and the "decision demon" chooses among the alternatives just as is done in the pattern classification stage of the fuzzy logical model. Pandemonium also shares with the present model the characteristics of parallel processing within each level of computation, of logical combination rules for featural integration, and of featural evaluation processes that result in information about the *degree* to which features are present in the stimulus. Selfridge gave particular emphasis to this latter aspect of Pandemonium and illustrated the concept with diagrams (Selfridge, 1959, p. 525) that are strikingly similar to those of fuzzy predicates that appear in the fuzzy logic literature (e.g., Zadeh, 1972).

However, whereas the fuzzy logical model is intended to describe the cognitive processes that are actually used by humans to identify speech sounds, Selfridge (1959) was primarily interested in the problem of learning to make correct identifications. Thus, Selfridge concentrated on how Pandemonium might be made to come to discover the appropriate features and feature integration rules over the course of training trials. In contrast, the present article relied on intuition and on the analytic linguistic description of the phonemes to formulate the corresponding prototype specifications.

Another closely related approach is that of Morton (1969), who developed a model of word identification based on signal detection theory (Green & Swets, 1966). In this model, the evidence for each candidate word is accumulated in parallel; for each word, the accumulation is compared against a criterion that is determined in part by the degree to which that word is expected. While Morton does not explicitly consider how the information from separate features is put together to determine the evidence for a given word, the obvious assumption is that each feature contributes independent evidence that is simply accumulated along with the rest. With this assumption, Morton's model leads to equations of the same form as those of the simple version of the fuzzy logical model.

In the area of speech perception, the models of Sawusch and Pisoni (1974) and of Repp (1977) are similar in several respects to the fuzzy logical model. Sawusch and Pisoni propose a model with three stages: auditory analysis that results in a set of acoustic cues, acoustic cue combination that produces the phonetic features such as those of place and voicing, and phonetic feature combination that results in the final identification. The phonetic featural information is considered to be continuous proportions of the nominal features. Sawusch and Pisoni propose and test a number of rules for combining the phonetic features. In all of these rules, the proportions are weighted by multiplicative weights and then combined to produce the predicted response probabilities directly. The combination rules that were considered were either strictly additive or else contained cross-product terms added in with the individual featural information. Of most importance is that all of these rules treat the featural information as continuous rather than discrete.

Repp's (1977) model of featural integration in speech perception is based on a spatial representation of speech sounds. In this model, the features correspond to dimensions in a euclidean space. Thus, both the prototypes and the stimulus can be considered to be points in this space, and the degree of match of each prototype to the stimulus is, therefore, an inverse function of the distance from prototype to stimulus. Thus, this model is identical to the fuzzy logical model in overall structure in that it considers identification to be a

process of choosing the phoneme with the highest degree of match. However, Repp's model differs from the present model primarily in its use of the euclidean distance function to describe the combined influence of the features.

Feature Independence

The assumption of independent acoustic features has been supported for a variety of speech stimuli. Not all acoustic manipulations will result in independence, however. Massaro and Cohen (1977) manipulated independently the duration of the frication period and the amplitude of vocal-cord vibration during the frication period on a /si/ to /zi/ continuum. Intuitively, these manipulations might not be expected to have independent effects, since other experiments have shown that the cue value of frication duration depends on whether or not vocal-cord vibration is present. Increasing the duration of frication without vocal-cord vibration makes the sound more voiceless, whereas frication duration with vocal-cord vibration has very little effect on perceived voicing. Accordingly, it seems reasonable to assume that the acoustic feature to voicing is the composite sound spectrum composed of both the high frequency noise and the harmonic energy produced by vocal-cord vibration. A version of the fuzzy logical model based on this single-composite feature idea was contrasted with one based on independent acoustic features. The description of the data was better for the composite features model than for the independent features model. This result shows two things: First, the independent features model is not too general, that is, it can be disconfirmed; second, a logical analysis of the stimulus situation is helpful in understanding how various acoustic manipulations will be processed in terms of acoustic features.

Phoneme Versus Syllable Prototypes

We have assumed that the relevant prototypes in LTM correspond to consonant–vowel syllables rather than stop-consonant phonemes. Given that no test of this assumption is possible in the present experiment, it is important to consider other sources of evidence. Acoustic cues to consonant phonemes depend critically on vowel context, and any normalization process that depends on vowel context is not easily handled by phoneme–unit models. In contrast, the necessary normalization is easy to build into syllable prototypes in LTM. Consider the classic example of the large differences in the second formant transitions in the stop consonant syllables /di/ and /du/. The second formant rises from approximately 2,200 Hz to 2,600 Hz in /di/, whereas it falls from approximately 1,200 Hz to 700 Hz in /du/. The second formant then remains relatively constant at these values during the steady-state vowel.

This example makes it apparent that the stop consonant /d/ cannot be invariantly defined by a phoneme prototype in LTM. However, the problem is easily solved by consonant–vowel syllable prototypes. Simplifying the situation, assume that two ordered features, the onset frequency of the second formant and its steady-state value, are sufficient acoustic features for distinguishing the syllables /di/ and /du/. In this case, /di/ would be defined as having a rising transition and a high steady-state vowel, whereas /du/ would be defined as having a falling transition and a low steady-state vowel. Consonant–vowel prototypes, then, solve the gross problem of the lack of acoustic invariance of stop consonants.

Another problem with phoneme prototypes is that the vowel sometimes provides direct acoustic cues to the identity of the consonantal portion of the syllable. As an example, vowel duration has a large influence on the perception of voicing of a vowel–consonant syllable in word-final position. Denes (1955), for example, 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 independently varied with five durations of frication taken from real speech. No vocal-cord vibration was present during the frication period. The results showed that the proportion of voiceless responses decreased with increases in vowel duration and increased with increases in frication duration. Massaro and Cohen (1977) provided a quanti-

tative description of these results in terms of the simple fuzzy logical model. It was assumed that vowel duration and frication duration are perceived independently and combined multiplicatively as in the fuzzy logical model. The good fit of the model and the meaningful parameter estimates simultaneously support the fuzzy logical model and the accompanying assumption of syllable prototypes in LTM.

Prototype modifiers. The inclusion of modifiers in the prototypes represents a fairly large deviation from the traditional way of thinking about phonemes. However, it would seem to be unlikely that there is no interaction at all between the articulations associated with the various acoustic dimensions. For voiced stops, for example, VOT increases as the place of articulation moves from labial to alveolar to velar points of closure. The difference in VOT follows directly from the time course of the pressure developed across the oral closure following release (Klatt, 1975). The release period is longer for velars than for labials, since the velar release involves the whole tongue body, whereas only the lips move in labials. Given that there are effects of this sort, then it also would seem most reasonable for the listener to come to expect some voiced phonemes to be subjectively more voiced than others. It is this kind of knowledge that the phoneme prototype modifiers represent.

An additional phenomenon that the phoneme prototype modifiers may be able to help interpret is the learning of dialects. It is, at first, difficult to understand people whose dialects are strongly different than one's own. However, after a period of listening, it becomes much easier and automatic. This process of "educating your ear" might be a matter of changing the modifiers on various phonemes, that is, restructuring the prototypes of perceptual units in long-term memory.

Phonetic boundary changes. Previous work (e.g., Lisker & Abramson, 1970) has found evidence for changes in the voicing boundary as a function of place of articulation. This result was obtained in the present experiment in terms of the quantitative predictions of the complex fuzzy logical model. The present experiment also obtained evidence of the same sort in support of the existence of changes in *place* boundaries as a function of *voicing*. In addition, these latter changes were also indicated by a qualitative effect: The crossover point between labial and alveolar phonemes occurred between different levels of the place factor depending upon the particular value of VOT.

Such boundary changes are sometimes (e.g., Haggard, 1970) taken to be evidence of interaction in the perception of the acoustic features themselves. However, once it is recognized that the qualities of place and voicing are continuous, then featural interaction becomes potentially an infinitely complex problem. If we had to allow arbitrary interaction between the feature evaluation operations, so that the perception of one feature depended in an idiosyncratic way not only on the value of its own specific acoustic cues (which under natural conditions are highly correlated) but also on all of the other cues that might be varied independently, then the task of specifying how phoneme identification takes place would be even more formidable than it is at present. Happily, as the present article has demonstrated, the existence of changes in the boundaries need not lead us to accept the featural interaction hypothesis. Rather, it appears that the acoustic featural information is obtained independently but combined together by an integration rule of a form that produces the overall observed interaction. In fact, the success of the complex fuzzy logical model in accounting for the data, including all of the phonetic boundary changes, while still maintaining complete noninteraction of feature evaluation, may be taken to be strong evidence for the independence of acoustic feature perception during phoneme identification.

Conclusions

The following three main conclusions may be reached from the present work:

1. The fuzzy logical model provides a good description of the processes used in integrating information about voicing and place of articulation during phoneme identification.

2. Some phonemes require more extreme values on one or both acoustic dimensions than do other phonemes, and therefore, phoneme prototype definitions must allow for modifiers.

3. There are changes in the voicing boundary

as a function of place and also changes in the place boundaries as a function of voicing. However, these effects do not require that there be any interaction in the perception of the acoustic features but rather may result simply from the nature of the prototype representations of the speech sounds in long-term memory.

Reference Notes

1. McNabb, S. D. *Using confidence ratings to determine the sensitivity of phonetic feature detectors.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1976.
2. Repp, B. H. *Interdependence of voicing and place decisions.* Unpublished manuscript, Haskins Laboratories, New Haven, Conn., September 1977.

References

Abramson, A. S., & Lisker, L. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, *1*, 1–8.

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: W. H. Freeman, 1974.

Barclay, J. R. Non-categorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, *11*, 269–273.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961–970.

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 & Row, 1968.

Cohen, M. M., & Massaro, D. W. Real-time speech synthesis. *Behavior Research Methods and Instrumentation*, 1976, *8*, 189–196.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 769–773.

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

Goguen, J. A. The logic of inexact concepts. *Synthese*, 1969, *19*, 325–373.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Haggard, M. P. The use of voicing information. *Speech Synthesis and Perception*, 1970, *2*, 1–15.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. *Journal of the Acoustical Society of America*, 1958, *30*, 122–126.

Hoffman, H. S. Studies of some cues in the perception of the voiced stop consonants. *Journal of the Acoustical Society of America*, 1958, *30*, 1035–1041.

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. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, *18*, 686–706.

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

Liberman, A. M., Delattre, P., & Cooper, F. S. Distinction between voiced and voiceless stops. *Language and Speech*, 1958, *1*, 153–167.

Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, *20*, 384–423.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences* (1967). Prague, Czechoslovakia: Academic, 1970.

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

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. The contribution of voice-onset time and fundamental frequency as cues to the /zi/–/si/ distinction. *Perception & Psychophysics*, 1977, *22*, 373–382.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, *26*, 338–352.

Miller, J. L. Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research*, 1977, *20*, 519–528.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165–178.

Oden, G. C. Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 565–575.

Oden, G. C. Integration of place and voicing information in the identification of synthetic stop consonants. *Journal of Phonetics*, in press.

Pisoni, D. B., & Lazarus, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, *55*, 328–333.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, *15*, 285–290.

Repp, B. H. Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of*

Experimental Psychology: Human Perception and Performance, 1977, *3*, 37–50.

Sawusch, J. R., & Pisoni, D. B. On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, 1974, *2*, 181–194.

Selfridge, O. G. Pandemonium: A paradigm for learning. In D. Blake & A. Utteley (Eds.), *Symposium on the mechanization of thought processes*. London: H. M. Stationary Office, 1959.

Smith, P. T. Feature-testing models and their application to perception and memory for speech. *Quarterly Journal of Experimental Psychology*, 1973, *25*, 511–534.

Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, *34*, 273–286.

Zadeh, L. A. A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 1972, *2*, 4–34.

Zadeh, L. A. The concept of a linguistic variable and its application to approximate reasoning (II). *Information Sciences*, 1975, *8*, 301–357.

Received July 16, 1977 ■