Journal of Phonetics (1974) 2, 181–194

On the identification of place and voicing features in synthetic stop consonants

James R. Sawusch and David B. Pisoni

Department of Psychology, Indiana University, Bloomington, Indiana 47401, *U.S.A.*

Received 30*th January* 1974

Abstract: Two models of the interaction of phonetic features in speech perception were used to predict Ss' identification functions for a bi-dimensional series of synthetic CV syllables. The stimuli varied systematically in terms of the acoustic cues underlying the features of place of articulation and voicing. Model I assumed the additivity of phonetic features and their independent processing in perception. Model II assumed that the phonetic features interact and are not processed independently. The fit of Model II to the bidimensional series data was better than the fit of Model I suggesting that the phonetic features of place and voicing in stop consonants are not processed independently but rather show a mutual dependency on each other.

Theoretical accounts of speech sound perception have frequently proposed some type of articulatory-motor involvement during perceptual processing (Liberman, 1957; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Stevens, 1960; Stevens & Halle, 1967). One reason for this may be that research on speech sound perception has drawn its descriptive categories from the account of speech production offered by phoneticians. Thus, the articulatory dimensions that distinguished different classes of speech sounds in production served as the basis for uncovering the acoustic cues that distinguish different speech sounds in perception. Spectrographic analysis and perceptual experiments revealed that the sounds of speech were not arrayed along a single complex dimension but could be specified in terms of a few simple and independent dimensions (Gerstman, 1957; Liberman, 1957). Acoustic dimensions were found in early experiments with synthetic speech to provide distinctions in perception corresponding to the articulatory dimensions of speech production, suggesting that perceptual and articulatory dimensions of speech may be intimately linked (Delattre, 1951; Liberman, 1957).

Two articulatory features to receive a great deal of attention in the description of stop consonant production are place of articulation and voicing. Both these features have fairly well defined acoustic properties which presumably mirror the differences in production (Delattre, 1951). For example, the feature of place of production refers to the point of constriction in the vocal tract where closure occurs. The acoustic cues that underlie the place feature in CV syllables are reflected in the formant transitions into the following vowel, particularly the direction and extent of the second and third formant transitions (Liberman *et al.*, 1967). In contrast, the voicing feature is related to the presence or absence of periodic vibration of the vocal chords. The acoustic cues that underlie the voicing feature in stop consonants in initial position are reflected in terms of the relative onset of the first formant transition (i.e. F1 "cutback") and the presence of aspiration in the higher formants

(Liberman, Delattre & Cooper, 1958). This compound acoustic cue has been called "voice-onset time" by Lisker & Abramson (1964) and corresponds to the time interval between the release from stop closure and the onset of laryngeal pulsing.

Figure 1 presents schematized spectrographic patterns which show the acoustic cues for place and voicing features for the CV syllables /ba/, /da/, /pa/ and /ta/. There is a relatively simple relation between articulatory features of place and voicing and their respective acoustic cues when the vowel is held constant (see also Liberman, 1970). Consonants within a particular row share voicing; /ba/ and /da/ are voiced, /pa/ and /ta/ are voiceless. The major acoustic cue for voicing in these syllables is the cutback or elimination of the initial portion of the first formant. Consonants within a particular column share place of production; /ba/ and /pa/ are bilabial stops, /da/ and /ta/ are alveolar stops. The primary acoustic cue for place is the direction and extent of the second and third formant transitions.

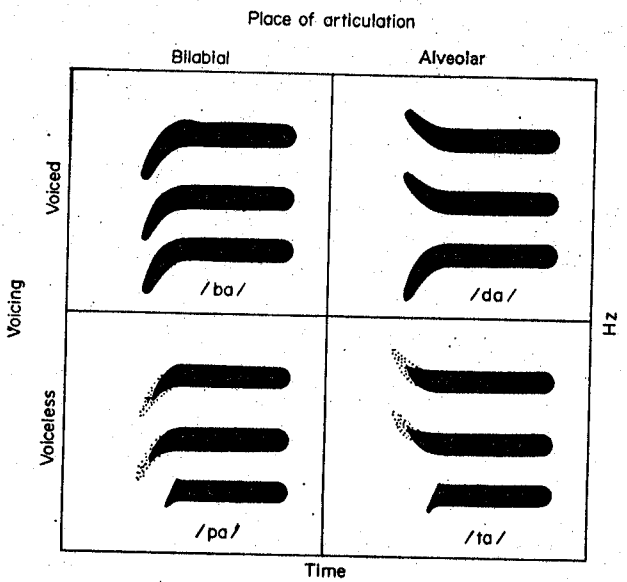

Figure 1 A schematized sound spectrogram of the syllables /ba/, /da/, /pa/ and /ta/ as used in the present experiment.

Several perceptual experiments employing stop consonant-vowel syllables have concluded that the features of place and voicing are processed independently of each other. For example, Miller & Nicely (1955) analyzed the perceptual confusions among 16 consonant-vowel syllables presented to listeners under various signal-to-noise ratios and filtering conditions. They computed the sum of the information transmitted by their several features separately and in combination. Since the two values were approximately equal, they concluded that the features used in their analysis were mutually independent. Among these features were place and voicing. As a part of a larger investigation of dichotic listening, Studdert-Kennedy & Shankweiler (1970) reached the same conclusion by a similar analysis of place and voicing confusions among stop consonants.

These studies imply that features are extracted separately during early perceptual processing and are later recombined in response. Figure 2 represents a simplified block diagram of this process. The output of the auditory analysis is a set of acoustic cues $\{c_i\}$. These cues are combined and from them a set of phonetic features $\{f_j\}$ is recognized. Finally, the

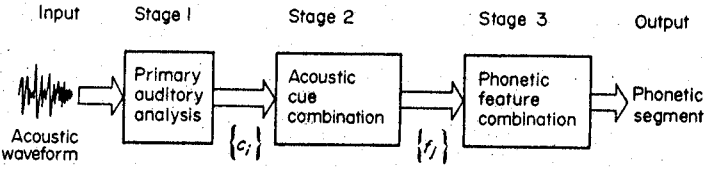

Figure 2 A block diagram of stages of perceptual analysis for phonetic segments. The input is the acoustic waveform and the output is a sequence of phonetic segments.

phonetic features are combined to yield the perception of the phonetic segment. Together, Stages 2 and 3 form what Studdert-Kennedy (1974) has described as the "phonetic" stage of processing. We assume that phonetic features are recognized or identified in short-term memory (STM) when the auditory patterns derived from the acoustic cues have made contact with some representation generated from synthesis rules residing in long-term memory. We assume that abstract phonetic features have an articulatory rather than acoustic reality in STM although we will not try to justify this assumption at the present time.

To say that phonetic features are independent implies independence of processing at all three stages of Fig. 2. That is, the acoustic cues are extracted from the acoustic waveform separately and independently of each other in Stage 1. Then the phonetic features are extracted separately and independently from the acoustic cues in Stage 2. Finally, the phonetic features are combined separately and independently of each other in Stage 3, resulting in a particular phonetic segment.

The independence of phonetic feature processing in Stage 3 may be described quantitatively by a simple linear or additive model. The phonetic features of place (f_1) and voicing (f_2) that are output from Stage 2 are weighted separately and then added together in Stage 3. Equation 1 expresses this concept algebraically:

$$X = a_1 f_1 + a_2 f_2. \tag{1}$$

Here, f_1 is the proportion of the place feature of stimulus X output from Stage 2 and a_1 is its associated weight. Similarly, f_2 is the proportion of the voicing feature of X output from Stage 2 and a_2 is its associated weight. Since these two features are sufficient to distinguish among the four stops b, d, p, and t, we will ignore other phonetic features as being redundant and nondistinctive.

However, evidence for non-independence of phonetic features, in particular the features of place and voicing, has also been presented by several investigators. This non-independence could come at any of the three levels mentioned. For example, Haggard (1970) put the dependence relationship of the features of place and voicing in the second stage, where phonetic features are extracted. In the model of Haggard (1970), the listener's decision on the voicing feature is partly determined by his prior decision on the place feature.

Lisker & Abramson (1964, 1967) reached the corresponding conclusion upon examination of their production data for stop consonants in initial position. The voicing feature as reflected in VOT depends on the feature of place of production; the VOT lag at the boundary between voiced and voiceless stops increases as place of production moves further back in the vocal tract (i.e. from /ba/ to /da/ to /ga/). Given the anatomical and physiological constraints on speech production, this position is *a priori* more plausible.

This particular concept of non-independence, where the feature of voicing partly depends on the feature of place, may also be expressed algebraically. We will again assume

independence of processing in Stages 1 and 2. The non-independence in Stage 3 may then be expressed as in equation 2:

$$X = a_1 f_1 + a_2 f_2 - b(1 - f_1) f_2. \tag{2}$$

Here, a_1, f_1, a_2, and f_2 are the same as in equation 1. However, the constant b represents the weight given to the interaction term of place and voicing $[(1 - f_1)f_2]$. This is not a general dependence term but rather represents a specific dependency of the feature of voicing on the feature of place. It has the effect of shifting the phonetic boundary of X to longer voicing values as place goes from bilabial to alveolar.

The purpose of the present study was to reexamine the identification of place and voicing features in stop consonants and to determine by means of a new experimental paradigm whether these two phonetic features (i.e. place and voicing) are combined additively or non-additively in Stage 3 as shown in Fig. 2. Stimuli for the experiment were three sets of synthetic speech sounds that varied systematically in the acoustic cues underlying the two phonetic features of place and voicing. One series of stimuli varied in the acoustic cues that underlie the phonetic features of place, while holding the voicing feature constant (/ba/ to /da/ with VOT at 0 ms). A second series varied the acoustic cues underlying the phonetic feature of voicing, while holding the place feature constant (/ba/ to /pa/ with F2 and F3 always rising). The final series varied the acoustic cues underlying both phonetic features simultaneously (/ba/ to /ta/). These three sets of speech sounds were presented separately to listeners for identification into the categories /ba/ – /da/, /ba/ – /pa/, and /ba/ – /ta/, respectively. By using synthetically produced stimuli, the correlation between place of production and voicing that Lisker & Abramson (1964, 1967) had found in natural speech could be controlled experimentally.

Our principal aim was to determine whether the probabilities of identification along the bidimensional continuum (/ba/ – /ta/) could be predicted from some combination of the probabilities along the separate unidimensional series. We consider below two possible models of how these separate features might be combined in the bidimensional case. Both Model I and Model II are concerned with the manner in which phonetic features are combined in phonetic perception. All processing up to Stage 3 of Fig. 2 is assumed to be independent according to the definition of independence for these stages given previously. We also assume that processing in Stages 1 and 2 takes place in parallel and is automatic in the sense that Ss do not have control over these stages of perceptual processing. (See also Shiffrin & Geisler, 1973 and Shiffrin, Pisoni & Castenada-Mendez, 1974.)

Model I: Linear Combination of Phonetic Features

Hereafter, if S identified a stimulus as /ba/ it will be denoted B and likewise, /da/ as D, /pa/ as P, and /ta/ as T. In the /ba/ to /da/ series only the acoustic cues underlying the phonetic feature of place of articulation were varied. Since processing in Stage 2 is assumed to be independent (i.e. separate for different phonetic features), the only variation in the output of Stage 2 on the /ba/ to /da/ (place) series should be in feature f_1, the phonetic feature of place of articulation. Accordingly, since the only variation in the input to Stage 3 is in f_1, the output of Stage 3 (a phonetic segment) is assumed to vary directly with the input (f_1) and thus accurately reflect f_1. However, due to noise in the acoustic waveform and the first two stages of processing, the outputs of Stage 2 are assumed to be probabilistic in nature. Thus, Ss' judgments of the stimuli from the /ba/ to /da/ series (the probability of responding D to a stimulus $Pr[D]$) may then be construed as accurately reflecting the input (f_1) to Stage 3. Similarly, $Pr[P]$ from the /ba/ to /pa/ (voicing) series may be construed as

accurately reflecting the input of the voicing feature (f_2) to Stage 3. Now, we can represent f_1 and f_2 from equations 1 and 2 as follows:

$$f_1 = Pr[D] \text{ on the /ba/ – /da/ series (PLACE)} \tag{3}$$

$$f_2 = Pr[P] \text{ on the /ba/ – /pa/ series (VOICING).} \tag{4}$$

Substituting equations 3 and 4 into equation 1 we obtain equation 5:

$$Pr[T] = a_1 Pr[D] + a_2 Pr[P]. \tag{5}$$

$Pr[T]$ in equation 5 represents the probability of a T response on the bidimensional /ba/ to /ta/ series.

One additional assumption will be made. This is shown in equation 6:

$$a_2 = 1 - a_1 \quad \text{where} \quad 0 \leqslant a_1 \leqslant 1. \tag{6}$$

This constraint is placed on a_1 and a_2 so that $Pr[T]$ will equal one when both $Pr[D]$ and $Pr[P]$ are equal to one. Since only one parameter is being used, we delete the subscript from parameter a.

If we now combine equations 5 and 6 and delete the subscript on parameter a we obtain equation 7:

$$Pr[T] = a\,Pr[D] + (1 - a)Pr[P]. \tag{7}$$

Equation 7 represents Model I. This model assumes independence of the features of voicing and place. If we estimate parameter a from the data by method of least squares, then Model I can be used to predict the bidimensional /ba/ to /ta/ identification function based on the unidimensional /ba/ to /da/ and /ba/ to /pa/ data.

Model II: Non-linear Combination of Phonetic Features

A development similar to that given for Model I may be applied to equation 2. If we combine equations 2, 3, 4 and 6, we obtain equation 8:

$$Pr[T] = a' Pr[D] + (1 - a')Pr[P] - b(1 - Pr[D])Pr[P]. \tag{8}$$

Here, a' is used to distinguish this parameter from parameter a of Model I. A major disadvantage of equation 8 is that it requires two different parameters, a' and b to be estimated from the data.

Equation 8 assumes that Ss employ information about both the phonetic features of place and voicing to make their decision on the /ba/ to /ta/ series. However, either of these features alone may be sufficient for an S to distinguish between /ba/ and /ta/ when only two response categories are permitted. For example, an S could identify /ba/ and /ta/ on voicing alone (i.e. if voiced respond /ba/, if voiceless respond /ta/) or place alone (i.e. if bilabial respond /ba/ if alveolar respond /ta/). Since these stimuli differ with regard to both voicing and place, Ss may use only one of these features in their decision. However, it is also possible that a particular decision on one feature necessarily entails a particular decision on the other. This is quite likely considering the constraints on production. In production, a shift in place of articulation entails a shift in VOT, but not vice versa. However, in perception, a shift in VOT may serve as a cue to place of articulation; a shift in place may also serve as a cue to voicing.

Previous investigators have found that decisions based on the voicing feature are more consistent and in some sense, easier than decisions based on other features, including place (Miller & Nicely, 1955; Studdert-Kennedy & Shankweiler, 1970; Shepard, 1972). One

reason for this finding may be the multiplicity of cues to the voicing feature (Liberman, Delattre & Cooper, 1958; Lisker & Abramson, 1964; Summerfield & Haggard, 1972), as compared with the relatively restricted number of cues to the place feature. If Ss were to use only one feature, it seems likely that they would use the feature of voicing for the /ba/ to /ta/ series. We can operationalize this assumption by setting a' from equation 8 to zero. This means that $1 - a'$ will be one and that the probability of a /ta/ response ($Pr[T]$) will be the result of the amount of the voicing feature present minus the interaction of place and voicing. This is summarized in equation 9, which represents Model II:

$$Pr[T] = Pr[P] - b(1 - Pr[D])Pr[P]. \tag{9}$$

Model II can be used to predict the bidimensional /ba/ to /ta/ series from the /ba/ to /da/ and /ba/ to /pa/ data by estimating the parameter b with the method of least squares. By setting parameter a' to zero Model II assumes that Ss categorize the stimulus as either /ba/ or /ta/ on the basis of the voicing feature alone. Thus, parameter b may be used as an estimate of how much an S's decision on the voicing feature depends upon the place information in the stimulus.

Method

Subjects

Ss were 24 students in introductory psychology participating as a part of the course requirement. All Ss were native American speakers of English, right-handed and reported no history of a speech or hearing disorder.

Stimuli

The three synthetic speech syllable series were /ba/ to /da/, /ba/ to /pa/, and /ba/ to /ta/. Each series contained 11 stimuli. The /ba/ to /da/ series varied in the initial frequencies of the second and third formant transitions. The second formant varied from an initial value of 1859 Hz (/ba/) to an initial value of 3530 Hz (/da/) in ten equal steps. The /ba/ to /pa/ series varied in VOT from 0 ms VOT (/ba/) to a +50 ms VOT (/pa/) in 5 ms steps. Aspiration replaced the harmonics in the second and third formant transitions for the duration of the F1 cutback. The /ba/ to /ta/ series combined the two component changes in a one-to-one fashion, resulting in the third 11 stimuli sequence. All stimuli were of 300 ms duration with a 50 ms transitional period followed by a 250 ms steady state vowel (/a/). The three series of synthetic stimuli were prepared on the speech synthesizer at Haskins Laboratories and recorded on magnetic tape.

Procedure

The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the headphones equivalent to 80 dB SPL re 0·0002 dynes/cm^2 for the steady state calibration vowel /a/.

On any one tape Ss heard 10 presentations of each of the 11 stimuli in random order with 4 s between stimuli. Ss were run in four different groups, six Ss in each group. Each group heard each tape three times resulting in 30 judgments of each stimulus for each S. In addition, the /ba/ to /ta/ tape was presented three more times with instructions to categorize the stimuli as any one of the four syllables ba, pa, da, or ta. The order of tape presentations was

counterbalanced; two groups received the /ba/ to /da/ tape first and two received the /ba/ to /pa/ tape first.

For each of the tapes Ss were told that they would hear synthetic speech syllables and they were to identify them as /ba/ or /da/, /ba/ or /pa/, /ba/ or /ta/, or /ba/, /pa/, /da/, or /ta/. Ss were told to record their identification judgment of each stimulus by writing down the initial stop consonant in prepared response booklets.

Results and Discussion

The identification probabilities for the /ba/ to /da/ (place) and /ba/ to /pa/ (voicing) series were in accord with previous experiments. All the stimuli at one end of the series were consistently categorized one way and all the stimuli at the other end were consistently categorized the other way. There were a few transition stimuli (generally one or two in the middle of the series) which were categorized both ways at a near chance (0·5) level. Identification functions from these two series for a typical S (S number 13) are shown in Fig. 3(a, b). The /ba/ to /ta/ identification function for the same S is shown in Fig. 3(c).

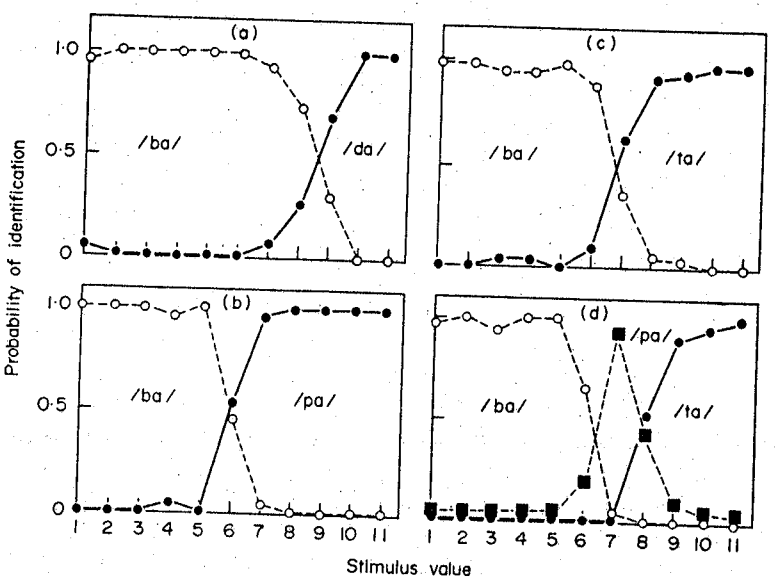

Figure 3 Identification functions for a representative S. Part (a) is for the /ba/ to /da/ series, part (b) the /ba/ to /pa/ series, part (c) the /ba/ to /ta/ series with two response alternatives, and part (d) is for the /ba/ to /ta/ series with four response alternatives.

The /ba/ to /da/ function [Fig. 3(a)] and /ba/ to /pa/ function [Fig. 3(b)] for each S were used to predict the /ba/ to /ta/ function [Fig. 3(c)] using both Models I and II. In order to estimate the weighting factor (a) from Model I for each S, a was allowed to vary from 0·0 to 1·0 in increments of 0·02. The squared error between the predicted and observed identification functions was then calculated for each value of a. The value which resulted in the minimum squared error for each S was chosen as the best estimate of a. These values of parameter a and their associated squared errors are shown in Table I. In 21 of 24 Ss the proportion of the variance accounted for by the predicted values exceeded 82%. The mean proportion of variance accounted for over all Ss by Model I was 91·9%.

Table I Variance accounted for by Model I

Subject	Parameters		Minimum squared error	Percent of variance accounted for
	a	1-*a*		
1	1·0	0·0	0·1729	92·4
2	0·52	0·48	0·1681	94·0
3	0·74	0·26	0·0093	99·6
4	0·56	0·44	0·0417	98·3
5	1·00	0·00	0·0724	97·5
6	1·00	0·00	0·1920	93·2
7	0·82	0·18	0·2599	90·5
8	0·60	0·40	0·0197	99·4
9	0·00	1·00	0·5145	82·6
10	0·86	0·14	0·0073	99·7
11	0·00	1·00	0·0635	98·1
12	1·00	0·00	0·4015	83·5
13	0·34	0·66	0·1035	95·5
14	0·08	0·92	0·0893	97·7
15	1·00	0·00	0·0720	97·6
16	1·00	0·00	0·0299	98·8
17	1·00	0·00	0·3421	85·9
18	0·00	1·00	0·1886	93·2
19	0·24	0·76	0·0961	95·9
20	0·16	0·84	0·0136	99·5
21	0·56	0·44	1·0517	73·3
22	1·00	0·00	0·7190	76·3
23	0·52	0·48	0·0233	99·3
24	1·00	0·00	1·1697	63·5

The data were analyzed a second time for Model II. Parameter a' was set to zero, in accord with the assumption that Ss would use only the voicing feature in making their judgment. Parameter b, the weight of the interaction term in Model II, was allowed to vary from 0·0 to 1·0 in increments of 0·02. The squared error between predicted and observed identification functions was also computed. The values for each S which resulted in minimum squared error are shown in Table II. The proportion of the variance accounted for by the predicted function was computed and is shown in Table II. The proportion of the variance accounted for by Model II is greater than or equal to that accounted for by Model I for every S except one. The overall mean proportion of variance accounted for was 93·1% in Model II.

Both Model I and Model II predict the identification probabilities along the bidimensional speech series reasonably well. However, predictions from Model II, the interaction model, fit the observed probabilities somewhat better than predictions from the additive model. There was an increase in the proportion of variance accounted for in 15 of the 24 Ss with Model II. The eight Ss for whom the proportion of variance accounted for remained the same from Model I to Model II showed the highest proportion of variance accounted for in Model I (over 95%).

We suggested earlier that identification of the bidimensional series /ba/ to /ta/ might be based on the use of only one feature—voicing, since Ss were constrained to only two response categories. Parameter a' in Model II was set at zero on the assumption that the

Table II Variance accounted for by Model II

Subject	Parameters			Minimum squared error	Percent of variance accounted for
	a'	1-a'	b		
1	0·00	1·00	0·66	0·2405	90·1
2	0·00	1·00	0·96	0·0887	96·5
3	0·00	1·00	0·64	0·0115	99·6
4	0·00	1·00	0·56	0·0370	98·5
5	0·00	1·00	1·00	0·0376	98·6
6	0·00	1·00	1·00	0·1451	94·7
7	0·00	1·00	0·82	0·2446	91·2
8	0·00	1·00	0·58	0·0182	99·4
9	0·00	1·00	1·00	0·4122	85·6
10	0·00	1·00	0·86	0·0073	99·7
11	0·00	1·00	1·00	0·0102	99·6
12	0·00	1·00	1·00	0·3801	84·1
13	0·00	1·00	0·34	0·1033	95·5
14	0·00	1·00	0·08	0·0893	97·7
15	0·00	1·00	1·00	0·0323	98·9
16	0·00	1·00	0·94	0·0166	99·4
17	0·00	1·00	1·00	0·2783	88·1
18	0·00	1·00	1·00	0·1500	94·3
19	0.00	1·00	0·00	0·1016	95·9
20	0·00	1·00	0·00	0·0167	99·5
21	0·00	1·00	1·00	0·6032	80·3
22	0·00	1·00	1·00	0·5107	81·9
23	0.00	1·00	0·46	0·0263	99·3
24	0·00	1·00	1·00	1·0355	66·2

place feature is based entirely on the voicing feature and would not contribute directly to the response decision. The strength of the interaction model, Model II, can be tested by letting parameter a' vary as in equation 8. Accordingly, when the squared error between the predicted and observed probabilities was obtained by equation 8, a' was estimated to be zero for 20 of the 24 Ss. The estimates of parameter b were identical to those obtained with equation 9 where a' was previously set to zero. This suggests that our original assumption was correct. Ss apparently relied more on the voicing feature than the place feature in the two category bidimensional series.

The extent to which place information enters into the voicing decision for each S is reflected in parameter b from Model II (equation 9). This parameter is greater than zero for all Ss except two, indicating that place information *does* affect the voicing decision, although only in terms of an interaction. Although the fit of the additive model (Model I) is good, the better fit of the interactive model (Model II) and the generally nonzero estimates of the interaction term support the notion that the phonetic features of place and voicing in stop consonants are not combined independently in Stage 3.

Probabilities for the second identification function generated by Ss for the /ba/ to /ta/ series with four response alternatives were also computed. Although this condition was included in the experiment almost as an afterthought, the results were not only surprising but consistent between Ss. The identification function for the same representative S as before in this condition is shown in Fig. 3(d). The high probability of a *P* response for

stimulus 7 and the distribution of *P* responses around this mode is of special interest. If Ss were responding *P* at random, the *Pr*[*P*] in this series should be 0·25 for all stimuli instead of approximately zero at the ends of the stimulus series and one in the middle. This same pattern of *P* responses was found for all Ss tested. The peak probability of a *P* and the stimulus at which it occurred are shown in Table III. When the data for each S is broken down by tape presentation, the same results are observed (see Table III).

In contrast, the *Pr*[*D*] was much lower for all Ss, and when split up by blocks showed greater variability. These data are shown in Table IV. Two Ss failed to report a single /da/ in 330 test trials.

There appears to be some consistency to the /da/ identifications as they generally peaked around stimulus 5. However, this peak reaches 0·5 for only four of the 24 Ss. On the other hand, the occurrence of /pa/ identifications is highly consistent both within and across Ss and the peak probability is less than 0·75 for only one S.

If the phonetic features of place and voicing combined separately and additively in Stage 3 as Model I would predict, the identification functions for this second series should resemble the data for the first /ba/ to /ta/ series. This did not occur as shown in Fig. 3(d). Ss showed consistent use of the /pa/ response in the second /ba/ to /ta/ series at levels well above chance expectation. The peak *Pr*[*P*] in this second bidimensional series occurred at a stimulus whose place value generally corresponded to a high *Pr*[*D*] in the /ba/ to /da/ (place) series (see Table V). Similarly, the peak *Pr*[*P*] stimulus in the bidimensional series

Table III Peak *Pr*[*P*] in the second /ba/ – /ta/ series

Subject	Peak *Pr*[*P*]	*Pr*[*P*] by blocks			Stimulus where peak occurs
		1	2	3	
1	0·967	1·0	1·0	0·9	7
2	0·967	1·0	1·0	0·9	7
3	1·000	1·0	1·0	1·0	7
4	1·000	1·0	1·0	1·0	7
5	0·833	0·9	0·8	0·8	7
6	1·000	1·0	1·0	1·0	7
7	0·867	0·9	1·0	0·7	7
8	1·000	1·0	1·0	1·0	7
9	0·967	1·0	1·0	0·9	7
10	1·000	1·0	1·0	1·0	7
11	0·567	0·2	0·8	0·7	7
12	0·933	1·0	0·9	0·9	7
13	0·933	0·8	1·0	1·0	7
14	1·000	1·0	1·0	1·0	7
15	0·900	0·8	1·0	0·9	7
16	0·933	0·8	1·0	1·0	7
17	0·900	0·9	0·9	0·9	7
18	0·766	0·8	0·6	0·9	7
19	1·000	1·0	1·0	1·0	7
20	0·993	0·9	0·9	1·0	7
21	0·967	1·0	0·9	1·0	7
22	1·000	1·0	1·0	1·0	7
23	0·900	0·8	1·0	0·9	7
24	0·967	0·9	1·0	1·0	7

Table IV Peak *Pr*[*D*] in the second /ba/ – /ta/ series

Subject	Peak *Pr*[*D*]	*Pr*[*D*] by blocks			Stimulus where peak occurs
		1	2	3	
1	0·500	0·9	0·6	0·0	1
2	0·100	0·3	0·0	0·0	8
3	0·500	0·4	0·7	0·4	5
4	0·133	0·0	0·0	0·4	5
5	0·433	0·7	0·3	0·3	5
6	0·467	0·7	0·4	0·3	5
7	0·033	0·1	0·0	0·0	8
8	0·367	0·3	0·5	0·3	5
9	0·067	0·2	0·0	0·0	1
10	0·067	0·1	0·0	0·1	5
11	0·067	0·1	0·0	0·1	5
12	0·533	0·3	0·7	0·6	5
13	0·133	0·2	0·1	0·1	6
14	0·367	0·4	0·4	0·3	5
15	0·567	0·5	0·5	0·6	5
16	0·167	0·3	0·2	0·0	5
17	0·133	0·1	0·0	0·3	5
18	0·467	0·2	0·4	0·8	8
19	0.267	0·1	0·3	0·4	5
20	0.133	0·2	0·1	0·1	8
21	0·000	0·0	0·0	0·0	—
22	0·000	0·0	0·0	0·0	—
23	0·400	0·7	0·3	0·2	5
24	0·067	0·2	0·0	0·0	4

corresponds to a stimulus in the /ba/ to /pa/ (voicing) series which exhibits a high *Pr*[*P*] (see Table V). A model that assumes separate, additive weighting of the features, such as Model I, would predict that the stimulus where the peak *Pr*[*P*] occurs in the second bi-dimensional condition would be categorized as /ta/ and not /pa/.

A model to fit these four response data was constructed based on the reasoning which led to equation 8. However, since Model II did not fit the four response data very well, even when parameter a' was allowed to vary, an additional term was added in which the place feature is dependent on the voicing feature. The term reflecting the dependence of place on voicing ($(1-f_2)f_1$) can now be combined with equations 2, 3, and 4 to yield equation 10:

$$Pr[T] = a_1 Pr[D] + a_2 Pr[P] - b_1(1 - Pr[D])Pr[P] - b_2(1 - Pr[P])Pr[D]. \quad (10)$$

By recombining terms equation 10 can be simplified into equation 11:

$$Pr[T] = (a_1 - b_1)Pr[D] + (a_2 - b_2)Pr[P] + (b_1 + b_2)Pr[P]Pr[D]. \quad (11)$$

If we now substitute a new set of parameter names for the old ones and make the further assumption that the parameters must sum to one, we obtain equation 12:

$$Pr[T] = a' Pr[D] + b' Pr[P] + (1 - a' - b')Pr[P]Pr[D]. \quad (12)$$

Table V Peak *Pr*[*P*] in the second /ba/ – /ta/ series with the corresponding *Pr*[*D*] in the /ba/ – /da/ series and *Pr*[*P*] in the /ba/ – /pa/ series to the corresponding stimulus (number 7)

Subject	Peak *Pr*[*P*]	Corresponding *Pr*[*D*]	Corresponding *Pr*[*P*]
1	0·967	0·933	1·000
2	0·967	0·900	0·833
3	1·000	0·833	1·000
4	1·000	1·000	0·967
5	0·833	0·933	0·967
6	1·000	0·967	0·967
7	0·867	0·233	0·867
8	1·000	0·733	1·000
9	0·967	0·967	1·000
10	1·000	0·567	1·000
11	0·567	0·933	0·967
12	0·933	0·933	1·000
13	0·933	0·067	0·967
14	1.000	0·300	1·000
15	0·900	0·933	1·000
16	0·933	0·833	0·933
17	0·933	0·700	0·933
18	0·767	0·933	0·733
19	1·000	0·600	0·933
20	0·933	0·700	1·000
21	0·967	1·000	0·867
22	1·000	0·800	0·867
23	0·900	0·400	0·867
24	0·967	0·967	1·000

The assumption that the parameters sum to one assures that *Pr*[*T*] will stay between zero and one. Equation 12 has two parameters that must be estimated from the data. Equation 12 generally failed to predict the magnitude of the *Pr*[*P*] in the four response condition, although equation 12 did generally predict a peak at stimulus seven. The fit of equation 12 to one S's data is shown in Fig. 4. The obtained identification function is shown in panel (a); the predicted function derived by equation 12 is shown in panel (b). The response data for all conditions for this same S were shown previously in Fig. 3.

The failure of equation 12 to predict the entire set of probabilities for the second /ba/ to /ta/ series may be attributed to the breakdown of one of our previous assumptions. The processing in Stages 1 and 2 of Fig. 2 may not be independent as we have assumed. Any non-independence of processing, especially in Stage 2 where the phonetic features are extracted, would affect the assumptions made in deriving both Model I and Model II.

In summary, an additive model which assumes independence in the processing of phonetic features cannot account for the identification functions when the acoustic cues underlying place and voicing features in stop consonants are varied systematically. Rather, it appears that an interaction model handles the data much better and provides additional support for evidence previously reported by Lisker & Abramson, and Haggard with different experimental procedures. The perception of an invariant acoustic cue underlying a particular phonetic feature (e.g. place or voicing) may not be invariant with changes in the

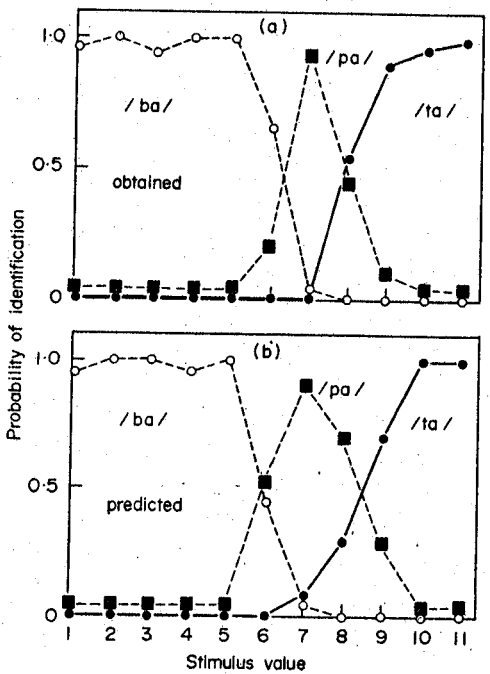

Figure 4 Observed identification function for a representative S in the four response /ba/ to /ta/ series [part (a)] and the predicted function for the same S using equation 12 [part (b)].

acoustic cues underlying other phonetic features. This conclusion is scarcely surprising since covariations in the acoustic cues derive directly from production constraints, and is added evidence of the close link between speech perception and production.

Conclusions

An additive model which assumes independence of processing at all stages did a creditable job in predicting the response probabilities along a bidimensional series of synthetic stop consonants when Ss were constrained to two responses. However, a model which does not assume additivity (i.e. non-independence) of processing in the stage where phonetic features are combined does even better than the independence (i.e. additive) model. When Ss are given four responses from which to choose, the additive model fails completely. In contrast, the non-additive model, while not yielding an excellent fit, does predict occurrence of /pa/ identifications in the /ba/ to /ta/ series. Based on these perceptual data with synthetic speech stimuli, we conclude that the acoustic cues underlying phonetic features in stop consonants are not combined independently to form phonetic segments.

This research was supported in part by a PHS Biomedical Sciences Grant (S05 RR 7031), NIMH research grant MH 24027, and PHS Training Grant (T01 MH 11219-04) to Indiana University and in part by a grant from NICHD to Haskins Laboratories. We wish to thank M. Studdert-Kennedy, R. M. Shiffrin and N. J. Castellan for comments and suggestions on the manuscript.

References

Delattre, P. (1951). The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association of America* **LXVI**, 864–875.

Gerstman, L. J. (1957). *Perceptual Dimensions for the Friction Portions of Certain Speech Sounds.* Unpublished doctoral dissertation, New York University.

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

Liberman, A. M. (1970). Some characteristics of perception in the speech mode. In (D. A. Hamburg, Ed.) *Perception and its disorders, Proceedings of A.R.N.M.D.* Baltimore: Williams and Wilkins.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America* **29**, 117–123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* **74**, 431–461.

Liberman, A. M., Delattre, P. C. & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* **1**, 153–167.

Lisker, L. & Abramson, A. S. (1964). A cross language study of voicing in initial stops: Acoustical measurements. *Word* **20**, 384–422.

Lisker, L. & Abramson, A. S. (1967). Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague.

Miller, G. A. & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* **27**, 338–352.

Shepard, R. N. (1972). Psychological representation of speech sounds. In (E. E. David, Jr. & P. B. Denes, Eds) *Human Communication: A Unified View.* New York: McGraw-Hill.

Shiffrin, R. M. & Geisler, W. S. (1973) Visual recognition in a theory of information processing. In (R. Solso, Ed.) *The Loyola Symposium: Contemporary Viewpoints in Cognitive Psychology.* Washington, D. C.: Winston.

Shiffrin, R. M., Pisoni, D. B. & Casteneda-Mendez, K. (1974). Is attention shared between the ears? *Cognitive Psychology*, **6**, (2), 190–215.

Stevens, K. N. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America* **32**, 47–55.

Stevens, K. N. & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In (W. Wathen-Dunn, Ed.) *Models for the Perception of Speech and Visual Form.* Cambridge, Mass.: M.I.T. Press.

Studdert-Kennedy, M. (1974). The perception of speech. In (T. A. Sebeok, Ed.) *Current Trends in Linguistics*, Vol. XII. The Hague: Mouton.

Studdert-Kennedy, M. & Shankweiler, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America* **48**, 579–594.

Summerfield, A. Q. & Haggard, M. P. (1972). Perception of stop voicing. *Speech Perception*, Series 2, **1**, 1–14. Queen's University of Belfast, Northern Ireland.