

Sound to Representation:
An Information-Processing Analysis

DOMINIC W. MASSARO

University of Wisconsin

Whorf (1956) claimed that speech is the greatest show people put on and we are gathered here as part of our effort to gain some insights into a few acts of this great show. The information-processing approach provides a worthwhile analytic tool in the study of language processing. The goal is to dissect the complex show into its component parts. With this analysis, we hope to understand each of the parts and how they interact together in communication. The study of speech perception has been greatly influenced by the information processing approach in the last decade. According to the prototypical information-processing model, speech perception begins with the language stimulus and involves a sequence of internal processing stages before understanding occurs. The processing stages are logically successive although they overlap in time. Each stage of information processing operates on the information that is available to it and makes this transformed information available to the next stage of processing.

The speech stimulus consists of changes in the atmospheric pressure at the ear of the listener. The listener is able to experience the continuous changes in pressure as a set of discrete percepts and meanings. Our goal is to analyze the series of processing stages that allow this impressive transformation to take place. Figure 1 presents the model we have used to describe the temporal course of speech perception (Massaro, 1975a, 1975b, 1978, 1979, Oden & Massaro, 1978). At each

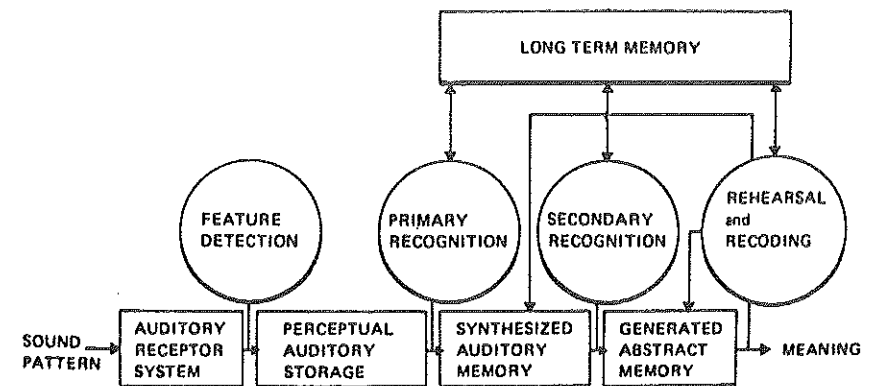


FIGURE 1 Schematic diagram of the general auditory information processing model.

stage the system contains structural and functional components: The structural component represents the information available at a particular stage of processing. The functional component specifies the procedures and processes that operate on the information held in the corresponding structural component. The model distinguishes four functional components; feature detection, primary recognition, secondary recognition, and rehearsal-recoding. The corresponding structural components represent the information available to each of these stages of processing. The present paper will provide a formal development of feature detection and primary recognition processes within this information-processing framework.

The changes in sound pressure set the eardrums in motion and these mechanical vibrations are transduced into a set of neural impulses. The neural impulses have a direct relationship to the changes in mechanical vibrations. We call the transformation from mechanical to neural information, feature detection and evaluation. The complex signal in the form of continuous changes in vibration pattern is transformed into a relatively small set of acoustic features. Features do not have to be relatively primitive such as the amount of energy in a particular frequency band, but they may include information about the duration of sound or silence, the rate of intensity change, and the rate of frequency change. It would be possible, for example, to have a feature detector that responds to the rising first formant transition that is characteristic of the class of stop consonants.

One traditional concern in speech research has been to determine the acoustic features that are utilized in perception. In terms of the model shown in Figure 1, the feature detection process places features in a brief temporary storage called preperceptual auditory storage, which holds information from the feature detection process for about 250 msec. The primary recognition process integrates these features into a synthesized percept which is placed in synthesized auditory memory. One critical question is what features are utilized and a second important question is how are all of the features integrated together. Does the listener only process the least ambiguous feature and ignore all others, or are the features given equal weight, and so on? Despite the overwhelming amount of research on acoustic features, very little is known about how the listener puts together the multitude of acoustic features in the signal in order to arrive at a synthesized percept.

The integration of acoustic features has not been extensively studied for two apparent reasons. The first is that research in this area was highly influenced by linguistic descriptions of binary all-or-none distinctive features (Jakobson, Fant, & Halle, 1951). Given the assumption of a binary representation of distinctive features the integration of information from two or more dimensions would be a trivial problem. Integrating binary features from voicing and place of articulation, for example, could be carried out by simple logical conjunction. If the consonant /b/ were represented as voiced and labial and /p/ were represented as voiceless and labial, the identification of voiced labial sound would be /b/ whereas the identification of a voiceless labial sound would be /p/. The simplicity of the logical conjunction of binary features may have discouraged psychologists and linguists from the study of the integration of acoustic featural information.

A second reason for the neglect of the integration problem is methodological. The primary method of study involved experiments in which the speech sound was varied along a single relevant dimension. For example, in a study of voicing all voicing cues were made neutral except one, such as voice onset time and then this dimension was varied through the relevant values. Similarly, place of articulation was studied by neutralizing all cues but one, and then varying the remaining dimension through the appropriate values. Very few experiments independently varied both voicing and place cues within a particular experiment. Those experiments that did (Hoffman, 1958) essentially reduced the data analyses to those of single dimension

experiments. Therefore, little information was available about how these cues were integrated into a synthesized percept.

More recently, we have initiated a series of experiments that are aimed more directly at the study of the integration of acoustic features in speech perception (Massaro & Cohen, 1976, 1977; Oden & Massaro, 1978). In contrast to the traditional linguistic description, we assume that the acoustic features held in preperceptual auditory storage are continuous, so that a feature indicates the degree to which the quality is present in the speech sound. Rather than assuming that a feature is present or absent, it is necessary to describe a feature as a function of its degree of presence. This assumption is similar to Chomsky and Halle's (1968) distinction between the classificatory and phonetic function of distinctive features. The features are assumed to be binary in their classificatory function, but not in their phonetic or descriptive function. In the latter, features are multivalued representations that describe aspects of the speech sounds in the perceptual representation. Similarly, Ladefoged (1975) has also distinguished between the phonetic and phonemic level of feature description. A feature describing the phonetic quality of a sound has a value along a continuous scale whereas a feature classifying the phonemic composition is given a discrete value. In our framework, the continuous features in preperceptual storage are transformed into discrete percepts in synthesized memory by the primary recognition process.

There now appears to be some evidence that listeners can transmit information about the degree to which a given acoustic feature is present in a speech sound (Barclay, 1972; Pisoni & Tash, 1974). Miller (1977a) asked listeners to monitor one ear during a dichotic presentation of a voiced stop to the monitored ear and a voiceless stop to the unmonitored ear. The voice onset time (VOT) of the voiceless stop significantly affected the identification of the voiced stop to the monitored ear; the likelihood of a voiceless response increased systematically with increases in the VOT values of the stop presented to the unmonitored ear. In addition, adaptation with voiceless stops decreased voiceless responses as a direct function of the VOT of the adapting stimulus. Miller (1977a) interpreted these results to indicate that the output of the feature detector for VOT is a graded signal whose magnitude is a direct function of VOT.

Although Miller's results are consistent with the idea of continuous or multivalued outputs of feature detectors, they do not disprove the possibility of all-or-none outputs. The findings of a relatively continuous identification function as a function of some stimulus property does not distinguish between the kinds of feature outputs. As assumed by the original categorical perception model (Lieberman, Harris, Hoffman, & Griffith, 1957), identification probability can reflect the proportion of times the listener heard the stimulus as a given speech sound, not the degree to which the stimulus represented that speech sound. Accordingly, Miller's finding that the identification of a monitored sound was influenced by the VOT of the stop presented to the other ear might simply reflect the likelihood of the VOT detector firing in an all-or-none manner. Increasing the VOT of the non-monitored ear would change the probability of firing. Similarly, the effectiveness of an adapting stimulus as a function of its VOT value can simply reflect the probability that the stop is heard as completely voiced or completely voiceless on each successive presentation in the adapting series.

Asking listeners for continuous rather than discrete judgments provides a more direct method of assessing the nature of the output of feature detectors. If a stop consonant is consistently rated as being .6 voiced, we would have evidence for the availability of information about the degree to which a feature is represented in a sound. Since an average rating of .6 may also result from averaging across discrete judgments, it is necessary to evaluate the distribution of rating responses in

differentiating between the possibilities of binary and continuous featural information. We have used rating responses to evaluate the nature of the featural information available to listeners in the perception of stops differing in VOT and F2-F3 formant transitions and vowels differing in their steady-state formant values (Massaro & Cohen, in preparation).

In these rating studies, subjects first listened to a stop consonant continuum changing in place (ba to da) or voicing (ba to pa) or a vowel continuum going from *i* to *ɪ*. The listeners were instructed about the nature of the continuum and were asked to rate the sounds according to where they fell on the continuum. The rating response was made by setting a pointer along a 5.5 cm scale. The ends of the scale were labeled with the two alternatives and subjects were told to place the pointer to the location that they think the sound belongs. The sounds were presented in random order and each subject rated each of the seven sounds on a given continuum 160 times. Figure 2 shows that the rating responses were a systematic function of the stimulus values for each of the three continua. The *i-ɪ* continuum was discriminated slightly better than the two stop continua as indexed by the larger range in ratings and the smaller standard deviations. However, we are confident that the vowel ratings could be made identical to the stops by either decreasing the stimulus range for the vowels or increasing the stimulus range for the stop consonants.

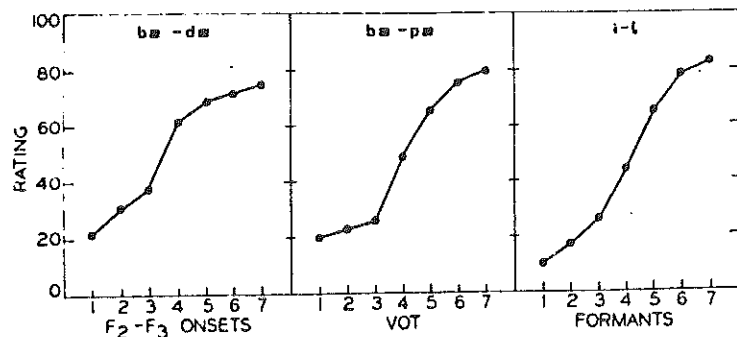


FIGURE 2 Rating responses for three speech continua.

The rating judgments demonstrate that listeners can transmit continuous information about acoustic properties of speech sounds. In the model developed by Oden and Massaro (1978), each acoustic feature is represented as a fuzzy predicate which describes the degree of presence or absence of the feature in the speech sound. The fuzzy predicate specifies the degree to which it is true that the sound has the acoustic feature. Truth values are expressed as continuous values between zero and one. For the dimension of VOT,

$$P(\text{long VOT}(S_{ij})) = .60, \quad (1)$$

would represent the statement that it is .60 true that the sound S_{ij} has a long VOT. According to this idea, each unique value of VOT could take on a unique truth value. The rating judgments provide a relatively direct index of these truth values.

Once the idea of continuous features is accepted, the integration of information across two or more features becomes an important issue. In traditional all-or-none classificatory schemes, logical conjunction was sufficient to combine values across dimensions. Pluses and minuses in the feature matrix were sufficient to classify the speech sound. Many possible classificatory schemes can be developed to represent

the integration of continuous featural information. Tests of models of featural integration require multifactor experiments rather than the more traditional single factor experiment. In a multifactor experiment, two or more acoustic dimensions are independently varied so that all combinations of the values of one dimension are paired with all combinations of the values of another property. The factorial design is optimal because it optimizes the number of data points relative to the number of parameters needed to test the various models of classification. Each level of each dimension requires a free parameter since the psychophysical function between the acoustic dimension and the truth value representing the feature is not usually known. However, Oden and Massaro (1978) and Derr and Massaro (submitted) have had some success in using an ogival relationship between the acoustic dimension and the truth values.

The ogival relationship predicts a reasonable relationship between changes in the acoustic dimension and changes in truth value. At extreme stimulus values, changes in the stimulus should have little consequence for the truth values. At intermediate stimulus values, however, small changes in the stimulus should produce relatively large changes in the truth values. A quantification of this relationship can be expressed as

$$t(x_i) = \frac{x_i^c}{x_i^c + (1-x_i)^c} \quad \text{where } y \geq 1 \quad (2)$$

where the x_i values are a simple linear function of the integer value i of the stimulus factor

$$x_i = a_i + b \quad (3)$$

except that values less than zero are set to zero and values greater than one are set to one. The form of the ogival function is specified by the three parameters a , b , and c . The values a and b allow the ogival function to shift along the stimulus dimension to position it at the optimal place. The exponent c determines the steepness of the ogival curve. Given that the ogival relationship requires 3 free parameters, a stimulus dimension of seven levels can be specified by just three rather than seven free parameters.

In order to illustrate how the present model is applied and tested, consider an experiment carried out by Massaro and Oden (submitted). Seven levels of voice onset time (VOT) were crossed with seven levels of the onsets of the F2-F3 transitions in stop consonant-vowel syllables. The VOT ranged from a completely voiced to a completely voiceless sound. The values were 10, 15, 20, 30, 25, and 40 msec. The seven levels of the F2-F3 onset frequencies ranged from 1345 to 1796 Hz for F2 and 2397 to 3200 Hz for F3 to give a continuum of sounds going from a labial to an alveolar place of articulation. Subjects made repeated identifications of the 49 unique syllables from the alternatives $/b\alpha l/$, $/d\alpha l/$, $/p\alpha l/$, and $/t\alpha l/$.

The four panels of Figure 3 present the percentage of $/b\alpha l/$, $/p\alpha l/$, $/d\alpha l/$, and $/t\alpha l/$ identifications, respectively, as a function of the two independent variables. The levels along the abscissa are not equally spaced but rather have been adjusted to be proportional to the differences between the respective marginal means across the levels of the F2-F3 transitions. The differences were computed separately for each of the four response alternatives and then averaged over response types so that all four of the panels have the same spacing along the abscissa.

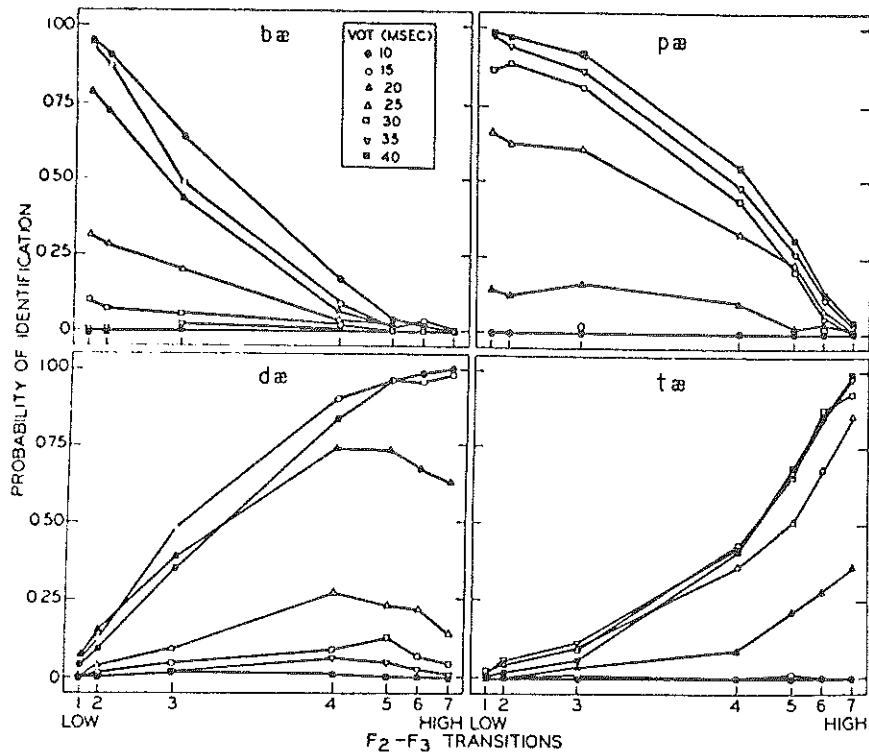


FIGURE 3 Percentage of /bæ/, /pæ/, /dæ/, and /tæ/ identifications as a function of VOT and F2-F3 transitions.

Evaluation of the results is facilitated by developing the basic model of featural integration given by Oden and Massaro (1978). Each of the four response alternatives is specified as a prototype corresponding to a proposition

- /bæ/: (short VOT) and (low F2-F3 onsets) (4)
- /pæ/: (long VOT) and (low F2-F3 onsets) (5)
- /dæ/: (short VOT) and (high F2-F3 onsets) (6)
- /tæ/: (long VOT) and (high F2-F3 onsets) (7)

The properties for the prototypes would also include other acoustic features characterizing stop consonants and the vowel /æ/. These are not included in the propositions since they are present in all of the four alternatives. These propositions specify the ideal values of each of the acoustic features for the particular speech sound.

Upon presentation of a speech sound, the feature detection process applies a truth value for each acoustic feature. The truth value specifies the degree to which it is true that the sound has the relevant acoustic feature. For example,

$$t[\text{short VOT}(S_{ij})] = .60 \quad (8)$$

represents that it is .60 true that the speech sound S_{ij} , from the i th row and j th column of the factorial stimulus design, has a short VOT. To simplify the notation,

let SV_j and LV_j correspond to short and long VOTs, respectively. The subscript j signifies that the values change only with changes in the row variable VOT. Similarly, LO_j and HO_j correspond to low and high F2-F3 onsets, respectively. The values change only with changes in the column j variable of the F2-F3 onset frequencies.

It is possible to simplify the descriptions of the prototypes by allowing each feature to be symmetrical along its acoustic dimension. In this case, a long VOT could be specified as the negation of the short VOT

$$\text{long VOT} = \text{NOT}(\text{short VOT}) \quad (9)$$

It is also reasonable to define the truth of the negation of a feature as one minus the truth value of the feature. In this case, if .6 specifies the truth value of a short VOT, then $1 - .6 = .4$ would specify the truth value of a long VOT. In general, the value $LV_j = 1 - SV_j$. A similar symmetry could be assumed between high and low F2-F3 onsets

$$\text{low F2-F3 onsets} = \text{NOT}(\text{high F2-F3 onsets}) \quad (10)$$

Following this logic, the value LO_j would be equal to $1 - HO_j$. Within limits, the assumption of symmetrical features in the prototypes leads to equivalent predictions as the more complex model given by Equations 4-7. Incorporating the assumption of symmetrical features in the prototype definitions gives Equations 11-14 in place of Equations 4-7

- /bæ/: (short VOT) and NOT (high F2-F3 onsets) (11)
- /pæ/: NOT (short VOT) and NOT (high F2-F3 onsets) (12)
- /dæ/: (short VOT) and (high F2-F3 onsets) (13)
- /tæ/: NOT (short VOT) and (high F2-F3 onsets) (14)

At the prototype match stage, the truth values derived from feature detection are inserted in the prototypes. High truth values in the prototypes would represent a good match of the speech sound with the prototype alternative whereas low truth values would represent a poor match. However, a precise index of the degree to which the speech sound matches each prototype requires the conjunction of truth values across the acoustic features. Two possible conjunction rules are addition and multiplication. Given an additive conjunction rule, the degree to which the sound S_{ij} matches the prototype /b/ is given by the matching function

$$bæ(S_{ij}) = SV_j + LO_j \quad (15)$$

The multiplicative rule gives the matching function

$$bæ(S_{ij}) = SV_j \times LO_j \quad (16)$$

Given the matching functions for each of the alternative prototypes, the speech sound is identified on the basis of the relative degree of match. Following the rationale of Luce's (1959) choice model it is assumed that the probability of identifying a stimulus to be a particular syllable is equal to the relative degree to which that syllable matches the stimulus compared to the degree of match of the other syllables under consideration. In our example, the person must identify the speech sound as either /bæ/, /pæ/, /dæ/, or /tæ/.

The probability of a /bæ/ identification will, therefore, be given by

$$P(bæ | S_{ij}) = \frac{bæ(S_{ij})}{bæ(S_{ij}) + pæ(S_{ij}) + dæ(S_{ij}) + tæ(S_{ij})} \quad (17)$$

where the variables in the ratio represent the matching functions for the four alternative speech sounds.

Returning to the results of the Massaro and Oden study shown in Figure 3, we can evaluate them in terms of the simple fuzzy logical model. The results provide a qualitative test between additive and multiplicative conjunction rules. An additive conjunction rule in the present model predicts a series of parallel lines within each of the panels. A multiplicative combination predicts a fan of diverging lines. For each response alternative, the pattern of results is that of a gradually diverging fan of curved lines. The fact that the lines come together at a point at zero percent and that the bottom curves are relatively flat near zero percent indicate that the information about the features associated with the two independent variables are combined in a multiplicative rather than additive or other fashion (see Massaro & Cohen, 1976; Oden, 1977) for each of the candidate phonemes.

The multiplicative combination rule has an appealing property not given by the additive rule. In the additive rule, a feature carries the same weight regardless of its value or the value of other features. The multiplicative rule, on the other hand, allows the least ambiguous feature to carry more weight with respect to the final truth value (see Massaro & Cohen, 1976).

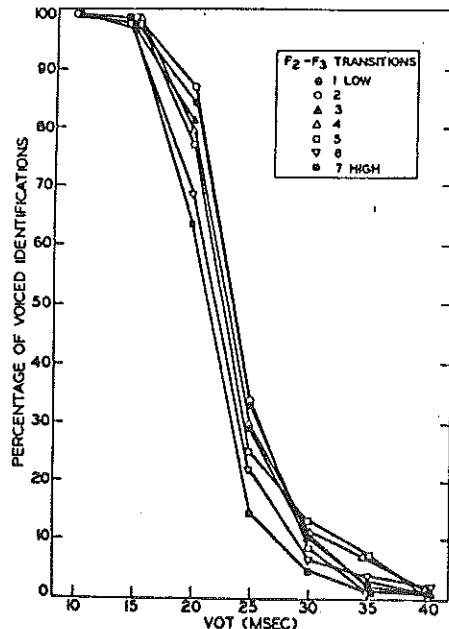


FIGURE 4 Percentage of voiced identifications (*/bæ/* or */dæ/*) as a function of VOT; the level of F2-F3 transitions is the curve parameter.

Although the multiplicative combination rule does a better job than the additive rule, something in addition to a simple multiplication of the acoustic features of VOT and F2-F3 onsets must be involved. The simple model given by Equations 16 and 17 predicts straight lines when the abscissa is spaced in the manner of Figure

3. To better evaluate how the results deviate from the predictions of the simple fuzzy logical model, the responses can be replotted in terms of the percentage of voiced identifications on the one hand, and the percentage of labial identifications on the other.

Figure 4 presents the percentage of times the stimuli were identified to be voiced phonemes (*/bæ/* or */dæ/*) plotted as a function of VOT; F2-F3 transitions is the curve parameter. The results show that VOT is a sufficient cue for voicing; the data go from completely voiced at a 10 msec VOT to completely voiceless at a 40 msec VOT. Phoneme identifications were most ambiguous with respect to voicing at the 20 and 25 msec VOTs. Although VOT was a sufficient cue to voicing, Figure 4 shows that F2-F3 transitions also had a consistent influence on voicing: the percentage of voiced phoneme identifications increased as the onset frequencies of the F2-F3 transitions decreased. This result contrasts with our previous finding of no voicing boundary shift (Oden & Massaro, 1978) or the findings that the voicing boundary is at shorter values of VOT for labial stops than it is for alveolar stops (Lisker & Abramson, 1970; Miller, 1977)

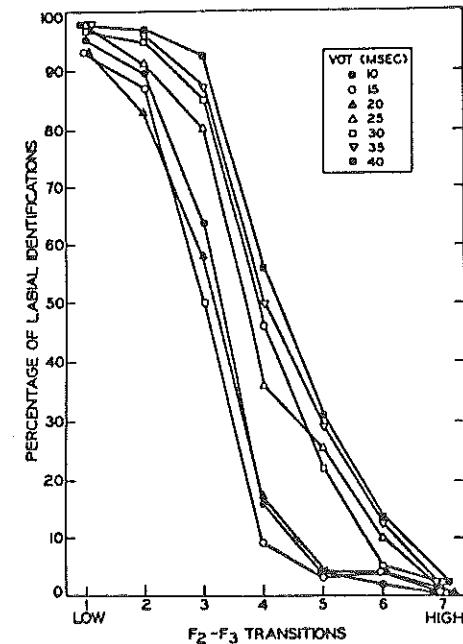


FIGURE 5 Percentage of labial identifications (*/pæ/* or */bæ/*) as a function of F2-F3 transitions; VOT is the curve parameter.

Figure 5 presents the percentage of times the stimuli were identified to be labial phonemes (*/bæ/* or */pæ/*) as a function of the F2-F3 transitions; VOT is the curve parameter. Replicating previous findings, phoneme identification with respect to place changed from labial to alveolar with increases in the onset frequencies of the F2-F3 transitions. Sounds with low F2-F3 onset frequencies were consistently heard as labial phonemes whereas those with high frequencies were heard as

alveolar phonemes. Intermediate frequencies gave more ambiguous identifications. Although the F2-F3 transitions were sufficient cues to place, VOT also had a relatively large influence on place, especially at the intermediate levels of the F2-F3 transitions. Decreasing the VOT decreased the likelihood of a labial identification. The place boundary was at level 3 of the F2-F3 transitions for a VOT of 15 msec whereas it was at level 4 for a VOT of 35 msec. The decrease in the percentage of labial identifications with decreases in VOT was highly consistent except for an inversion with VOT of 10 msec. This general result agrees with those of Miller (1977b) and Repp (1977). Any possible explanations for the reversal at a VOT of 10 msec is secondary to the more important issue of the independence of labial identifications on VOT.

When the identification of a speech dimension such as voicing is also dependent on an acoustic feature for another dimension such as place, we called the outcome a boundary shift. That is to say, the transition from voiced to voiceless identifications as a function of VOT is shifted by the values of the F2-F3 onsets. We have distinguished three alternative explanations for the boundary shifts that have been observed. Although we are focusing on a specific result, the experiments have the promise of illuminating general processes in featural evaluation and integration in pattern recognition. The three explanations are (1) non-independence of feature evaluation, (2) multiple features, and (3) modifiers in long-term memory prototypes.

Feature Nonindependence One possible type of feature nonindependence is that there are complex low-level auditory interactions so that, for example, the perceptual realization of VOT is modified by the F2-F3 transition frequencies. Analogously, the perceptual realization of the F2-F3 transitions could be modified by VOT. According to this view, VOT and F2-F3 transitions maintain their role as primary cues for voicing and place respectively, but the value of each acoustic cue is dependent on the stimulus value of the other. As an example of such an interaction in nonspeech stimuli, changing the perceived hue of a color from green to blue by changing the wavelength also changes the perceived brightness since we are less sensitive to wave-lengths in the blue than in the green part of the visible spectrum. An experiment in color perception which would be analogous to the present studies would independently vary the wavelength and intensity of the color and the results would presumably show that wavelength not only influences the perception of hue but also the perception of brightness.

Multiple Features The second possible explanation for the boundary shifts is that changes in a single stimulus dimension may change more than one acoustic feature. Since any manipulation of an arbitrarily defined stimulus dimension will also produce changes along other dimensions, it may be that some of the co-varying changes are other acoustic features. When a given acoustic characteristic functions primarily as an acoustic feature for one distinction but also cues a second distinction; it is called a multiple feature. For example, increasing VOT not only increases the time between burst onset and vocal-chord vibration onset but also increases the total amount of aspiration noise. The quality of the aspiration may function as an independent acoustic feature for place of articulation. It is possible that the aspiration during the VOT period of most synthetic speech without bursts is more representative of the burst and aspiration of a voiceless labial (/p/) than a voiceless alveolar (/t/). The spectrograms in Figures 6 and 7 show that the aspiration during the VOT of our synthetic speech were more representative of the burst and aspiration of a natural /p/ than of a natural /t/. If this were the case, longer VOTs would have produced more aspiration which would have produced a more labial sound.

In terms of the formal model of Oden and Massaro (1978), both the F2-F3 transitions and the degree of aspiration during the VOT period contribute to

perception of the place distinction. In this case, the prototype for /t/ would be defined as:

T:[NOT(aspiration)] and [high F2-F3 onsets] and [NOT(short VOT)] (18)

where /t/ is now characterized by not having low-frequency aspiration.

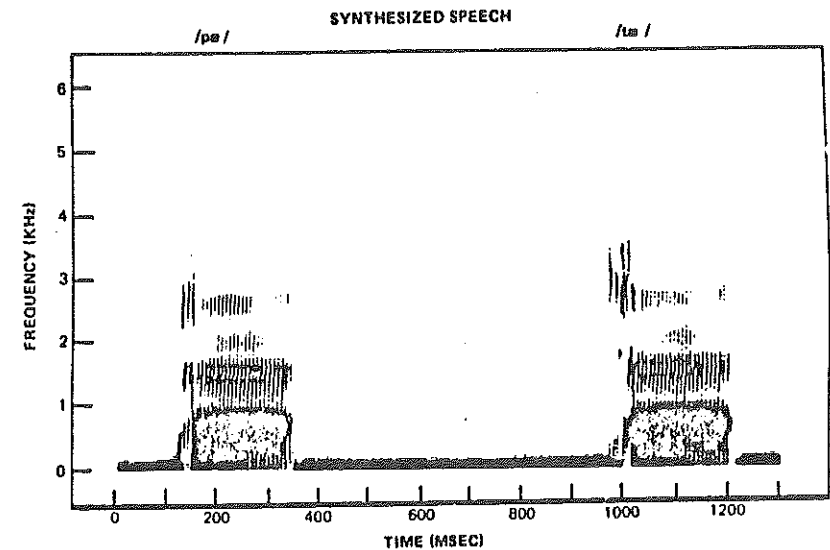


FIGURE 6 Spectrograms of two synthesized speech sounds from the present Experiment 1.

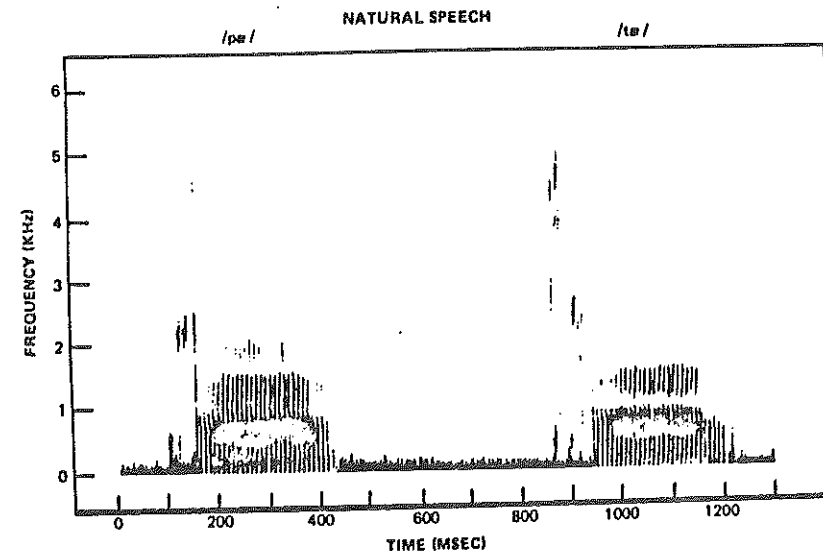


FIGURE 7 Spectrograms of two naturally occurring speech sounds.

Prototype Modifiers The third explanation of the boundary shifts is that the prototypes of the speech sounds in long-term memory include modifiers of the featural information. Rather than defining /t æ / as

/t æ /: (long VOT) and (high F2-F3 onsets) (19)

it could be assumed that /tae/ requires relatively extreme values of F2-F3 onsets

/t æ /: (long VOT) and quite (high F2-F3 onsets) (20)

where the modifier "quite" expresses the extremity of this feature in the prototype.

The central assumption of this explanation is that more extreme values of features are required for some phonemes. One way for this to come about could be through listener's experience with natural speech. For example, Fant (1973, Chapter 11) points out that, with most vowels, the locus of F2 at the instant of release is higher for /t/ than it is for /d/. Thus, since a high F2-locus is a cue to alveolarity, it would be reasonable for listeners to expect /t/ to be more strongly alveolar relative to /d/. If listeners use this information, then for a given level of F2-F3 transitions, they should make fewer /t/ identifications than they would if they did not have this expectation of higher F2 values for /t/. Producing fewer /t/ responses would mean that the place boundary would be shifted toward the alveolar end when the speech sounds are voiceless. This is exactly the result that was obtained in Figure 5.

The model has been developed and utilized to provide a framework for research in the identification of speech sounds varying on two or more dimensions. The model can also be extended to include other theoretical and empirical issues in speech perception research. In addition to a treatment of other multifactor experiments in speech, extensions of the model allow an account of auditory context effects, higher-order context effects, and normalization in speech perception. Auditory context sometimes produces contrast as in selective adaptation and anchor effects. Higher-order context is effective when information at a higher (more abstract) level influences lower-level decisions. Finally, normalization refers to the relative nature of featural information.

Summary In summary, we believe that the present framework offers a productive approach to the study of acoustic features in speech perception. Factorial designs have shown that a simple identification task of sounds differing on just a few dimensions produces relatively complex results. The results reject the idea of all-or-none binary features and, in addition, show that a single acoustic feature for a given articulatory distinction is probably not adequate. Furthermore, it seems likely that prototypes are defined in a more complex manner than might be expected from the binary assumption of linguistic features. We have not yet determined whether the descriptions of prototype speech sounds are relatively flexible and modifiable. Although other important issues remain unresolved, the present framework appears to offer a formal theoretical and empirical approach to the study of these issues.