Speech Communication 2 (1983) 15–35
North-Holland Publishing Company

CATEGORICAL OR CONTINUOUS SPEECH PERCEPTION: A NEW TEST

Dominic W. MASSARO and Michael M. COHEN

Program in Exp. Psych., University of California, Santa Cruz, Santa Cruz, CA 95064, USA

Received 4 October 1982
Revised 5 January 1983

Abstract. An important question in speech perception is whether listeners have continuous or categorical information about the acoustic signal in speech. Most traditional experimental studies have been interpreted as evidence for categorical perception. It is also argued in the present paper that more recent results taken as evidence against categorical perception are not unequivocally negative. Accordingly, further tests between continuous and categorical views of speech perception are necessary. In the present experiments, listeners were asked for continuous rather than discrete judgments in order to provide a more direct answer to this question. Subjects were asked to rate speech sounds according to where they fell on a particular speech continuum. The continua consisted of stop consonants varying in place (/bæ/ to /dæ/) or voicing (/bæ/ to /pæ/) or a vowel continuum varying from /i/ to /I/. The distributions of rating responses of individual subjects were used to test quantitative models of categorical and continuous perception of acoustic features in speech. The results provide strong evidence against the categorical perception of speech contrasts, and contribute additional evidence for the role of continuous acoustic feature information in speech processing.

Zusammenfassung. Eine wichtige Frage im Zusammenhang mit der Sprachperzeption ist, ob Hörer kontinuierliche oder kategorische Information über das akustische Sprachsignal verarbeiten. Die meisten traditionellen experimentellen Untersuchungen sind als Nachweis kategorischer Perzeption gedeutet worden. Im vorliegenden Artikel wird aber auch darauf hingewiesen, dass neure Ergebnisse, die als Beweis gegen eine kategorische Perzeption ausgelegt wurden, ernst zu nehmen sind. Nur weitere Tests können daher zur Klärung der unterschiedlichen Auffassungen über die Vorgänge bei der Sprachperzeption beitragen. Um eine direktere Antwort auf die aufgeworfene Frage zu erhalten, hatten die Hörer in den im folgenden vorgestellten Experimenten eher kontinuierliche als diskrete Urteile zu fällen. Sie mussten im einzelnen angeben, an welcher Stelle bestimmte Sprachlaute in kontinuierlichen Versuchsreihen erschienen. Die kontinuierliche Veränderung bestand bei Explosivlauten in der Variation ihrer Artikulationsstelle (/bæ/ → /dæ/) oder ihrer Stimmhaftigkeit (/bæ/ → /pæ/); bei Vokalen ging es dagegen um den Übergang von /i/ zu /I/. Die Verteilungen der individuellen Hörerurteile wurde zur Überprüfung quantitativer Modelle kategorischer und kontinuierlicher Perzeption akustischer Merkmale herangezogen. Die Ergebnisse sprechen stark gegen eine kategorische Perzeption von Kontrast im Sprachsignal und liefern weitere Beweise für die Rolle kontinuierlicher Information akustischer Merkmale in der Sprachverarbeitung.

Résumé. Dans la perception de la parole, une question importante consiste à savoir si les auditeurs disposent d'une information continue ou catégorielle sur le signal acoustique de parole. La plupart des expériences traditionnelles ont été interprétées comme supportant la thèse de la perception catégorielle. Ainsi que nous l'exposons dans cet article, des résultats récents, pris comme arguments en défaveur de la perception catégorielle, ne sont pas univoquement négatifs. Des tests supplémentaires permettant de séparer ces deux optiques sont nécessaires. Dans nos expériences, il a été demandé à des auditeurs de porter des jugements continus plutôt que discrets afin de fournir une réponse plus directe à cette question. Les sujets devaient évaluer la localisation de sons de la parole sur un continuum particulier. Les continuums étaient constitués de consonnes occlusives variant selon le lieu d'articulation (de/bæ/vers/dæ/) ou le voisement (de/bæ/vers/pæ/), ou de voyelles variant de /i/ vers /I/. Les distributions individuelles des réponses ont été utilisées pour tester des modèles quantitatifs de perception catégorielle ou continue de traits acoustiques de la parole. Les résultats vont fortement à l'encontre de la perception catégorielle et fournissent des arguments supplémentaires en faveur du rôle de l'information continue sur les traits acoustiques dans le traitement de la parole.

Keywords: Speech perception, categorical perception, ratings, psychophysics.

0167-6393/83/$03.00 © 1983 North-Holland

Introduction

A strong hypothesis in psychological science is rare but valuable [1,2]. One strong hypothesis was presented by Studdert-Kennedy, Liberman, Harris, Cooper when they defined categorical perception as "a mode by which stimuli are responded to and *can only be responded to, in absolute terms*" [3, p. 234]. According their hypothesis, discrimination of certain speech sounds is limited by identification; two different stimuli can be discriminated only to the extent that they are identified differently. Thus, categorical perception can be considered to be the opposite of continuous perception in which discrimination of a set of test stimuli is much better than what might be expected from differential identification. Studdert-Kennedy et al. concluded that there was good evidence for the categorical perception of stop consonants, which supported the involvement of a special decoding device in phonetic perception. During the last 25 years, many experiments have been interpreted as supporting the categorical perception hypothesis [e.g. 4]. More recently, however, negative evidence has been published and the status of categorical perception is much more controversial [e.g. 5]. In our view, the issue is not settled and our goal is to provide an independent test between categorical and continuous perception of speech.

The original experimental study of categorical perception involved the observed relationship between the labeling and discrimination of speech sounds along a synthetic continuum. As an example, Liberman et al. [6] used the pattern playback to generate a series of 14 consonant–vowel syllables going from /be/ to /de/ to /ge/ (/e/ as in gate). The onset frequency of the second formant transition of the initial consonant was changed in equal steps to produce the continuum. In the identification task, observers labeled random presentations of the sounds as b, d, or g. The discrimination task used the ABX paradigm. Three stimuli were presented in the order ABX; A and B always differed and X was identical to either A or B. Observers were instructed to indicate whether X was equal to A or B. This judgment was supposedly based on auditory discrimination in that observers were instructed to use whatever auditory cues they could perceive.

This experiment tested the hypothesis that listeners can discriminate stimuli only to the extent that they can identify them as different phoneme categories. The hypothesis was quantified in order to predict discrimination performance from the identification judgments. According to this formalization, stimuli can be discriminated only to the extent that they are identified as different. The original results were reasonably consistent with the hypothesis, although discrimination was significantly better than that predicted from identification.

Since discrimination was better than that predicted by identification, the results could have been taken as evidence against the categorical perception interpretation. However, researchers at Haskins Laboratories never really accepted this discrepancy as negative evidence. This confirmation bias was proved to be justified 25 years later when Healy and Repp [7] observed that a separate identification task may not be an appropriate measure of the putative identification occurring in the discrimination task. Healy and Repp propose to measure identification directly by also requiring subjects to make identification judgments during the actual discrimination task. Extending their logic even further, however, it might be argued that overt identification can not be taken as an index of the putative identification in the discrimination task. Accordingly, the results showing an advantage of observed discrimination over that predicted by identification are only equivocal evidence against categorical perception.

There have been many studies of categorical perception since the original study of Liberman et al. [6]. Some of the issues have been the differences between speech and nonspeech, differences in consonants and vowels, developmental effects, human versus non-human differences, the sensitivity of the discrimination task, the training and practice of the observers, and the appropriate test of categorical perception. Many of the controversies that have arisen are centered on the use of labelling and discrimination tasks. In many cases, the relationship between the two tasks is the important index of categorical perception. As an example, Macmillan, Kaplan, and Creelman [8] have expanded on this relationship within the framework of Thurstone's Theory of Comparative

Judgment and Signal Detection Theory. Macmillan et al. propose that categorical perception occurs if the perceived spacing of signals along a dimension is the same for discrimination and for identification. We believe that this proposal and other recent research have lost sight of the original question. The central question should be whether continuous changes along some stimulus dimension result in relatively continuous or discrete changes along a perceptual dimension. In the first case, we call perception continuous, in the second case, perception is referred to as discrete or categorical.

More recent research has been interpreted as evidence that listeners can transmit information about the degree to which a given acoustic feature is present in a speech sound. Barclay [9,10] tested a /b/, /d/, /g/ continuum under two different sets of conditions. First, listeners were asked to identify each stimulus as one of the three alternatives b, d, or g. Second, the possible response alternatives were reduced to b or g. The results of interest were the identification responses in the second task to the stimuli identified as d in the first task. Barclay argued that if perception is categorical, the probability of a b or g response should be independent of the acoustic representation of sounds identified as d in the first task. In fact, the results revealed that the likelihood of a b response was greater for sounds toward the /b/ end of the continuum. Barclay interpreted the results to imply that the listener could discriminate between stimuli that belong to the response category d.

Miller [11] 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 VOT values of the stop presented to the unmonitored ear. Miller [11] interpreted this result 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 multi-valued outputs of feature detectors, they do not disprove the possibility of all-or-none outputs. The finding of a relatively continuous identification function as a function of some stimulus property does not distinguish between continuous and categorical feature outputs. As assumed by the original categorical perception model [6], 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 the VOT of the stop presented to the other ear might simply reflect the likelihood of the appropriate feature detector firing in an all-or-none manner. Increasing the VOT of the non-monitored ear would change the probability of firing.

In a second study by Miller, adaptation with repeated presentations of a voiceless stop decreased voiceless responses as a direct function of the VOT of the adapting stimulus. However, the effectiveness of an adapting stimulus as a function of its VOT value can simply reflect the probability that the stop is categorically perceived as completely voiced or completely voiceless on each successive presentation in the adapting series. McNabb [43] also interpreted some selective adaptation results in terms of the continuous output of a phonetic feature detector. The first three sounds along a seven-step /ba/ to /da/ continuum were used as adaptors in different adaptation sessions. The results showed more adaptation with the more extreme adaptor. This result was interpreted in terms of a continuous output of a phonetic feature detector. Regardless of the explanatory status of a phonetic feature detector, the result provides no firm evidence for continuous information in speech perception. Analogous to our interpretation of Miller's [11] results, differential effectiveness of the adaptors might reflect only the probability that a given adaptor is heard as a given binary alternative on each adapting presentation. In this case, hearing the adaptor as /ba/ would be more likely with the more extreme than with the less extreme adaptor. Accordingly, more adaptation would be expected with the more extreme adaptor.

Reaction times to speech stimuli along a continuum might be relevant to the issue of discrete versus continuous speech perception. Studdert-Kennedy, Liberman, and Stevens [12], Pisoni and

Tash [13], and Repp [14] have all found an increase in identification times as the sound moves closer to the category boundary. If sounds within a category can not be discriminated, then the time for a particular category identification should not differ as a function of the different sounds in the category. Differences in identification reaction times would indicate that, at some level, the sounds were not processed equivalently. However, a proponent of discrete perception could always argue that the listener does *not* have access to the processing time information or does not use it, and therefore, identification reaction times are not relevant to the issue of discrete versus continuous speech perception. Although this explanation places an additional burden on the discrete view, additional results are needed for a convincing test.

Pisoni and Tash [13] used reaction times in a discrimination task to look for evidence for continuous perception of speech sounds. Subjects were given two successive speech sounds and told to respond according to whether the sounds had the same or different names. Given that the listeners were asked to respond on the basis of category class, the acoustic similarity between the two sounds should have no effect if the comparison is actually being made at an abstract phonetic level. That is, responding 'same' to two different /ba/'s should take no longer than responding 'same' to two identical /ba/'s. In fact, the results showed a faster 'same' response to physically identical sounds than to sounds acoustically different but within the same speech category. In addition, 'different' responses to sounds of two different categories were slower with smaller acoustic differences between the sounds. Once again, reaction times seem to challenge the notion of discrete perception, although the idea that listeners do not have access to the processing time (or some correlate of it) can salvage the discrete view.

Other results indicate that subjects can discriminate between different members of a speech category [15,16,17]. For example, Samuel [17] found that discrimination in an ABX task was better than that predicted from a separate identification task. These results have been interpreted as evidence against the traditional categorical perception view. However, in all cases of an advantage of observed discrimination over predicted, it remains possible that the discrepancy may be due to an inappropriate measure of identification [7]. Therefore, previous results are not definitive in their rejection of categorical perception theory.

A new test

Our goal in the present research is to offer a new approach to the question of categorical perception. Our approach is similar to one taken in psychophysics to test between threshold and continuous theories of signal detection [18]. The distinguishing feature of this approach is the use of continuous rather than discrete perceptual judgments. Although rating judgments have been used in a number of previous studies [19,20] they have not been analyzed in such a way to test between categorical and continuous models of speech perception. Relative to discrete judgments, continuous judgments may provide a more direct measure of the listener's perceptual experience. For example, McNabb [43] found that a binary response proved insensitive to the manipulation of an independent variable whereas confidence ratings

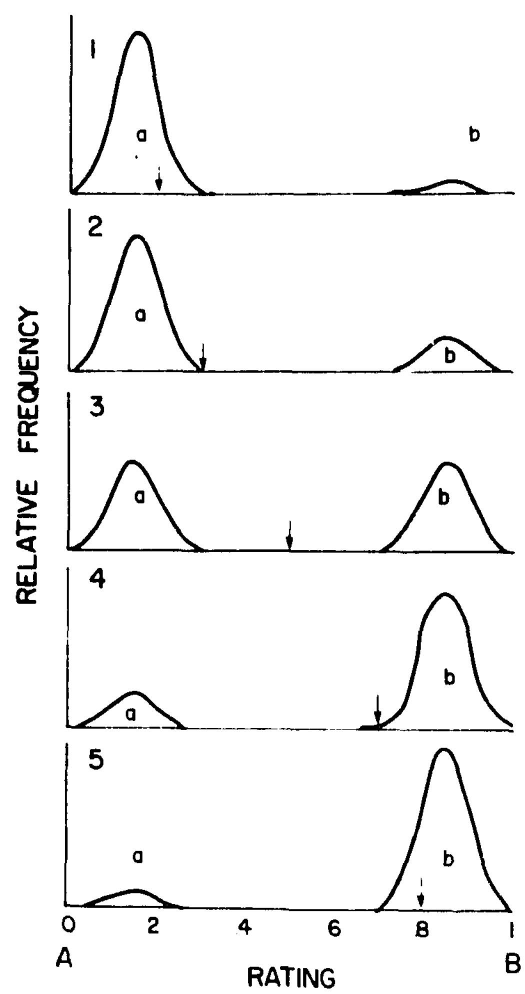

Fig. 1. Illustration of the categorical perception rating model.

revealed significant effects of this variable. In the present study, subjects were asked to rate the degree to which they felt that the speech stimulus represented one alternative or the other, rather than simply asked to indicate which alternative was present. Although there has been considerable debate on the interpretation of rating judgments [21,22] this method has been extended in our research to reduce the ambiguity of interpretation. More specifically, we have formalized categorical and continuous models of speech perception and have contrasted their predictions with respect to the distribution of repeated rating responses to the test stimuli.

Consider the assumptions of the categorical perception model illustrated in Fig. 1. It is assumed that the listener has only two perceptual states, *a* or *b*, along a sound continuum of five levels. At level 1, the likelihood of an *a* percept is very high whereas the likelihood of a *b* percept is very low. As the levels increase, the relative likelihood of the two percepts change so that at level 5, *b* is the most likely percept. But in all cases, the sound is heard as either *a* or *b*. If perception is truly categorical, any sound along the continuum can be heard only as *a* or *b* and nothing in between. What does a categorical perception subject do when asked to make continuous rating judgments? He or she might note the foolishness of the request, but most likely would attempt to comply in a reasonable manner. The subject would choose a rating towards the *A* end of the response scale for the perception of *a* and a rating towards the *B* end for the perception of *b*. If there is variability in memory and response, however, the subject would generate a distribution of rating responses for each of the two percepts. That is, subjects may not remember where they last rated the *a* category and they may also only approximate the intended rating because of response variability. Furthermore, given the demand characteristics of the task, subjects might actually generate additional variability in their ratings if their percepts were categorical but they were expected to make a range of rating responses. Accordingly, the rating responses to the percept *a* would be normally distributed, with a mean *Xa* and a variance *Sa* and similarly for the *b* percept.

The important question is how the rating responses are expected to differ as a function of the speech stimulus. Consider the speech continuum of five levels as illustrated in Fig. 1. Although perception is categorical, a stimulus is more likely to produce the percept *a* to the extent that it is away from the categorical boundary and towards the *A* end of the continuum. Variation in the categorical boundary, or variation in the perceptual system, or both allows the percept to have only a probabilistic relationship to a given stimulus. A given stimulus produces the percept *a* with probability *Pa* and produces the percept *b* with probability $1 - Pa$. Therefore, the distribution of rating responses to a given stimulus will actually be a mixture of ratings generated by the two percepts. The proportion of ratings generated from the percept *a* will increase with increases in *Pa*, the likelihood of the percept *a*. Similarly, the proportion of ratings generated from the percepts *b* will decrease with increases in *Pa*. The arrows in Fig. 1 give the mean rating responses resulting from the mixture of the two distributions over trials. Fig. 2 illustrates the continuous changes in the mean rating response with changes in stimulus level predicted by this model. Thus, the continuous changes in mean ratings with changes in stimulus level shown in Fig. 2 might occur even though perception is categorical.

The continuous model is illustrated in Fig. 3. In this model, the rating given to a given stimulus is a direct function of the percept generated by that stimulus. This model is similar to Thurstone's [23]

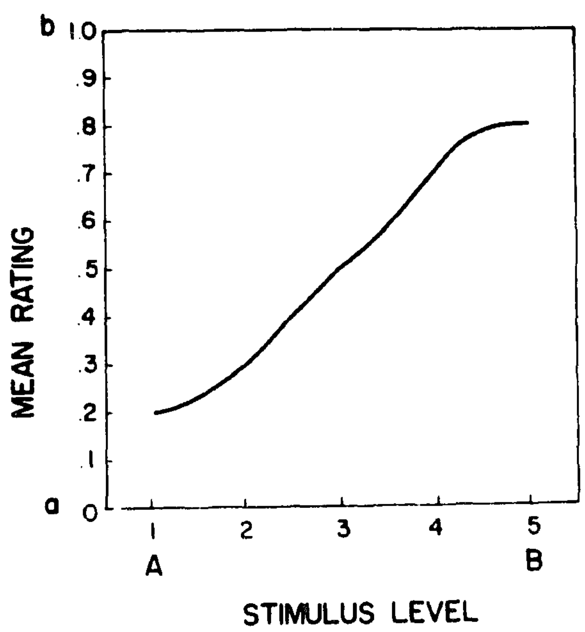

Fig. 2. Hypothetical mean rating responses predicted by both categorical and continuous models.

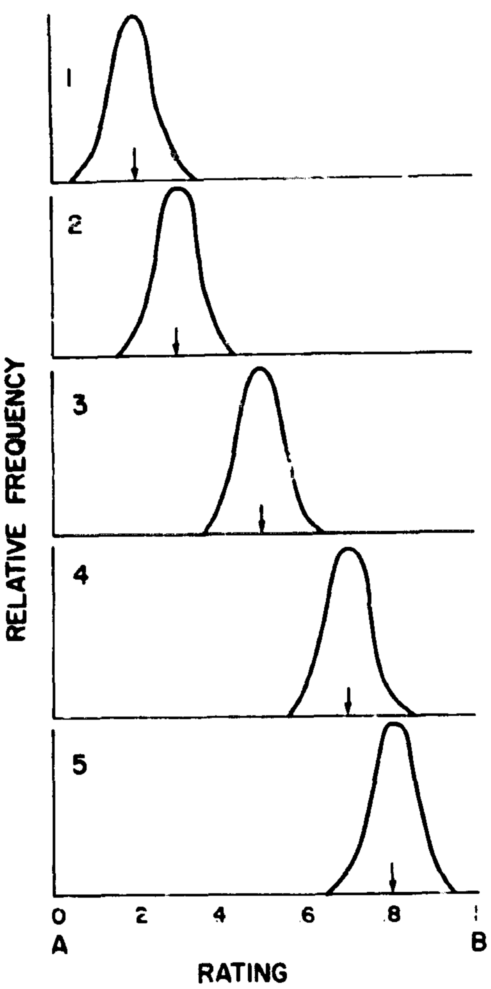

Fig. 3. Illustration of the continuous perception rating model.

law of comparative judgment in which each stimulus is seen as giving rise to a normal distribution along an internal dimension. In the continuous model, the percepts of two adjacent stimuli will usually differ from each other and the rating responses will reflect this fact. The percept of a stimulus towards the *A* end of the continuum will be more *A*-like than that of a neighboring stimulus towards the *B* side of the continuum. Random variability in the perceptual, memory, or response systems will also result in a distribution of rating responses to any given stimulus. Fig. 3 shows how the continuous model predicts a systematic change in average rating responses with changes in stimulus level. Therefore, the hypothetical results in Fig. 2 are equally consistent with the continuous model.

We have seen that the categorical and continuous perception models make similar predictions about the average rating judgments. Therefore, the mean judgments are not capable of distinguishing between the two models. The models might, however, be distinguished on the basis of the actual distribution of rating responses. The final distribution of rating responses is predicted to differ for the two models. Figs. 1 and 3 illustrate the overall form of the predicted distribution of rating responses for each of the two models. As can be seen in the figures, although the average rating function (indicated by the arrows) is identical for the two models, the distribution of rating responses are not. For example, at level 3 on the stimulus continuum, the continuous model would predict a single, central distribution, while the categorical model would predict a bimodal distribution with a central trough. In actual experimentation, we might be able to discriminate between the models by evaluating the distribution of rating responses and, therefore, gain some insight into the issue of categorical perception of speech.

In the actual tests of the models we make the following assumptions. The ratings for a given percept are assumed to be normally distributed. In addition, any potential rating response that would fall outside the range of the rating scale, given the normal distribution, was assumed to be placed at the appropriate end of the rating scale. This phenomenon has been termed the *end effect* in previous scaling research [24]. For example, if the percept generated by a stimulus was even more *A*-like than the most extreme *A* rating, than the rating would be placed at the *A* end of the scale. In addition, subjects might remember their rating to a given percept as being farther to the right than the actual right end of the scale. In this case, subjects would simply respond at the right end of the scale.

Experiment

To gather data for the model tests we had three groups of subjects rate continua of synthetic speech stimuli. One group rated consonants differing in place of articulation from /bæ/ (as in bat) to /dæ/, a second group rated consonants differing in voicing from /bæ/ to /pæ/, and a third group rated vowels on a continuum from /i/ (as in heat) to /I/ (as in hit).

Method

Subjects. The subjects were three groups of 12 introductory psychology students who participated to fulfill a course requirement.

Stimuli. The stimuli were produced during the experiment proper by a formant series resonator speech synthesizer (FONEMA OVE-IIId) under the control of a PDP-8/L computer [25]. Each stimulus was specified as a series of lists of parameter vectors. Each parameter vector specified the target value of a parameter, the transition time, and the transition type. Transitions could be linear or negatively accelerated. Each list specified the amount of time until the next list would take control. Time values were specified and parameters calculated in 5-msec. increments. The first pulse of voicing was synchronized to begin at the onset of the test stimulus.

Three continua of speech sounds were generated. For the place continuum the stimuli were consonant–vowel (CV) syllables simulating labial and alveolar stop consonants and the vowel /æ/. The CV can be represented as a consonant transition followed by a final vowel segment. The final vowel segment had $F1$, $F2$, $F3$, $F4$ and $F5$ set at 734, 1600, 2851, 3500, and 4000 Hz, respectively, and the amplitude of the buzz source simulating vocal cord vibration (AV) was set at 16 dB. (The amplitudes given here are the OVE-IIId amplitude parameters used, and not the levels at the subjects' ears. They are given for the benefit of those readers who are familiar with the OVE.) For the CV transition $F1$, $F2$, and $F3$ moved to the vowel configuration in 40 msec. following a negatively accelerated path. The initial $F2$ and $F3$ frequencies for the seven levels of place are given in Table 1. The initial value of $F1$ was always 200 Hz. The onset of voicing energy for the consonant was instantaneous and the offset of the vowel had a linear amplitude drop taking 30 msec. The fundamental frequency ($F0$) of the initial vowel was set at 133 Hz, and fell linearly to 126 Hz during the last 100 msec of the final vowel. The vowel segment duration was always 170 msec, from the end of the CV transition to the beginning of the final fall of AV.

It should be noted that the changes in the formant values along the place continuum were not in equal steps. However, we believe that the concern about equal step size is misguided. The important issue is whether continuous changes along the stimulus dimension produce continuous changes along the perceptual dimension.

For the voicing continuum the stimuli were also CV syllables, but varied in VOT with place fixed at the most /bæ/-like level from the place condition. To create syllables with VOT values greater than 0 msec a period of aspiration was created by sending the noise source through the vowel formants at 14 dB. The seven VOT values used were 0, 10, 20, 30, 40, 50, and 60 msec.

For the vowel continuum the stimuli were seven 205 msec vowels varying between /i/ and /I/. The rise and fall times (included in the 205 msec) for the vowel were each 30 msec. Initially the $F0$ of the vowel was 126 Hz. After 150 msec $F0$ fell linearly to 112 Hz over a 40 msec period. Table 2 gives the $F1$, $F2$, and $F3$ values used for the seven vowels.

Procedure. All experimental events were controlled by a PDP-8/L computer. The output of the speech synthesizer was bandpass filtered 20–5000 Hz (KROHN-HITE Model 3500 R), amplified (McIntosh Model MC-50) and presented over KOSS PRO-4AA headphones at a comfortable listening level (about 72 dB SPL-B). Four subjects could be tested simultaneously in individual sound attenuated rooms.

Table 1
$F2$ and $F3$ values used for the Place Continuum

Place	$F2$	$F3$
1 b	1199	1958
2	1307	2198
3	1425	2397
4	1510	2614
5	1600	2851
6	1745	3109
7 d	1903	3390

Table 2
$F1$, $F2$ and $F3$ values used for the Vowel Continuum

Vowel	$F1$	$F2$	$F3$
1 i	267	2329	3109
2	291	2263	2934
3	308	2198	2851
4	327	2136	2770
5	346	2075	2691
6	367	2016	2614
7 I	389	1958	2540

Each trial began with the presentation of a stimulus, selected randomly without replacement in blocks of 7 trials. Each observer then made a response by setting the pointer of a linear control 5.5 cm long. For place judgments, the left end of the control was labeled 'B' and the right end 'D'. Similarly, in the labels were 'B' and 'P' for voicing, and 'EE' and 'IH' for vowel judgments. Once the observer was satisfied with the position of the pointer he or she pressed a small button to the right of the scale. The computer then read the position of the control via an analog-to-digital converter, recording the response as an integer value between 0 and 49. The computer waited until all observers had responded (on the average about 1.5 seconds) before proceeding. The next trial began 1.0 second later.

Independent groups of 12 subjects were tested under each speech condition. Each subject participated in two 15 minute sessions on each of two successive days, with a 5 minute break between the two sessions. Subjects rated 249 stimuli during each of the 4 sessions. The first 25 trials were unscored practice trials, selected randomly without replacement in blocks of 7. Responses were recorded for 32 blocks of 7 stimuli, sampled randomly without replacement. Thus, a total of 896 ratings were collected for each subject, 128 measures for each of the 7 stimuli.

Before the first session subjects listened to the set of 7 stimuli played in order ten times. The subjects were first told, for example, "I am going to play for you some of the syllables going from ··· the best /bæ/ to the best /dæ/ ··· we want you to use this sequence to help anchor your ratings. That is to say, the first stimulus /bæ/ is as /bæ/-like as any that you will hear later ··· so if you hear something as /bæ/-like as this you should set your pointer all the way to the left. Similarly, the seventh stimulus /dæ/ is as /dæ/-like as any ··· The other five stimuli in the sequence are equally spaced between the two extremes. ··· We want you to use the whole scale to respond with, not just the two end points and middle, for example. For the syllables you will hear you should be using the entire scale and all of the points in it". Also, the subjects were told that the stimuli were presented in random order, with no patterns for them to guess.

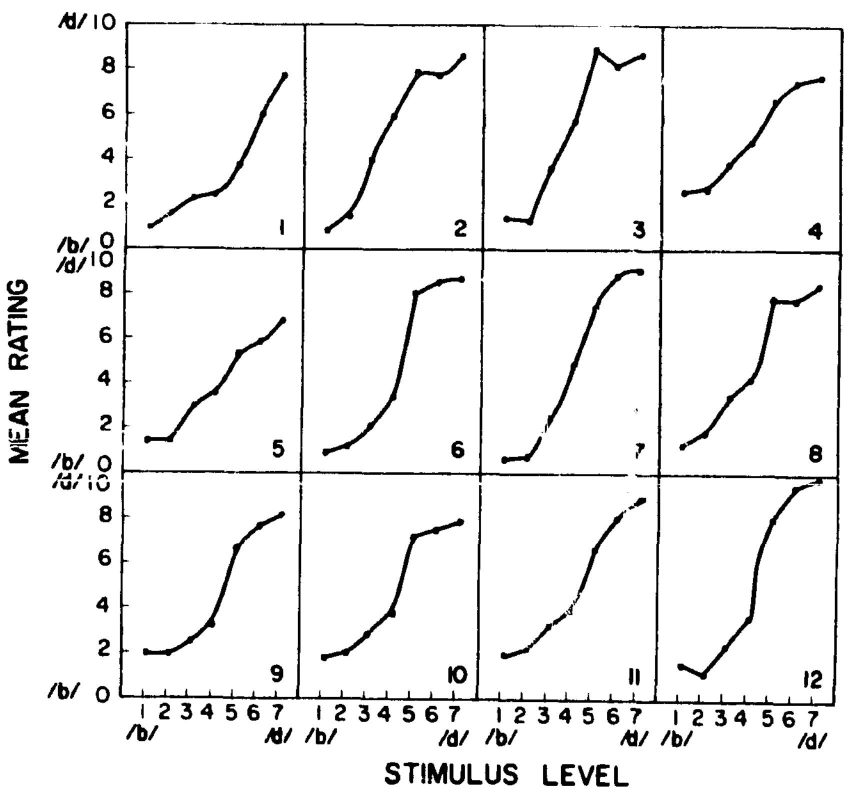

Fig. 4. Mean /bæ/–/dæ/ ratings for each of the 12 individual subjects from the place condition.

Results and analysis

Figs. 4, 5 and 6 show the mean ratings for individual subjects in the place, voicing and vowel conditions, respectively. As can be seen, the ratings changed relatively continuously as a function of the stimulus level. In Fig. 4, for example, the ratings changed relatively gradually from the B

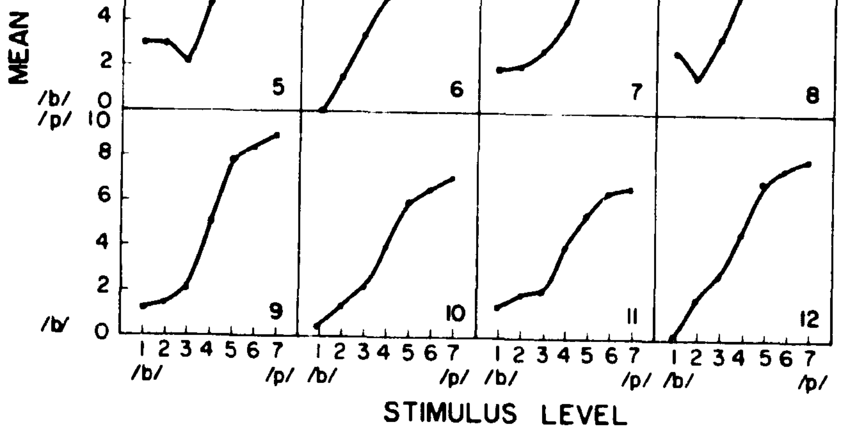

Fig. 5. Mean /bæ/–/pæ/ ratings for each of the 12 individual subjects from the voicing condition.

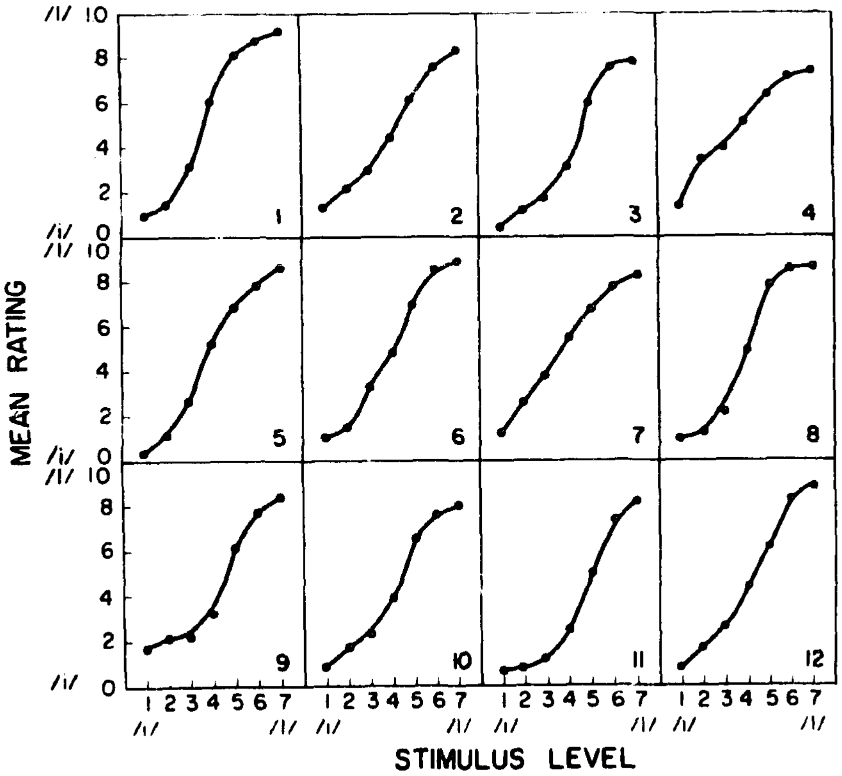

Fig. 6. Mean /i/–/I/ ratings for each of the 12 individual subjects from the vowel condition.

end of the response scale to the D end with changes of the stimulus from /bæ/ to /dæ/.

As discussed in the Introduction, however, the critical feature of our analysis is not the examination of the marginal means of the ratings but rather the exact nature of their distribution of occurrence. In order to determine whether the observed distributions of ratings were best fit by the continuous or categorical models the computer program STEPIT [26] was used. A model is represented to the analysis program STEPIT as a set of prediction equations which contain a set of unknown parameters. By iteratively adjusting the parameters of the model, STEPIT minimizes the chi-square deviations between the observed and predicted values. Thus, what STEPIT does is to find a set of parameter values which when put in a model, come closest to describing the observed data. We can discriminate between competing models on the basis of the overall goodness of fit. For the continuous model there were 14 parameters which included 7 means and 7 standard deviations for the 7 normal distributions corresponding to the 7 stimuli. The initial values of the STEPIT parameters were set equal to the observed means and standard deviations. The binary model had 11 parameters which included 2 means, 2 standard deviations for the 2 normal distributions and 7 sampling probabilities. For the initial values of the STEPIT parameters in the fit of the binary model, the means were set to 0.85 and 0.15, the standard deviations were both set to 0.15 and the probabilities were set equal to the observed means. It should be noted that the initial values chosen for the parameters did not influence the goodness of fit of either model. They were simply chosen in order to expedite the operation of STEPIT. Before modeling, the observed data were grouped into 25 bins for each stimulus level. This reduction was done to increase the number of cases per bin and to decrease somewhat the variability of the data.

The most straightforward method of computing the total chi-square for each subject would be simply to take the sum of chi-squares for each of the 25 cells in each of the 7 stimulus distributions. Unfortunately, this method of computing chi-square is inappropriate, because some cells are predicted to contain 0 instances. In order to obtain a good chi-square measure it is usually necessary to include at least 5 predicted occurrences [27, p. 293]. To meet this restriction, the chi-square terms within each stimulus level were derived by pooling together (if necessary) several bins of data to ensure that the predictions include 5 or greater cases. The exact method was as follows: Starting at the 0 end of the distribution each predicted cell was checked to see if it contained 5 or more ratings. If it did, a chi-square term was computed. If there were fewer than 5 cases, the cell was pooled with the next cell or cells until at least 5 ratings were obtained. Once 5 or more were obtained, a chi-square term was computed on this pooled prediction versus the number of observed ratings pooled across the same bins. If the last bin was reached with an insufficient number of predicted ratings, the final data were pooled with the cell (original or pooled) just below it and the chi-square term for that cell was recomputed. On the average, about 8–10 such pooled bins were created out of the original 25 bins for each stimulus level distribution. It should be noted that this dynamic pooling method sometimes resulted in a different number of bins being used for the same observed data depending on the theoretical model being tested.

Tables 3 and 4 give the optimal parameter values found by STEPIT for the three sets of data for the continuous and binary models, respectively.

Table 3
Continuous model parameters for place

S	x1	x2	x3	x4	x5	x6	x7	sd1	sd2	sd3	sd4	sd5	sd7	sd7
1	0.013	0.031	0.045	0.076	0.310	0.688	0.935	0.265	0.354	0.520	0.579	0.547	0.554	0.475
2	0.046	0.045	0.385	0.591	0.860	0.827	0.867	0.152	0.287	0.474	0.378	0.170	0.006	0.122
3	0.117	0.101	0.327	0.571	0.990	0.968	0.999	0.159	0.163	0.312	0.351	0.267	0.154	0.122
4	0.261	0.246	0.388	0.466	0.656	0.724	0.772	0.129	0.136	0.249	0.265	0.211	0.155	0.143
5	0.143	0.148	0.278	0.334	0.558	0.603	0.729	0.078	0.097	0.276	0.287	0.305	0.229	0.119
6	0.061	0.074	0.196	0.324	0.825	0.867	0.871	0.170	0.189	0.204	0.316	0.163	0.142	0.144
7	0.215	0.066	0.134	0.452	0.770	0.899	0.948	0.001	0.064	0.141	0.421	0.283	0.142	0.130
8	0.112	0.162	0.344	0.416	0.818	0.806	0.882	0.174	0.195	0.229	0.218	0.242	0.207	0.223
9	0.208	0.199	0.271	0.330	0.678	0.765	0.813	0.108	0.122	0.177	0.188	0.226	0.153	0.121
10	0.175	0.220	0.292	0.384	0.709	0.742	0.790	0.139	0.116	0.135	0.104	0.123	0.109	0.143
11	0.207	0.219	0.314	0.376	0.657	0.813	0.912	0.149	0.183	0.177	0.215	0.370	0.166	0.131
12	0.121	0.031	0.172	0.329	0.847	0.999	0.999	0.326	0.300	0.338	0.386	0.274	0.133	0.078
X	0.140	0.128	0.262	0.387	0.723	0.808	0.876	0.154	0.184	0.269	0.309	0.265	0.179	0.163

Continuous model parameters for voicing

S	x1	x2	x3	x4	x5	x6	x7	sd1	sd2	sd3	sd4	sd5	sd6	sd7
1	0.135	0.162	0.177	0.440	0.741	0.765	0.280	0.129	0.113	0.137	0.295	0.147	0.111	0.101
2	0.231	0.342	0.455	0.730	0.792	0.804	0.807	0.051	0.194	0.213	0.142	0.070	0.068	0.070
3	0.000	0.000	0.000	0.799	0.999	0.999	0.999	0.361	0.182	0.745	0.915	0.000	0.204	0.159
4	0.161	0.172	0.189	0.493	0.585	0.644	0.702	0.067	0.105	0.181	0.189	0.178	0.184	0.194
5	0.308	0.292	0.217	0.476	0.753	0.791	0.819	0.119	0.142	0.182	0.324	0.178	0.132	0.153
6	0.210	0.137	0.316	0.504	0.680	0.765	0.792	0.004	0.074	0.184	0.218	0.182	0.207	0.200
7	0.175	0.194	0.242	0.381	0.703	0.776	0.821	0.110	0.132	0.158	0.292	0.195	0.146	0.135
8	0.298	0.122	0.317	0.572	0.824	0.846	0.839	0.197	0.158	0.321	0.368	0.218	0.117	0.133
9	0.099	0.140	0.166	0.507	0.766	0.832	0.937	0.138	0.113	0.152	0.243	0.169	0.132	0.183
10	0.000	0.113	0.180	0.402	0.568	0.641	0.707	0.113	0.138	0.128	0.206	0.159	0.181	0.174
11	0.079	0.148	0.176	0.379	0.544	0.649	0.678	0.236	0.227	0.173	0.389	0.401	0.323	0.322
12	0.009	0.167	0.256	0.452	0.684	0.698	0.791	0.018	0.150	0.106	0.166	0.139	0.136	0.148
X	0.142	0.165	0.224	0.511	0.720	0.767	0.808	0.129	0.144	0.223	0.312	0.170	0.162	0.164

Table 3 (contd.)
Continuous model parameters for vowels

S	x1	x2	x3	x4	x5	x6	x7	sd1	sd2	sd3	sd4	sd5	sd6	sd7
1	0.006	0.103	0.288	0.593	0.796	0.874	0.921	0.195	0.169	0.222	0.258	0.155	0.146	0.143
2	0.115	0.199	0.288	0.431	0.594	0.748	0.824	0.106	0.090	0.118	0.142	0.159	0.109	0.114
3	0.000	0.088	0.161	0.299	0.569	0.746	0.756	0.094	0.114	0.127	0.174	0.196	0.141	0.126
4	0.219	0.323	0.373	0.491	0.630	0.701	0.724	0.155	0.180	0.220	0.221	0.180	0.148	0.126
5	0.000	0.020	0.229	0.515	0.692	0.799	0.899	0.083	0.234	0.286	0.315	0.288	0.226	0.193
6	0.076	0.122	0.307	0.472	0.684	0.857	0.923	0.124	0.149	0.189	0.223	0.199	0.168	0.183
7	0.099	0.239	0.356	0.541	0.665	0.772	0.809	0.113	0.163	0.199	0.187	0.178	0.148	0.164
8	0.073	0.095	0.193	0.486	0.785	0.832	0.848	0.129	0.166	0.173	0.274	0.177	0.132	0.127
9	0.148	0.197	0.205	0.304	0.583	0.759	0.816	0.150	0.129	0.126	0.150	0.193	0.156	0.144
10	0.059	0.150	0.201	0.400	0.648	0.751	0.781	0.075	0.111	0.130	0.204	0.141	0.099	0.093
11	0.049	0.061	0.112	0.213	0.481	0.732	0.818	0.063	0.078	0.119	0.165	0.205	0.165	0.124
12	0.073	0.155	0.247	0.428	0.610	0.807	0.875	0.095	0.085	0.109	0.171	0.145	0.086	0.117
X	0.076	0.146	0.247	0.431	0.645	0.782	0.833	0.115	0.139	0.168	0.207	0.185	0.144	0.138

Table 4
Binary model parameters for place

S	x1	x2	sd1	sd2	p1	p2	p3	p4	p5	p6	p7
1	0.8788	0.0187	0.4614	0.3851	0.0120	0.0001	0.1142	0.1535	0.3457	0.7905	0.9999
2	0.8593	0.0715	0.1309	0.1833	0.0001	0.0264	0.3631	0.6098	0.9433	0.8999	0.9714
3	0.9672	0.1665	0.2399	0.2131	0.0001	0.0002	0.2090	0.6180	0.9912	0.9999	0.9999
4	0.7216	0.2400	0.1436	0.1362	0.0135	0.0069	0.2728	0.4811	0.8812	0.9999	0.7589
5	0.7169	0.1532	0.1164	0.1147	0.0003	0.0038	0.2401	0.3464	0.6667	0.7978	0.8799
6	0.8458	0.1313	0.1591	0.2100	0.0001	0.0001	0.0481	0.2183	0.9975	0.9969	0.9999
7	0.8940	0.0905	0.2140	0.0900	0.0001	0.0004	0.1874	0.6401	0.9207	0.9773	0.9999
8	0.8067	0.2213	0.2254	0.2400	0.0004	0.0001	0.2648	0.3910	0.9848	0.9789	0.9999
9	0.7777	0.2412	0.1590	0.1503	0.0003	0.0001	0.0311	0.0764	0.8689	0.9935	0.9999
10	0.7523	0.2792	0.1344	0.1587	0.0001	0.0001	0.0001	0.0007	0.9808	0.9997	0.9997
11	0.8493	0.2412	0.1807	0.1744	0.0001	0.0001	0.0369	0.0819	0.7626	0.9806	0.9998
12	0.9808	0.1387	0.1650	0.3969	0.0087	0.0003	0.0090	0.2199	0.7797	0.9999	0.9999
X	0.8375	0.1661	0.1942	0.2044	0.0030	0.0032	0.1481	0.3198	0.8436	0.9512	0.9674

Table 4 (contd.)
Binary model parameters for voicing

S	x1	x2	sd1	sd2	p1	p2	p3	p4	p5	p6	p7
1	0.7824	0.1622	0.1081	0.1233	0.0003	0.0230	0.0781	0.3986	0.8868	0.9805	0.9999
2	0.7904	0.2996	0.0770	0.1715	0.0001	0.0441	0.1940	0.7995	0.9999	0.9999	0.9999
3	0.9969	0.0060	0.4094	0.3998	0.0002	0.0002	0.0127	0.7314	0.9998	0.9997	0.9901
4	0.6042	0.1555	0.1956	0.0919	0.0312	0.0474	0.2394	0.9313	0.9999	0.9999	0.9999
5	0.7923	0.2584	0.1491	0.1690	0.0001	0.0001	0.0001	0.4793	0.9564	0.9999	0.9999
6	0.6809	0.1041	0.2242	0.1894	0.0003	0.0001	0.1706	0.5022	0.9999	0.9972	0.9999
7	0.7655	0.2057	0.1608	0.1393	0.0001	0.0005	0.0155	0.3446	0.8852	0.9999	0.9999
8	0.8529	0.1841	0.1043	0.1911	0.0246	0.0002	0.1800	0.5276	0.8286	0.9136	0.9268
9	0.8201	0.1666	0.1503	0.1500	0.0001	0.0001	0.0468	0.5232	0.9655	0.9908	0.9998
10	0.6149	0.0879	0.1879	0.1680	0.0001	0.0036	0.0433	0.5906	0.9999	0.9980	0.9999
11	0.7737	0.1498	0.2028	0.2108	0.0001	0.0270	0.0251	0.3668	0.5426	0.7776	0.7362
12	0.6863	0.1376	0.1512	0.2103	0.0002	0.0005	0.0001	0.4492	0.9999	0.9997	0.9999
X	0.7634	0.1598	0.1767	0.1845	0.0048	0.0122	0.0838	0.5537	0.9220	0.9714	0.9710

Binary model parameters for vowels

S	x1	x2	sd1	sd2	p1	p2	p3	p4	p5	p6	p7
1	0.8459	0.1426	0.1763	0.2406	0.0002	0.0003	0.0727	0.5782	0.9679	0.9999	0.9999
2	0.7044	0.2280	0.1639	0.1603	0.0001	0.0001	0.0001	0.3563	0.8904	0.9999	0.9999
3	0.6926	0.1211	0.1630	0.1793	0.0002	0.0001	0.0001	0.2356	0.7736	0.9999	0.9999
4	0.6917	0.2492	0.1481	0.1200	0.0418	0.2141	0.2742	0.5777	0.8295	0.9994	0.9999
5	0.7992	0.0128	0.3414	0.2579	0.0002	0.0008	0.1700	0.6788	0.9022	0.8447	0.9023
6	0.8031	0.1891	0.2198	0.2002	0.0001	0.0001	0.0777	0.3894	0.9659	0.9999	0.9999
7	0.7308	0.2135	0.1746	0.1800	0.0003	0.0192	0.2540	0.5344	0.9267	0.9998	0.9997
8	0.8096	0.1415	0.1524	0.1855	0.0001	0.0001	0.0340	0.4993	0.9666	0.9999	0.9999
9	0.7502	0.2185	0.1700	0.1551	0.0001	0.0001	0.0520	0.0520	0.6870	0.9999	0.9999
10	0.7241	0.1577	0.1226	0.1672	0.0001	0.0001	0.0124	0.2771	0.8783	0.9999	0.9999
11	0.7325	0.1040	0.1814	0.1467	0.0006	0.0001	0.0001	0.1422	0.5179	0.9999	0.9999
12	0.7456	0.2037	0.1500	0.1596	0.0001	0.0001	0.0001	0.2907	0.6077	0.9999	0.8487
X	0.7525	0.1651	0.1803	0.1794	0.0037	0.0196	0.0746	0.3843	0.8261	0.9869	0.9792

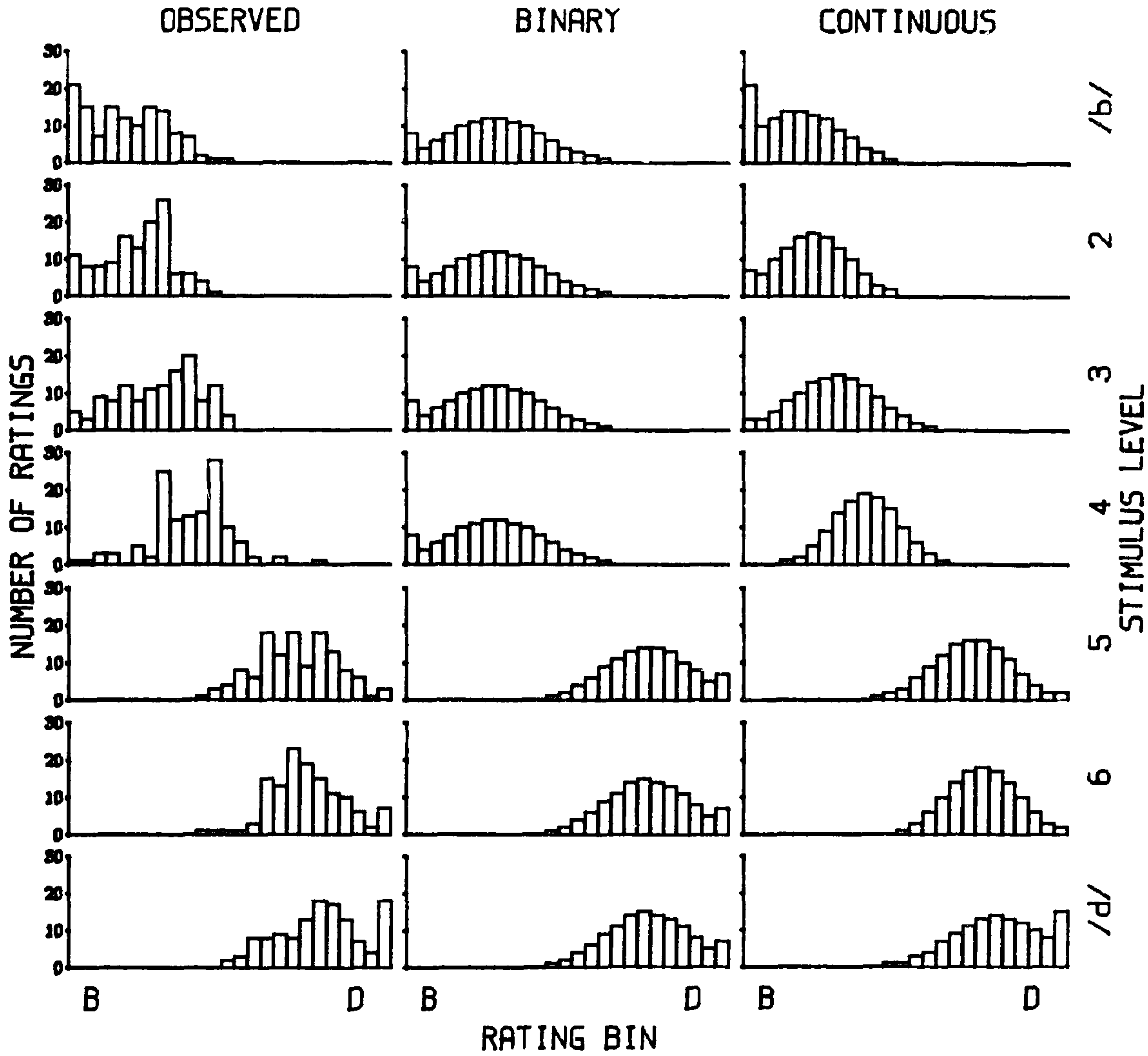

Fig. 7. Observed (left), predicted binary (center), and predicted continuous (right) distributions for a typical subject in the place condition.

These values gave the smallest chi-squares between the predicted and observed ratings, i.e. the best fit to the observed data. Figs. 7, 8, and 9 show the observed distributions along with the predictions made by the binary and continuous models for three subjects (place subject 10, voicing subject 9, and vowel subject 6, respectively). As can be seen, the continuous model does a better job of fitting the observed data for all three subjects. What is most noticeable in the figures is how much better the continuous model does for the intermediate stimulus level. It is much better, of course, since the observed data appear to result from a single, central distribution rather than from a mixture of two distant distributions. The figures also illustrate how much better the continuous model does for the other levels of the distributions. In Fig. 8, for example, we can see that the continuous model much more accurately captures and reproduces the different number of observations occurring in the tails of the end-point stimuli.

To compare the goodness of fit for the two models the F statistic, i.e. the ratio of the obtained chi-square values, was computed for the fits of the both models for each subject and for the total experimental group. The significance of the F ratio may be tested according to the degrees of freedom of the two chi-squares, which in the present application are given by the total number of bins used minus the number of parameters minus 1 [27, p. 294). The binary/continuous F ratios are given in Table 5. For the few cases where the binary model was significantly better, the probability associated with the inverse F ratio (with reversed df) is given followed by an asterisk. As can be seen, the continuous model fit the data significantly better for most of the subjects. In 3 cases the binary model fit better ($p < 0.05$), in 22 cases the continuous model did best and in 11 cases the differences between the models did not reach significance. For each of the three speech continua an F ratio was calculated from the chi-squares pooled across the

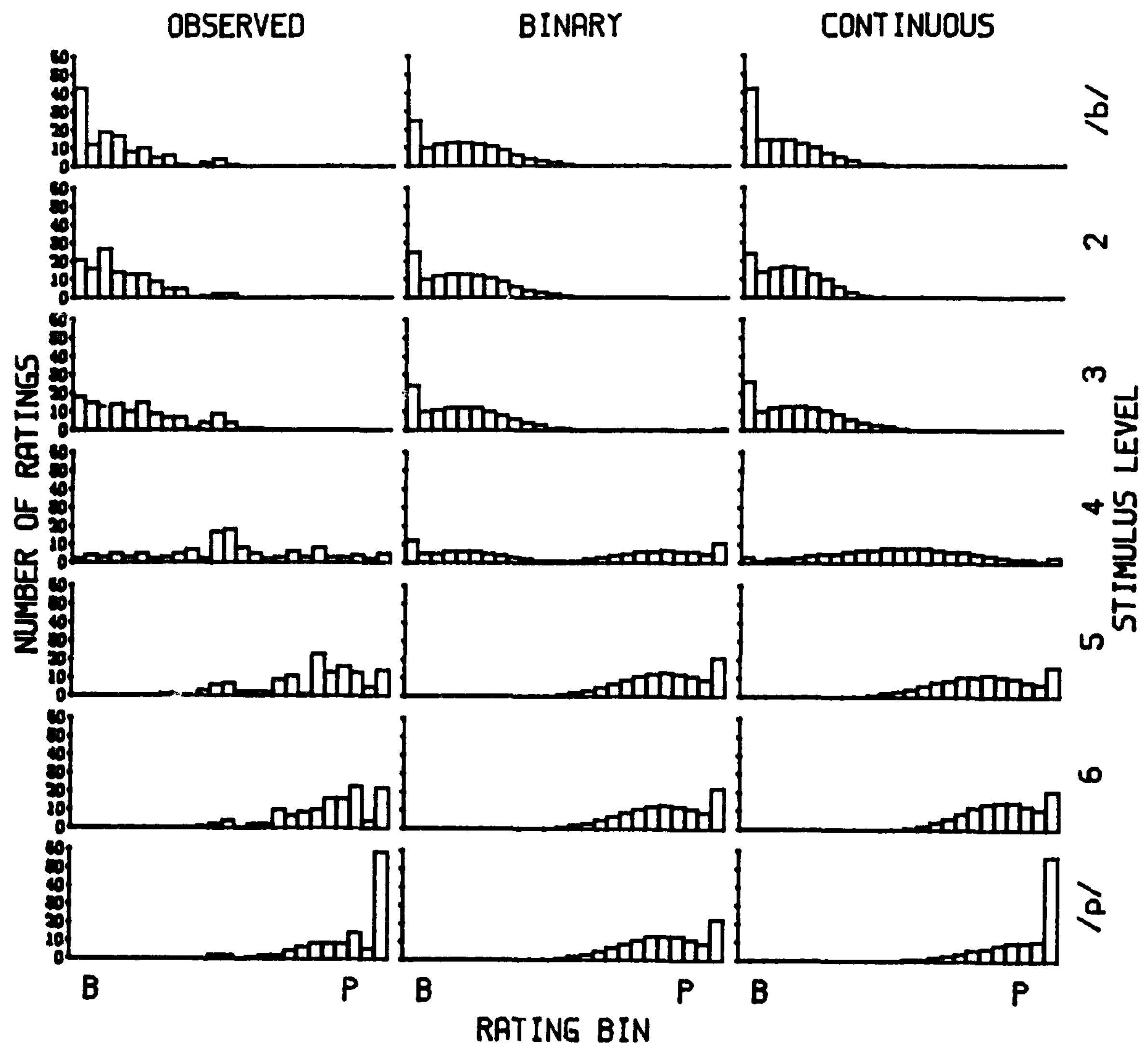

Fig. 8. Observed (left), predicted binary (center), and predicted continuous (right) distributions for a typical subject in the voicing condition.

12 subjects in each group. The continuous model fit better than the binary model in each case (*place*: $F(866,782) = 1.3923$, $p < 0.0355$, *voicing*: $F(900,773) = 1.6327$, $p < 0.0038$, *vowel*: $F(49,771) = 2.3611$, $p < 0.0001$).

In order to be scrupulously fair to theories of categorical speech perception, the data were also tested with a more elaborate categorical model which we may call the Guessing model. This model is the same as the binary model but with the additional feature that for each level of the stimulus there is a certain probability that the response is based on a neutral *Guessing* distribution. One way of viewing the guessing model would be in terms of a threshold or criterion for making a particular response. In this conceptualization, if the evidence for one or the other alternative does not exceed the criterion level, then a guessing response would be made. Naturally, it would be expected that the guessing response would occur most frequently for the more central, ambiguous levels of the stimulus dimension. This model has 20 parameters – 3 means and 3 standard deviations for the two alternative distributions and the guessing distribution, and 14 sampling probabilities. Each of the 7 stimulus levels needs only 2 probability parameters to compute the two alternative probabilities and the guessing probability since the 3 probabilities sum to 1. The starting values for the sampling probability parameters were initially set to 0.98 of the final values found for the binary model, the two percept distributions remained the same, and the guessing distribution (mean 0.5, *sd* 0.2) was initially sampled with a 0.02 probability uniformly across stimulus level.

Table 6 gives the Binary/Guessing F ratios computed after fitting the guessing model with STEPIT. As can be seen, the guessing model was not significantly better than the straight binary model. Only in 5 out of 36 cases was there any improvement in fit. In no case was the overall group fit significantly improved.

Table 5
Comparison of Binary and continuous model fits *

Place	1	$f(62, 60) = 1.0718$	
	2	$F(64, 51) = 0.7992$	
	3	$F(70, 59) = 1.8369$	$p = 0.0087$
	4	$F(88, 83) = 1.2153$	
	5	$F(77, 78) = 0.5100$	$p = 0.0017$ *
	6	$F(68, 62) = 1.8335$	$p = 0.0082$
	7	$F(62, 44) = 1.3421$	
	8	$F(87, 76) = 2.8730$	$p < 0.0001$
	9	$F(78, 72) = 1.5390$	$p = 0.0326$
	10	$F(72, 61) = 2.5593$	$p = 0.0001$
	11	$F(79, 75) = 1.8489$	$p = 0.0039$
	12	$F(59, 61) = 1.6062$	$p = 0.0324$
	total	$F(866,782) = 1.3923$	$p < 0.0355$
Voicing	1	$F(70, 64) = 0.6245$	$p = 0.0275$ *
	2	$F(67, 54) = 1.3120$	
	3	$F(53, 36) = 1.1705$	
	4	$F(84, 71) = 1.1027$	
	5	$F(79, 75) = 1.8059$	$p = 0.0053$
	6	$F(83, 64) = 1.8253$	$p < 0.0001$
	7	$F(76, 71) = 1.4696$	$p = 0.0513$
	8	$F(75, 73) = 0.6667$	$p = 0.0414$ *
	9	$F(71, 63) = 2.7709$	$p < 0.0001$
	10	$F(76, 65) = 2.0585$	$p = 0.0016$
	11	$F(90, 82) = 0.9146$	
	12	$F(76, 55) = 2.5600$	$p = 0.0002$
	total	$F(900,773) = 1.6327$	$p < 0.0038$
Vowel	1	$F(73, 62) = 2.6251$	$p = 0.0001$
	2	$F(81, 58) = 4.4677$	$p < 0.0001$
	3	$F(74, 62) = 4.3819$	$p < 0.0001$
	4	$F(90, 83) = 0.9006$	
	5	$F(72, 70) = 1.3142$	
	6	$F(85, 66) = 2.6485$	$p < 0.0001$
	7	$F(90, 75) = 2.5126$	$p < 0.0001$
	8	$F(72, 65) = 1.7784$	$p = 0.0097$
	9	$F(80, 68) = 2.7309$	$p < 0.0001$
	10	$F(74, 57) = 3.2230$	$p < 0.0001$
	11	$F(72, 53) = 1.4471$	
	12	$F(68, 52) = 3.6620$	$p < 0.0001$
	total	$F(949,771) = 2.3611$	$p < 0.0001$

* Chi-square of Binary model divided by chi-square of the Continuous model. The asterisk denotes cases in which the Binary model did significantly better than the Continuous model.

Table 6
Comparison of Binary and Guessing model fits *

Place	1	$F(62,52) = 1.0501$	
	2	$F(64,58) = 0.9557$	
	3	$F(70,51) = 1.2764$	
	4	$F(88,78) = 0.9467$	
	5	$F(77,66) = 1.0037$	
	6	$F(68,58) = 1.5582$	$p = 0.0424$
	7	$F(62,47) = 1.3573$	
	8	$F(87,77) = 0.8761$	
	9	$F(78,69) = 0.9238$	
	10	$F(72,59) = 1.7598$	$p = 0.0131$
	11	$F(79,74) = 1.0268$	
	12	$F(59,44) = 1.0641$	
	total	$F(866,733) = 1.1435$	
Voicing	1	$F(70,60) = 0.9576$	
	2	$F(67,53) = 1.3847$	
	3	$F(53,47) = 0.9258$	
	4	$F(84,60) = 0.9238$	
	5	$F(79,66) = 1.4856$	$p = 0.0490$
	6	$F(83,68) = 1.3217$	
	7	$F(76,69) = 0.2447$	
	8	$F(75,69) = 0.9147$	
	9	$F(71,63) = 0.9113$	
	10	$F(76,64) = 0.9856$	
	11	$F(90,79) = 1.0261$	
	12	$F(76,60) = 1.3686$	
	total	$F(900,768) = 1.1271$	
Vowel	1	$F(73,60) = 1.4614$	
	2	$F(81,71) = 0.9468$	
	3	$F(74,58) = 1.6056$	
	4	$F(90,79) = 1.3765$	
	5	$F(72,60) = 1.3765$	
	6	$F(85,66) = 1.5757$	$p = 0.0280$
	7	$F(90,81) = 0.9244$	
	8	$F(72,59) = 1.3246$	
	9	$F(80,66) = 1.2618$	
	10	$F(74,65) = 0.8900$	
	11	$F(72,58) = 0.8888$	
	12	$F(86,64) = 1.1730$	
	total	$F(949,787) = 1.1222$	

* Chi-square of Binary model divided by chi-square of the Guessing model.

Table 7 gives the Guessing/Continuous F ratios. In only two out of 36 cases was the guessing model significantly better than the continuous model. In 16 of the 36 cases the continuous model was significantly better and in the remaining cases the comparison did not reach significance. For each of the three groups of subjects, the continuous model provided a better fit of the data than the guessing model (*place*: $F(733,782) = 1.2177$, $p < 0.0443$, *voicing*: $F(768,773) = 1.4485$, $p <$

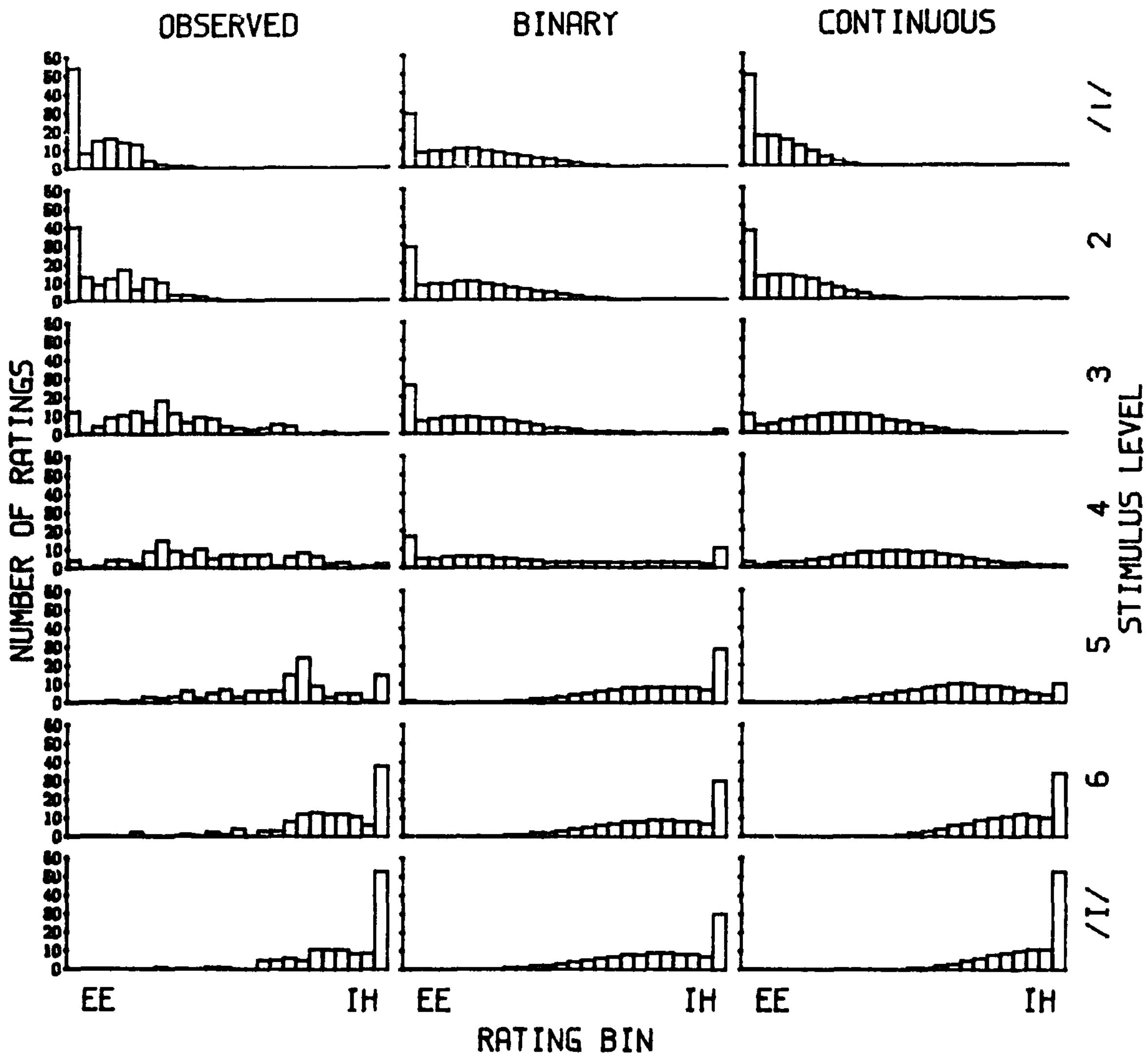

Fig. 9. Observed (left), predicted binary (center), and predicted continuous (right) distributions for a typical subject in the vowel condition.

0.0217, *vowel:* $F(787,771) = 2.1040$, $p < 0.0001$). With reference to the earlier discussion of Figs. 7, 8, and 9, it seems that it is not enough to account simply for the responses to the intermediate stimulus level; the guessing model remains inferior in representing the responses to the other levels of the sound continuum.

The fit of the continuous model model was somewhat better for the vowel than for the consonant series. It has been noted that a vowel continuum spans a greater perceptual range in terms of just noticeable differences than a stop consonant continuum [28]. Therefore, the difference might simply be due to the differences in the perceptual ranges for the two types of continua. It is possible that the difference might be eliminated or even reversed if the range of vowels was made smaller.

Discussion

The present study provides a new converging line of evidence supporting continuous, as opposed to strictly categorical perception of speech. In recent years several other studies have been carried out showing the importance of continuous information in speech processing [29]. Cohen [44] studied the combination of cues to voicing of medial velar stops. He varied preceding vowel duration, silent closure interval, and voice onset time and had subjects rate the sounds on a scale from /aga/ to /aka/. He found that the results were best described by a model which assumed the combination of three independent, continuous voicing cues. In similar experiments, Massaro and Cohen [30,31] varied frication duration, voice onset time and voice pitch, showing how continuous information is functional in the perception of /si/ and /zi/. In two further studies, Oden and Massaro [32] and Massaro and Oden [33] have demon-

Table 7
Comparison of Guessing and Continuous model fits *

Place	1	$F(52,60) = 1.0206$	
	2	$F(58,51) = 0.8362$	
	3	$F(51,59) = 1.4392$	
	4	$F(78,83) = 1.2837$	
	5	$F(66,78) = 0.5081$	$p = 0.0026$ *
	6	$F(58,72) = 1.1766$	
	7	$F(47,44) = 0.9888$	
	8	$F(77,76) = 3.2791$	$p < 0.0001$
	9	$F(69,72) = 1.6659$	$p = 0.0166$
	10	$F(59,61) = 1.4543$	
	11	$F(74,75) = 1.8006$	$p = 0.0060$
	12	$F(44,61) = 1.5094$	
	total	$F(733,782) = 1.2177$	$p < 0.0443$
Voicing	1	$F(60,64) = 0.6521$	$p = 0.0482$ *
	2	$F(53,54) = 0.9475$	
	3	$F(47,36) = 1.2644$	
	4	$F(70,71) = 1.1936$	
	5	$F(66,75) = 1.2156$	
	6	$F(68,64) = 2.8943$	$p < 0.0001$
	7	$F(69,71) = 1.1807$	
	8	$F(69,73) = 0.7288$	
	9	$F(63,63) = 3.0405$	$p < 0.0001$
	10	$F(64,65) = 2.08845$	$p = 0.0018$
	11	$F(79,82) = 0.8913$	
	12	$F(60,55) = 1.8705$	$p = 0.0099$
	total	$F(768,773) = 1.4485$	$p < 0.0217$
Vowel	1	$F(60,62) = 1.7964$	$p = 0.0120$
	2	$F(71,58) = 4.7187$	$p < 0.0001$
	3	$F(58,62) = 2.7291$	$p = 0.0001$
	4	$F(79,83) = 0.8681$	
	5	$F(60,70) = 0.9548$	
	6	$F(66,66) = 1.6809$	$p = 0.0180$
	7	$F(81,75) = 2.7180$	$p < 0.0001$
	8	$F(59,65) = 1.3426$	
	9	$F(66,68) = 2.1642$	$p = 0.0009$
	10	$F(65,57) = 3.6214$	$p < 0.0001$
	11	$F(58,53) = 1.6282$	$p = 0.0370$
	12	$F(64,52) = 3.0879$	$p = 0.0002$
	total	$F(787,771) = 2.1040$	$p < 0.0001$

* Chi-square of Guessing model divided by chi-square of the Continuous model. The asterisk denotes cases in which the Guessing model did significantly better than the continuous model.

strated how continuous feature information is important in the perception of initial stop consonant syllables.

The utilization of continuous acoustic information is also central to the integration of higher-order context in speech perception. As an example, Massaro and Cohen [45] have shown the importance of continuous information in the listener's utilization of certain phonological constraints. In this study, a speech continuum between the liquid-vowel syllables /ri/ and /li/ was generated by varying the onset frequency of the third formant. Each sound along the continuum was placed in a consonant-cluster vowel syllable after an initial consonant /p/, /t/, /s/, and /v/. In English, both /r/ and /l/ are phonologically admissible following word initial /p/ but are not admissible following /v/. Only /l/ is admissible following /s/ and only /r/ is admissible following/t/. The identifications of these sounds revealed significant effects of both the acoustic information and the phonological context. The results were described quantitatively with the assumption that the listener integrates *continuous* sources of information in speech perception. Similarly, Ganong [34] has demonstrated that continuous, rather than categorical information seems to be available at the stage of integration of lexical information in speech processing [cf. 29].

Categorical perception

Our experiments tested the strong form of categorical perception and is neutral with respect to the weak form. The latter states simply that there is an enhanced discriminability between stimuli across a phoneme category. The results of the present experiment indicate that continuous changes along some speech dimension produce relatively continuous and not discrete perceptual changes. The experiments do not address the issue of whether the changes along the perceptual dimension are proportional to those along the stimulus dimension. The average rating responses given in Fig. 6 show some hint of larger differences between successive stimuli in the middle relative to the extremes of the stimulus continuum. This result can not be taken as evidence for better discriminability across the phoneme boundary, however. Firstly, we do not have a measure of the phoneme boundary, and secondly, there is reason to believe that the observed mean ratings would be closer together at the extremes even if the discrimination was constant across the continuum. As was noted in the formulation and tests of the models, a potential rating response might fall outside the range of the rating scale and in practice the rating would be placed simply at the end of the rating scale. Since it is more likely that two adjacent stimuli will have some of their distributions off the scale when they are at the extremes of the continuum, the average rating responses to two extreme stimuli will be more similar to each other than the similarity of the responses to two stimuli in the middle of the stimulus continuum.

Our results indicate that the listener has continuous information about speech sounds. These results are consonant with a continuous perception interpretation of the early studies of Barclay [9,10] Miller [11] and McNabb [43]. Fujisaki and Kawashima [35,36]. Pisoni and Lazarus [16], and Pisoni and Tash [13]. More recently, Carney, Widlin, and Viemeister [15] and Samuel [17] have demonstrated continuous discrimination of voice onset time in stop consonants. In contrast to these demonstrations, other studies of speech perception have reached just the opposite conclusion [37,3].

How do we resolve the discrepancy between these two sets of studies? Our proposal is as follows: Although the listener has continuous information available, this information is not always transmitted to the experimenter. When the continuous information is not transmitted, the results take on the form of categorical perception. For example, Hary and Massaro [38] created a qualitative change along a nonspeech continuum in order to create a traditional categorical result with sounds that can be perceived continuously. Replicating the Rosen and Howell [39] study, they showed that a continuum of sawtooth stimuli whose rise times varied in a single direction were perceived continuously, not categorically. In contrast, a bipolar continuum of positive and negative rise times yielded traditional categorical results. Therefore, the finding of categorical results along the bipolar continuum, when some sounds were shown to be perceived continuously in another context, argues against the use of similar results as evidence for categorical perception. Rather than referring to these results as categorical perception, categorical communication or categorical responding might be a better descriptor. In order to resolve the discrepancies across experiments, we briefly discuss some important factors in experimental studies of categorical perception.

Abstract codes used in discrimination

The prototypical discrimination test used to assess categorical perception in the ABX task. Given the fragility of auditory memory, however, the observed phenomenon of categorical perception may, in fact, be categorical memory. Subjects in the ABX task may try to remember both the percept and the label assigned to the A and B sounds. When X is presented, they try to match the sound of X with the remembered sounds of A and B. Given that perceptual memory is so fragile in this task, however, they more often than not forget the sounds of A and B. In this case, the subjects must rely on the labels they assigned to the sounds A and B and choose the one that matches the label they assigned to X [35,36,40,41]. This strategy will produce the results usually attributed to categorical perception. Sounds assigned different labels will be more likely to be discriminated in the ABX task simply because the subject depends on the name of the sounds rather than auditory memory for the sounds themselves. Accordingly, there is nothing categorical about the *perception* of speech sounds when the role of memory is accounted for in the ABX task.

In agreement with this analysis, Pisoni and Lazarus [16] found categorical perception of stop consonants with the ABX task, even if observers are trained in listening to the stimuli in sequential order along the continuum. Subjects given similar training did not show categorical perception when tested with a more sensitive discrimination task. The discrimination test involved the presentation of two pairs of sounds; one pair was the same and one pair was different. This task reduces the load on auditory memory and encourages a direct auditory comparison between the two members of the a pair of sounds.

Categorical perception might also be found when a floor effect exists in the discrimination task. For example, a subject might have continuous information available in the discrimination task but this information might be relatively uninformative. This would produce poor discrimination for a given speech contrast and continuous perception would not be found. Given poor continuous information, subjects might fall back on their abstract categorization of the sounds. If this strategy is followed, the results would follow the predictions of categorical perception.

When categorical perception results are found, a burden of proof still remains with the investigator. Some converging evidence should be provided to demonstrate that subjects were utilizing auditory memory in the discrimination task. If this evidence is not provided, the results might be interpreted as simply due to utilization of abstract codes rather than auditory memory in the discrimination task. Therefore, the results can not be taken as unambiguous support for categorical perception.

Binary stimulus dimensions

Categorical perception results may also be found when a single acoustic continuum gives the listener two perceptual dimensions: one that is continuous, and a second that is binary and easily categorized as one of two alternatives. This situation appears to have been present in an experiment carried out by Pastore, Ahroon, Baffuto, Friedman, Puleo, and Fink [42]. Each stimulus was a positive or negative amplitude change in a continuous tone. Subjects were tested in both an ABX task and a standard labeling identification task. The results replicated the prototypical results of categorical perception. The labeling boundary was very sharp, with a corresponding peak in the discrimination function; the labeling results were capable of predicting the discrimination results using the classical categorical perception formula [3].

However, the Pastore et al. [42] results simply reflect the reasonable observation that listeners can discriminate the direction of a loudness change without necessarily discriminating the exact magnitude of the change. Therefore, a listener can hear a positive or negative change but still be unclear about whether it was a 2- or 4-dB change. Given reliable information about the discrete (two-valued) dimension, and relatively poor information about the continuous dimension, we can expect a discontinuity at the point of change along the discrete dimension. Therefore, categorical perception results will be observed [29]. Although there may be certain binary stimulus dimensions in natural speech, we believe that most of the contrasts in speech include continuous dimensions. For the place continuum between /bæ/ and /dæ/ however, the direction of the $F2$–$F3$ transitions might be considered to be a binary stimulus dimension. The $F2$–$F3$ transitions are rising at the /bæ/ end of the continuum and falling at the /dæ/ end. Therefore, this continuum might be characterized by both this binary stimulus dimension and the continuous dimension of the magnitude of the formant transitions. The present results show that continuous information is available along the dimensions of formant transitions, voice onset time, and vowel formant continua.

It might still be argued that it is necessary to demonstrate categorical results in the traditional labeling-discrimination tasks using the stimuli and subjects from the present experiment. However, our point is that the traditional task is not diagnostic of categorical versus continuous perception. A match or mismatch between labeling and discrimination can result from either categorical or continuous perception. Perception may be continuous but one might find a match between the two tasks simply because subjects use abstract labels rather than auditory differences in the discrimination task. That is, subjects simply respond on the basis of whether the sounds were classified as one alternative or the other rather than on the basis of the relative degree of match to each of the two alternatives. Similarly, perception might be discrete or categorical even though one obtains a mismatch between labeling and discrimination performance. Healy and Repp [7], for example, illustrate how context effects in the discrimination task can contribute to a mismatch between the labeling and discrimination results. Accordingly, we rest our case on the evaluation of discrete and continuous models of perception of common speech continua.

Concluding comment

An important question in speech perception has been to what level is continuous information preserved? No speech theorist would argue against the continuous representation of acoustic information on the basilar membrane. Early categorical theories suggested that this continuous information does not get much further. Later theories [e.g. 35,36,41] suggested that both acoustic, i.e. continuous, and phonetic, i.e. categorical, processing occurs, but that the acoustic information is available only for a short time from a rapidly decaying auditory memory. The results of the present study indicate that some usable continuous information is available for a perceptual judgment. Several questions remain about continuous speech information. First, what is the mechanism which makes available the continuous information? Second, how is this information evaluated and integrated by the listener? And finally, to what extent is this continuous information utilized in the natural perception of continuous speech.

Acknowledgement

The preparation of this paper was supported in part by NIMH grants MH-19399 and MH-35334. We would like to thank Neil Macmillan, Hector Javkin and Ian Maddieson for commenting on an earlier version of this paper.

References

[1] J.R. Platt, "Strong inference", *Science*, Vol. 146, 1964, pp. 347-353.
[2] K.R. Popper, *The Logic of Scientific Discovery*, Huchinson, London, 1959. Originally published 1934.
[3] M. Studdert-Kennedy, A.M. Liberman, K.S. Harris, and F.S. Cooper, "The motor theory of speech perception: A reply to Lane's critical review", *Psych. Rev.*, Vol. 77, 1970, pp. 234-249.
[4] M. Studdert-Kennedy, "Speech perception", in: N.J. Lass, ed., *Contemporary issues in Experimental Phonetics*, Academic Press, New York, 1976, pp. 243-293.
[5] M.E.H. Schouten, "The case against a speech mode of perception", *Acta Psychologica*, Vol. 44, 1980, pp. 71-98.
[6] A.M. Liberman, K.S. Harris, H.S. Hoffman, and B.C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries", *J. Exp. Psych.*, Vol. 54, 1957, pp. 358-368.
[7] A.F. Healy, and B.H. Repp, "Context independence and phenetic mediation in categorical perception", *J. Exp. Psych.: Hum. Percep. Perf.*, Vol. 8, 1982, pp. 68-80.
[8] N.A. Macmillan, H.L. Kaplan, and C.D. Creelman, "The psychophysics of categorical perception", *Psych. Rev.*, Vol. 84, 1977, pp. 452-471.
[9] J.R. Barclay, "Non-categorical perception of a voiced stop consonant", *Proc. 78th Annual Convention Amer. Psych. Assoc.* 1970, pp. 9-10.
[10] J.R. Barclay, "Non-categorical perception of a voiced stop: A replication", *Percep. Psychophys.*, Vol. 11, 1972, pp. 269-273.
[11] J.L. Miller, "Properties of feature detectors for VOT: The voiceless channel of analysis", *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 641-648.
[12] M. Studdert-Kennedy, A.M. Liberman, and K.N. Stevens, "Reaction times to synthetic stops and vowels at phoneme centers and at phoneme boundaries", *J. Acoust. Soc. Am.*, Vol. 35, 1963, pp. 1900.
[13] D.B. Pisoni, and J. Tash, "Reaction times to comparisons within and across phonetic categories", *Percep. Psychophys.*, Vol. 15, 1974, pp. 285-290.
[14] B.H. Repp, "Perceptual equivalence of two kinds of ambiguous speech stimuli", *Bull. Psychonom. Soc.*, Vol. 18, 1981, pp. 12-14.
[15] A.E. Carney, G.P. Widlin, and N.F. Viemeister, "Noncategorical perception of stop consonants differing in VOT", *J. of the Acoust. Soc. Am.*, Vol. 62, 1977, pp. 961-970.
[16] D.B. Pisoni, and J.H. Lazarus, "Categorical and noncategorical modes of speech perception along the voicing continuum", *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 328-334.
[17] A.G. Samuel "The effect of discrimination training on speech perception: Non-categorical perception", *Percep. Psychophys.*, Vol. 22, 1977, pp. 321-330.
[18] C.S. Watson, M.E. Rilling, and W.T. Bourbon, "Receiver-operating characteristics determined by a mechanical analog to the rating scale", *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 283-288.
[19] J.L. Elman, "Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis", *J. Acoust. Soc. Am.*, Vol. 65, 1979, pp. 190-207.
[20] J.R. Sawusch, "Peripheral and central processes in selective adaptation of place of articulation in stop consonants", *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 738-750.
[21] D.W. Massaro, "The role of the decision system in sensory and memory experiments using confidence judgments", *Percep. Psychophys.*, Vol. 5, 1969, pp. 270-272.
[22] W.A. Wickelgren, "Testing two state theories with operating characteristics and a priori probabilities", *Psych. Bull.*, Vol. 69, 1968, pp. 126-131.
[23] L.L. Thurstone, "A law of comparative judgment", *Psych. Rev.*, Vol. 34, 1927, pp. 273-286.
[24] W.S. Torgerson, *Theory and Methods of Scaling*, Wiley, New York, 1958.

[25] M.M. Cohen, and D.W. Massaro, "Real-time speech synthesis", *Beh. Res. Meth. Inst.*, Vol. 8, 1976, pp. 189–96.

[26] J.P. Chandler, "Subroutine STEPIT – Finds local minima of a smooth function of several parameters", *Beh. Sci.*, Vol. 14, 1969, pp. 81–82.

[27] I. Guttman, S.S. Wilks, and J.S. Hunter, *Introductory Engineering Statistics,* Wiley, New York, 1971.

[28] A.E. Ades, "Vowels, consonants, speech, and nonspeech", *Psych. Rev.,* Vol. 84, 1977, pp. 524–530.

[29] D.W. Massaro, and G.C. Oden, "Speech perception: A framework for research and theory", in: N.J. Lass, ed. *Speech and Language: Advances in Basic Research and Practice*, Vol. 3, Academic Press, New York, 1980, pp. 129–165.

[30] D.W. Massaro, and M.M. Cohen, "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction", *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 704–717.

[31] D.W. Massaro, and M.M. Cohen, "The contribution of voice-onset time and fundamental frequency as cues to the /zi/-/si/ distinction", *Percep. Psychophys.*, Vol. 22, 1977, pp. 373–382.

[32] G.C. Oden, and D.W. Massaro, "Integration of featural information in speech perception", *Psych. Rev.*, Vol. 85, 1978, pp. 172–191.

[33] D.W. Massaro, and G.C. Oden, "Evaluation and integration of acoustic features in speech perception", *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 996–1013.

[34] W.F. Ganong, "Phonetic categorization in auditory word perception", *J. Exp. Psych.: Hum. Percep. Perf.*, Vol. 6, 1980, pp. 110–125.

[35] H. Fujisaki, and T. Kawashima, "On the modes and mechanisms of speech perception", *Annual Report Engin. Res. Inst.*, Faculty of Engineering, Univ. of Tokyo, Tokyo, Vol. 28, 1969, pp. 67–73.

[36] H. Fujisaki, and T. Kawashima, "Some experiments on speech perception and a model for the perceptual mechanism", *Annual Report Engin. Res. Inst.*, Faculty of Engineering, Univ. of Tokyo, Tokyo, Vol. 29, 1970, pp. 206–214.

[37] K. Miyawaki, W. Strange, R. Verbrugge, A.M. Liberman, J.J. Jenkins, and O. Fujimura, "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English", *Percep. Psychophys.*, Vol. 18, 1975, pp. 331–340.

[38] J.M. Hary and D.W. Massaro, "Categorical results do not imply categorical perception", *Percept. Psychophys.*, Vol. 32, 1982, pp. 409–418.

[39] S.M. Rosen and P. Howell, "Plucks and bows are not categorically perceived", *Percep. Psychophys.*, Vol. 30, 1981, pp. 156–168.

[40] K.R. Paap, "Theories of speech perception", in: D.W Massaro, ed., *Understanding Language,* Academic Press, New York, 1975, pp. 151–204.

[41] D.B. Pisoni, "Auditory and phonetic codes in the discrimination of consonants and vowels", *Percep. Psychophys.*, Vol. 13, 1973, pp. 253–260.

[42] R.E. Pastore, W.A. Ahroon, K.J. Baffuto, C. Friedman, J.S. Puleo, and E.A. Fink, "Common-factor model of categorical perception", *J. Exp. Psych.: Hum. Percep. Perf.*, Vol. 3, 1977, pp. 686–696.

[43] S.D. McNabb, "Must the output of the phonetic feature detector be binary?", *Res. Speech Perception Progress Report*, Vol. 2, Indiana University, Department of Psychology, 1974.

[44] M.M. Cohen, "Three cues to the voicing of intervocalic velar stops", MS Thesis 1979, University of Wisconsin-Madison.

[45] D.W. Massaro and M.M. Cohen, "Phonological constraints in speech perception", Talk presented at the Acoust. Soc. Am. 99th Meeting, Atlanta, GA, April 1980.