

Natural auditory sensitivities as universal determiners of phonemic contrasts¹

PETER HOWELL and STUART ROSEN

1. Introduction

Are there acoustic invariants that mediate the perception of phonemes? If there were, they might be shared by a number of languages. But this acoustic-phonemic relationship does not occur in a simple way for any of the languages that have been examined. Has, then, such an approach to speech perception been abandoned? Not entirely; some theorists argue that there are invariant acoustic properties which occur whatever context a phoneme is spoken in. On the other hand, some have considered that the speech signal is inherently variable and the invariant relations only occur at higher levels of processing (Liberman et al. 1967; Liberman and Studdert-Kennedy 1978).

Stevens has been the main protagonist of the view that the signal contains invariant acoustic correlates of phonemic categories (Stevens 1975, 1981; Stevens and Klatt 1976; Halle and Stevens 1980). He has considered all the problems that need to be accounted for by a theory which holds that acoustic invariants exist in the signal — how these invariant relations arise in speech production, what factors make it difficult to isolate them from acoustic records, and what evidence there is that these acoustic relations are used for the purpose of perceiving phonemic categories. Other workers have taken a similar theoretical line in explaining phonemic perception. For example, Kuhl and Miller (1975) have reported that chinchillas label synthetic speech in a remarkably similar way to English-speaking humans. This finding has been interpreted as showing that the animals are responding to acoustic factors in these sounds which underlie the phoneme categories common to the languages of the world. In this chapter we will assess whether invariant auditory factors can mediate perception of phonemic categories within and across the world's languages. The invariant acoustic factors associated with phonemes are supposed to give rise to the phenomena of categorical perception whether it occurs in human or other mammalian

species. We commence the assessment by describing what categorical perception is and how it was originally explained.

2. Categorical perception: the basic phenomena and theoretical positions

Investigations of speech perception originally commenced with an examination of the articulation of speech sounds to ascertain what articulatory movements determine them. The physical correlates of these cues are then sought. Finally, perceptual tests are performed using synthesis techniques. As an illustration we will consider the English voicing distinction in stop consonants (bilabial /b/ versus /p/, alveolar /d/ versus /t/, or velar /g/ versus /k/). There are several important articulatory differences which mark this distinction, all associated with the onset of laryngeal vibration relative to the release gesture for the stop consonant (Lisker and Abramson 1970; Klatt 1975). There are two main acoustic correlates associated with these articulatory events. If the initial portion of the sound is excited aperiodically and periodic excitation does not start until well after the upper vocal tract resonances (the upper formants) have appeared, the sound is heard as a voiceless stop (e.g. /k/). If the sound is excited periodically right from its start, it is heard as a voiced stop (e.g. /g/). The time at which the low-frequency resonances of the vocal tract appear relative to when the higher resonances appear differs between the voiced and voiceless stop consonants too. (The formants are numbered in ascending order of frequency and are designated F1, F2, F3, etc.) F1 starts earlier in voiced stop consonants than in voiceless stop consonants. These cues are illustrated in the set of schematic spectrograms shown in Figure 1 used for synthesis by Simon and Fourcin (1978) for the distinction 'coat' versus 'goat'. The width of the formant track is proportional to the formant amplitude. Voiced sections are colored black, while voiceless sections are white. Fx refers to the fundamental frequency of voiced excitation. Once these exemplars are constructed, it is simple to generate a continuum of sounds varying from one extreme to the other.

For example, the relative onset time of periodic vibration as a cue to the voiced-voiceless distinction has been intensively examined. With a speech synthesizer the onset of periodic excitation can be controlled directly and a continuum varying along this dimension constructed. Such a continuum is called a voice-onset time (VOT) continuum. VOT is usually given in time units representing the onset of periodic vibration relative to the appearance of the upper formants. Thus, 0 ms VOT means all formants are excited periodically at the start of the sound. A VOT of +10 ms

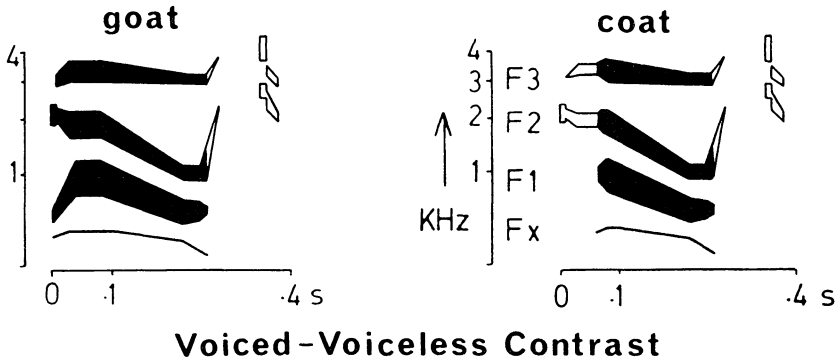


Figure 1. Schematic spectrograms of *goat* (left) and *coat* (right); these words are distinguished by the voice-onset time of the initial stop consonant (the way information is represented in these spectrograms is described in the text) (Simon and Fourcin 1978)

means periodic vibration starts 10 ms after the upper formants appear, which are excited in the first 10 ms, now, by an aperiodic source. If periodic excitation starts 10 ms before the upper formants appear, VOT would be -10 ms. A typical continuum which might be used to examine how the voiced-voiceless distinction is perceived might consist of sounds varying from 0 to $+80$ ms VOT in 10 ms steps. If a listener is asked to categorize a randomization of sounds drawn from this continuum as voiced or voiceless, he classifies sounds with short VOTs as voiced and sounds with longer VOTs as voiceless. One important aspect of perception on this type of continuum is the abrupt way that the categorizations change across the continuum from one category to another. These categorization curves are supposed to be sharper than those which occur with nonspeech (Liberman et al. 1967). The point on the categorization function where each category is heard equally often is called the phoneme boundary (in speech) or the category boundary (in nonspeech).

Discrimination of items along the VOT continuum was examined in the original studies by using an ABX discrimination task. Each trial consists of a triplet of sounds presented in sequence (designated A, B, and X respectively). A and B stimuli are always different and always have the same physical difference — for example, adjacent sounds on evenly spaced continua. The third (X) sound is the same as either the first (A) or the second (B) sound, and the subject has to indicate which he thinks the third sound is like. Performance with two sounds drawn from the same category is almost at chance but performance with stimuli straddling the phoneme boundary is considerably better. In other words, there is a discrimination peak at the phoneme boundary. This type of perception

has been found with several phonemic contrasts, most strongly stop consonants. We have already noted that the identification functions of speech continua are supposed to be sharper than those found in nonspeech continua common in psychophysical investigations. Nonspeech continua did not seem to show such discrimination peaks. (See Miller 1956 for a contemporary review of identification and discrimination experiments with nonspeech. Bear in mind that many of the studies discussed use different tasks to those employed in speech studies.) The unusual type of perception originally observed in certain speech continua was termed *categorical perception*.

Three characteristics of categorically perceived continua have been noted, which are summarized in Figure 2. These are a sharp identification function, chance discrimination within a phoneme category, and a peak in discrimination performance at the phoneme boundary.

The original interpretation of categorical perception posited that the listener employed phonemic labels during perception. Discrimination was poor for two stimuli in the same phoneme class because then the listener would give the first two sounds of an ABX triplet the same label and would have to guess which the third sound was most like. If the A and B sounds were labeled differently (as happens when a pair straddle the phoneme boundary) the listener only has to ascertain what the label of the third sound is to perform the discrimination correctly (Liberman et al. 1967). Thus all the features of categorical perception are predicted by this account. Since the use of phonemic labels was proposed, categorical perception was considered specific to speech.

It is important to note that this theory invokes a specific relationship between identification and discrimination. A second point to note is that no such relationship between identification and discrimination was thought to occur for any nonspeech continuum. Unlike the case in speech, where discrimination seemed to be limited by identification, discrimination in nonspeech was normally thought to be far better than identification.

Difference in 'sharpness' alone of identification functions among different continua cannot be used to argue that some of the continua are categorically perceived. There are no unambiguous criteria for measuring the relative 'sharpness' of identification since the stimulus values for different acoustic dimensions cannot be directly compared.

Identification data are ambiguous in another way when it comes to testing the claim that speech is special. The methodology used to assess performance with speech continua is quite different to that traditionally used in psychophysical studies. Identification experiments employing nonspeech continua are not usually performed with a small number of

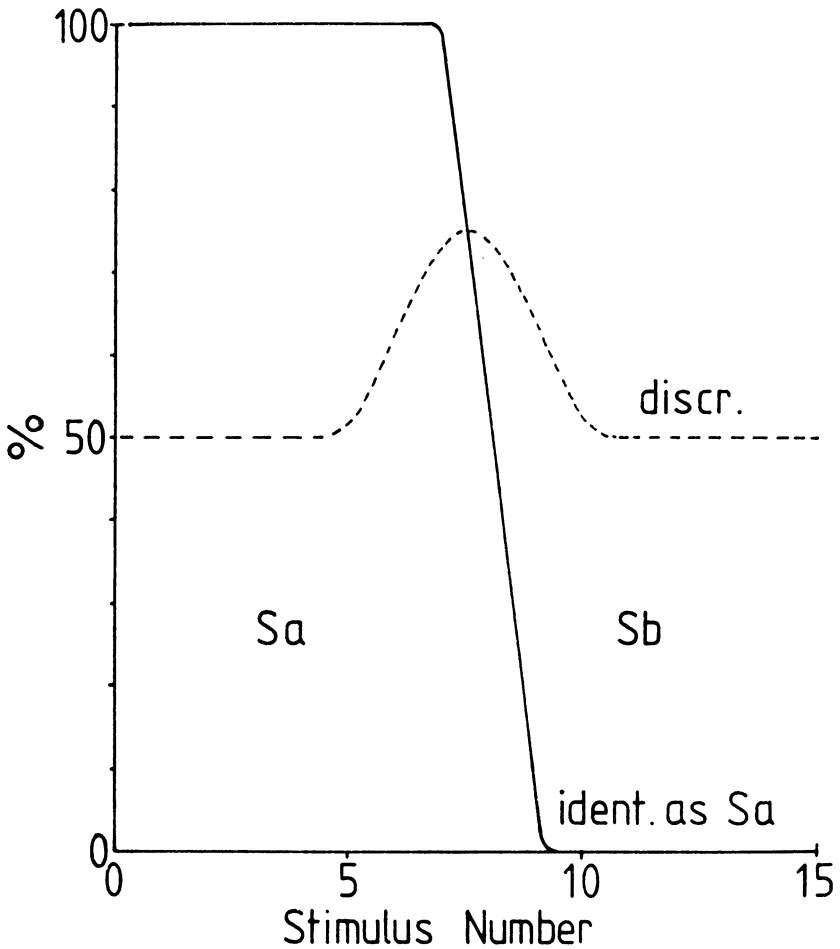


Figure 2. Illustration of the ideal identification and discrimination results expected for a continuum that is categorically perceived. For the identification function, the ordinate value gives the percent of times one of the response labels is used for each stimulus. For the discrimination curve, the values give percent correct

discrete labels (like the two used in the illustration we have given for the voiced-voiceless opposition) unless, of course, the experimenter is interested in ascertaining whether categorical perception occurs with such psychophysical continua. This difference in methodology causes the identification functions to vary in their steepness, as is shown in Figure 3 (Rosen 1979). Subjects were asked to use a six-point scale in the judgment of a set of stimuli from a voiced-stop place continuum /ba/-/da/. Three of

the responses were for varying goodness of 'ba' and three for 'da' (an excellent, a good, a poor). When the responses were analyzed as binary choices (either 'ba' or 'da'), as is typically the case in speech studies, the sharp categorization function evident at the top of the figure was obtained. When a number from 1 to 6 was applied to each of the response labels and a mean rating calculated for each of the stimuli, as is typically the case in nonspeech studies, the rather shallower function at the bottom of the figure was obtained. Note that this is only a difference in data analysis procedures and that both curves are results of the same experiment. Such examples clearly indicate that at least some of the difference in slopes of identification functions between speech and nonspeech is due to differences in procedure. This also illustrates the point that if you only give a subject two labels to use in such a task, stimuli that scale equivalently are not necessarily perceived equivalently. The use of more responses reveals that subjects are sensitive to differences within phonemic categories that they are supposedly not able to perceive. This effect also explains why actual discrimination performance in speech experiments is usually better within categories than is predicted from identification functions, since the identification functions derived from binary-choice experiments underestimate perceptual differences, particularly in within-category comparisons. This is clearly seen in Figures 5 and 6 of Liberman et al. (1957), where scatter plots of obtained versus predicted discrimination values are given. For relatively good performance (between-category judgments) obtained scores are close to those predicted, whereas for relatively poor performance (within-category judgments), obtained scores are much better than predicted.

Thus identification data alone provide little information about whether a particular continuum is categorically perceived or not. The literature contains many misconceived studies which assume that a sharp identification function alone tells us something about the categorical perception of the continuum. Another caution: if a phonemic contrast has been declared to be perceived categorically when signaled by one cue, it does not follow that the contrast is perceived categorically for other cues. Some preliminary investigations of our own have shown that stop consonants are not perceived categorically if they are cued by bursts rather than formant transitions.

Since, in the early experiments, the unusual discrimination performance was linked to the phonemic labels, categorical perception was thought to be specific to speech. Views changed in the middle 1970s when it was reported that categorical perception occurs with certain nonspeech continua more complex than those used in the classic psychophysical studies. At around the same time, it was reported that chinchillas showed similar

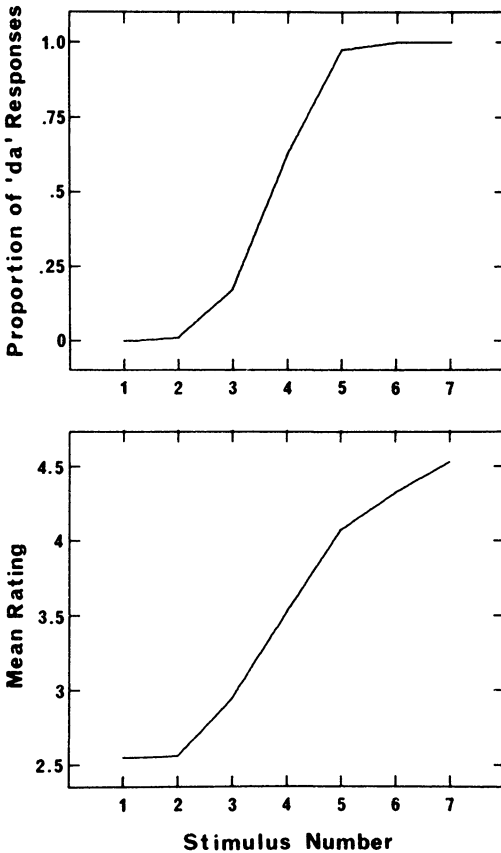


Figure 3. Results from an identification experiment on a $|ba|$ to $|da|$ continuum (Rosen 1979). In the lower half of the figure, mean ratings were calculated for each stimulus on the basis of a six-point rating scale used by the subjects. In the upper half of the figure, these same responses were analyzed as binary choices. Although the same data are employed in both analyses, the categorization function derived from the binary responses is considerably steeper than that based on the mean ratings

identification functions to humans on a speech continuum. The former finding indicates that categorical perception is not peculiar to speech and the latter finding suggests that even nonhumans may show categorical perception (though this had to be qualified since, at that time, no discrimination data were available to see whether a peak in the discrimination function occurred for this species).

These findings led to an alternative interpretation of categorical perception in which it was proposed that we have sensitivities to some

dimensions that auditory signals, including speech, vary on. To put it another way, the auditory system does not respond to equal physical differences equivalently. Changes at some points on physical continua are barely noticeable, while other equally large differences at different points on the same continuum are clearly distinct. Stevens (1981) has argued that speech has evolved to take advantage of these perceptual discontinuities to distinguish one phoneme from another. Peaks in the discrimination function of nonspeech continua are important for the clue that they may give to where discontinuities in the auditory system exist which may be used to advantage in the discrimination of phonemic classes. Such discontinuities might also serve to divide signals ranged along such continua into discrete categories and explain the 'sharp' identification functions on categorically perceived continua. This account is one version of what is called a 'natural-sensitivities' account. Though such an account could explain identification performance on categorically perceived continua, it is the discrimination results which are most important because they truly reflect perceptual discontinuities, while, as mentioned above, sharp identification functions alone do not definitely establish these discontinuities.

3. Evidence from speech continua and adult listeners

The evidence that has been taken as support for a natural-sensitivities account of speech perception is difficult, if not impossible, to account for by a theory that holds categorical perception to be a feature of a special speech-processing mechanism in the human brain. Those theorists who continue to maintain that speech is special have offered other evidence to show that it is necessary to propose a specialized speech-processing mechanism. They have, for example, pointed out that the auditory signal corresponding to a particular phoneme is drastically restructured when the phoneme is spoken in different contexts (e.g. stops before different vowels), at different rates, or by different speakers (see Howell and Harvey 1983 for a review). Different acoustic cues can be used for equivalent phonemic purposes, so none are invariantly related to particular phonemes. The remainder of this paper will examine whether the existence of auditory invariants which are used for perception can be maintained. In this section, data from human adults judging speech continua are examined. Stevens (1980) has presented data from acoustic analyses of stop consonants to show that there are invariant acoustic factors involved in the perception of such phonemes. The sensitivities that underlie perception are considered fixed; therefore, estimates of the parameters

that are thought to reflect these sensitivities should be fixed too. Operationally, then, the discrimination peaks are the particular things that we expect to be fixed across experimental tests because the natural-sensitivities account refers specifically to these peaks. Thus, to assess it best, we really need discrimination data. Often only identification data are available. Stevens (1981) leaves open the question of whether natural sensitivities should be evidenced in identification tasks. However, a theory that only applies to discrimination would not be much use, because identification is more like what listeners are called on to do in their everyday life. So we feel justified in assessing the account with respect to identification data.

A second problem is that the cues varied in experiments may not be the invariant ones but other cues the listener has learned are associated with the phonemic contrast (e.g. if it is considered that the onset of periodic vibration is the invariant underlying VOT perception, the onsets of the low-frequency vocal resonances may be learned as secondary cues, so experiments examining the latter would not allow an assessment of the natural-sensitivities account).

We will attempt an interim assessment of whether phoneme boundaries are fixed based mainly on identification data. If they are fixed, then there may be an invariant acoustic factor that underlies the perception of the contrast. If they are not, then, though an invariant may be present, it does not determine perception. We will consider the 'secondary-cues' argument further below.

A. Variation in the VOT phoneme boundary across languages

Boundary estimates for the voiced-voiceless distinction from three languages are summarised in Table 1. It is clear from inspection of this table that the boundaries exhibit considerable variation, indicating that an invariant acoustic factor is not used in the perception of this contrast.

Table 1. *Phoneme boundaries for VOT continua for three languages, presented for three places of articulation (bilabial, alveolar and velar); boundary estimates are presented in milliseconds (data taken from Lisker and Abramson 1971)*

	Bilabial	Alveolar	Velar
English	+25	+35	+42
Thai	+40	+46	+44
Spanish	+14	+22	+24

A similar point can be made within subjects in the case of bilinguals. Spanish–English bilinguals are an especially convenient group for study as the boundary for the voicing distinction occurs for quite different values of VOT in the two languages. Elman et al. (1977) have shown that language set affects the position of the phoneme boundaries for such bilinguals in an identification task. These cognitive influences on identification performance are difficult to account for if natural auditory sensitivities are meant to be responsible for phonemic distinctions.

B. Variation of VOT phoneme boundary with languages — rate effects

As speakers speak faster they produce more phonemes in a given time. The length of each phoneme decreases and the acoustic characteristics of the phonemes change too. Do listeners compensate for these changes or is the signal perceived invariantly? If the latter occurs then boundary estimates should fall at the same point whatever rate the speech is spoken at. Test syllables can be inserted into speech which is spoken at different rates and this question examined. When such tests are performed, listeners compensate for the perceived difference in rate as indicated by a change in the estimated phoneme boundary (Summerfield 1981; Diehl et al. 1980).

C. Range effects

It is well known from traditional psychophysics that the judgment of a particular stimulus is influenced by the other stimuli presented. For example, consider a loudness-rating task where subjects are asked to classify stimuli as ‘loud’ or ‘soft’. A stimulus that is judged ‘loud’ in the context of less intense stimuli will be judged ‘soft’ in the context of more intense stimuli. This is known as a range effect since it is the overall range of stimuli presented which causes the shifts in judgment. Other effects of context can be demonstrated, though they will not concern us here (Poulton 1977; Parducci 1965). The critical finding of these studies is that judgments are always relative.

If, however, sharp categorization of speech sounds is due to a natural sensitivity on the continuum, the categorization function should not alter with changes in the range of stimuli presented. It has in fact been asserted by Strange and Broen (1980), among others, that range effects do not occur for speech contrasts, especially not for stop consonants. Figure 4 shows a range effect for a place continuum (/ba-/da/) cued by second-

and third-formant transitions reported by Rosen (1979). Stimulus number 1 represented a good /ba/ and stimulus number 8 a good /da/. For the categorization function on the left, only stimuli 1 through 5 were presented, whereas for the function on the right, stimuli 4 through 8 were used. The phoneme boundary shifts from somewhat less than stimulus 4 to nearly stimulus 5. Stimulus 5, a good da in the lower subrange (94% da judgments) becomes, in the upper subrange, a highly ambiguous sound (53% da judgments). Similar effects of range have also been shown for an alveolar (/d/-/t/) voicing distinction by Brady and Darwin (1978). They obtained a maximum shift of about 10 ms for the phoneme boundary in their two extreme conditions.

Admittedly, these shifts are small though easily reproducible. They are only a problem to one version of the auditory-sensitivities hypothesis, which posits a stable reference as responsible for the various aspects of categorical perception (Pastore et al. 1977). Stevens's (1981) notion of a

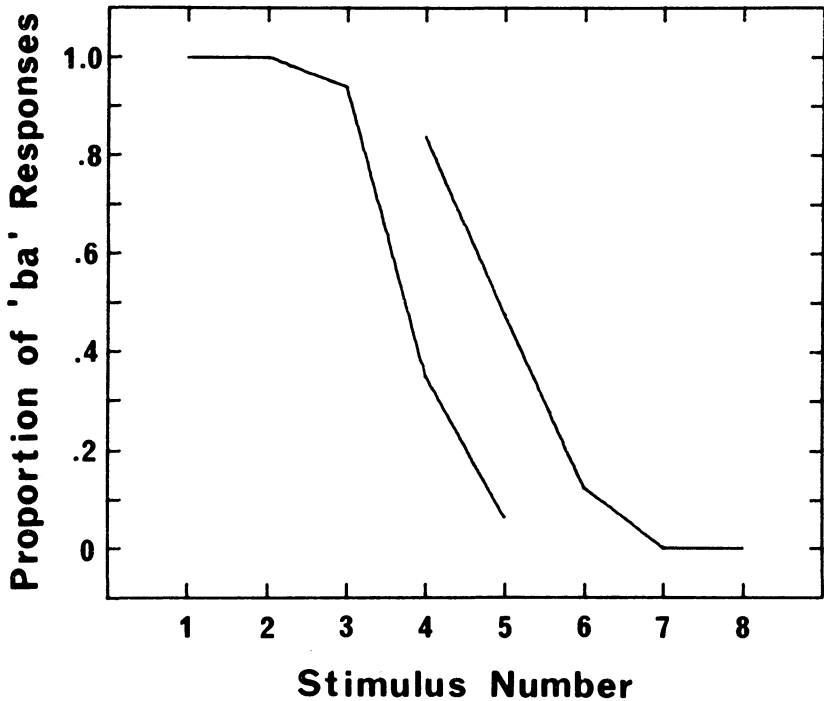


Figure 4. Identification functions for two different subranges of a |ba| to |da| continuum (Rosen 1979). The function on the left was determined from responses to the five most |ba|-like stimuli on the continuum, while for the function on the right, the five most |da|-like stimuli were used. Note the shift in identification function with the shift in stimulus range

region of increased sensitivity on a continuum compared to other regions easily accommodates the range effects by making this region of increased sensitivity give the limits over which the phoneme boundary may vary.

Such an escape is not possible for the range effects reported by Howell and Rosen (1983) on an affricate/fricative continuum cued by rise time. This continuum was considered particularly strong evidence for the natural-sensitivities hypothesis because Cutting and Rosner (1974) showed both it and an analogue nonspeech continuum to be categorically perceived. Furthermore, the discrimination peak occurred at nearly the same physical value of rise time (about 40 ms) for the two continua. An affricate/fricative continuum ($/tʃa/-/ʃa/$) was created by artificially contouring the onset envelope of a naturally produced $/ʃa/$. Two different subranges were presented to subjects: 0 to 80 ms and 40 to 120 ms. As seen in Figure 5, there is an enormous change in the position of the phoneme boundary. In this case, the range of stimuli seems to be the primary determinant of the placing of the boundary. No single region of sensitivity can account for these results. Even more damning, it turns out (as we will discuss in detail later) that the 40 to 120 ms range is more representative of real speech sounds than the 0 to 80 ms used by Cutting and Rosner

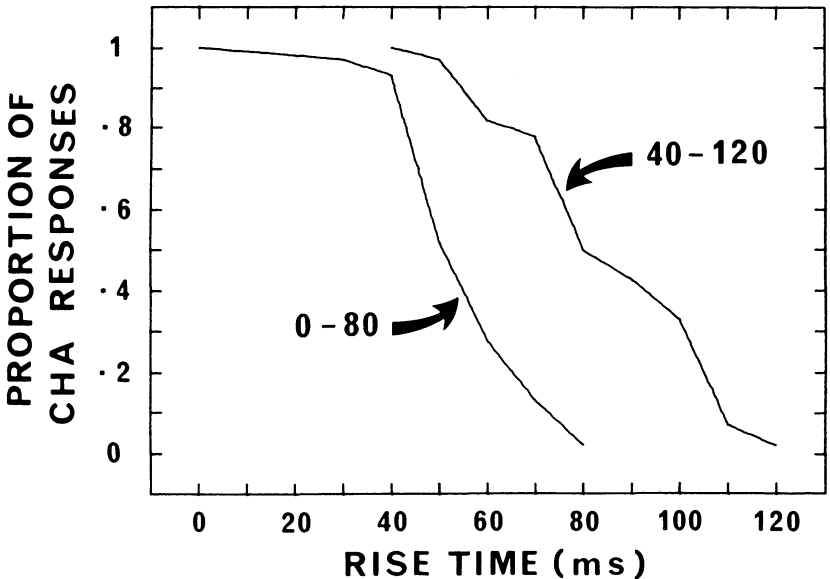


Figure 5. Categorization functions on a $/tʃa/-/ʃa/$ continuum for two subranges. The proportion of 'cha' responses is plotted as a function of stimulus rise time. Points from the two subranges (0-80 and 40-120 ms) are connected together and so labeled

(1974). A natural sensitivity at 40 ms, even if it existed, is irrelevant to real-life speech contrasts. A final point to note is the relative shallowness of the categorization curve for the longer (and more natural) range of rise times. If sharpness of categorization is desirable, and if speech contrasts evolved to make the best use of auditory sensitivities, natural rise times should occur in the shorter range, which they do not.

The central difficulty of all studies involving adult listeners' perception of speech continua is the near-impossibility of separating innate factors from the effects of learning and experience. It may well be that there are natural acoustic sensitivities in the infant that enable it to perceive some speech contrasts at first, but that these innate sensitivities are overshadowed by the learning of secondary acoustic aspects associated with the primary cues. These secondary cues would vary from language to language, even though the primary cues would not. Such a finding has already been reported by Simon and Fourcin (1978), who showed the development of a sensitivity to F1 transitions in the labeling of a voicing continuum for English children (where such a cue is distinctive) but not for French children (where it is not).

There are three ways of getting around this problem of mixed innate and experiential factors, and we will discuss each of them in turn. With adult listeners, we can avoid the effects of language experience by using stimuli that are not perceived as speech, but which have analogous acoustic characteristics. The use of infants as experimental subjects means that some perceptual processes can be investigated before developmental processes are completed. Finally, we can use animals with auditory systems similar to our own, where auditory properties can be expressed unaffected by linguistic factors.

4. Evidence from nonspeech continua and their supposed speech analogues

The important demonstrations of categorical perception for nonspeech continua from the point of view of assessing the natural-sensitivities account of speech perception are those in which the cues contrasting the endpoints of the nonspeech continuum can be directly linked with a categorically perceived speech contrast. There are two principal demonstrations which meet this criterion. The first is the categorical perception of rise time, which has been linked to perception of the affricate-fricative distinction (Cutting and Rosner 1974). The second is the perception of relative onset time of complex signals that contain two components (Miller et al. 1976; Pisoni 1977), which has similarities to the VOT distinction.

A. Categorical perception of rise time

Cutting and Rosner (1974) generated a nonspeech continuum of sawtooth wave forms which varied only in rise time. A wave form which reaches its maximum amplitude right at the start of the sound has a rise time of 0 ms and sounds like a note produced by plucking a string. The rise time can be systematically varied as illustrated in Figure 6, and at longer values of rise time the stimulus sounds like that produced by bowing a string. Cutting and Rosner constructed a continuum of such sounds varying in rise time from 0 to 80 ms and required listeners to classify a randomization of them as 'plucks' or 'bows'. They went on to assess discrimination using the ABX discrimination task. All of the criteria for declaring the continuum categorically perceived were met by these results. As we have noted, rise time is an important cue for the affricate/fricative distinction. For example, the affricate /tʃ/ has a more rapid rise than the fricative /ʃ/ (Gerstman 1957). The categorical perception of the speech and nonspeech rise-time continua and the close correspondence of the boundaries for the two continua (in each case the boundaries were at about 40 ms) has been regarded by Stevens (1981) as being particularly important since it appears to point to a natural sensitivity of the auditory system to distinguish interrupted consonants (such as /tʃ/) from continuant consonants (such as /ʃ/).

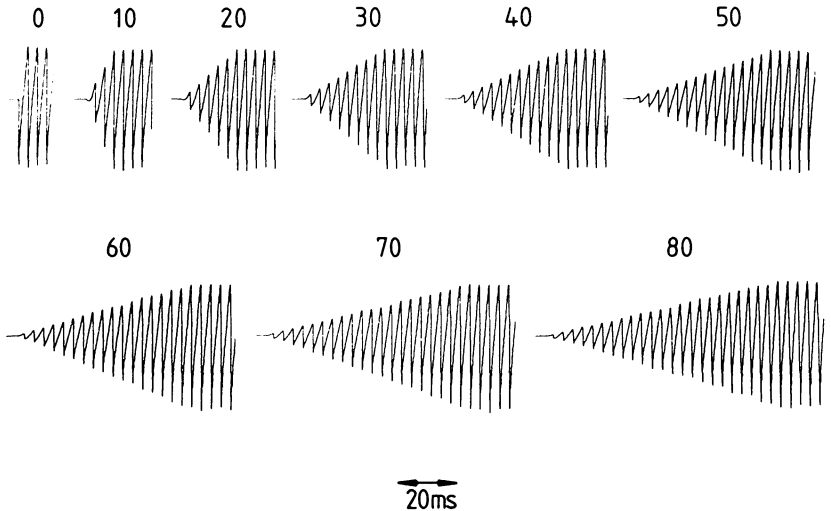


Figure 6. Initial portions of oscillographic traces of a sawtooth wave form varying in rise time. Presented above each stimulus is its rise time. Subjects are able to label sounds with short rise times as plucks and those with longer rise times as bows (Rosen and Howell 1981 with permission of Perception and Psychophysics)

Rosen and Howell (1981) investigated the role of this so-called natural sensitivity at 40 ms in the identification of stimuli from the pluck-bow continuum by varying the range of stimuli presented to the subject. A set of sawtooth stimuli with rise time varying from 0 to 80 ms was synthesized digitally, those shown in Figure 6. One subject was asked to label the stimuli as plucks or bows in three different conditions. The only difference between the conditions was the range of stimuli presented: 0-40 ms, 20-60 ms, or 40-80 ms. The results are shown in Figure 7, where a clear range effect is seen. The boundary for the range 0-40 is less than the boundary for the range 20-60, which, in turn, is less than the boundary for the range 40-80 ms. This result indicates that the boundary Cutting and Rosner estimated from an identification task (about 40 ms) is not a natural sensitivity at all but a value near the midpoint of the stimulus range (0-80) they used.

Stevens's natural-sensitivities account of speech perception emphasizes the naturalness of the discrimination peak. While range effects have been shown on identification functions, they have not, so far, been shown to

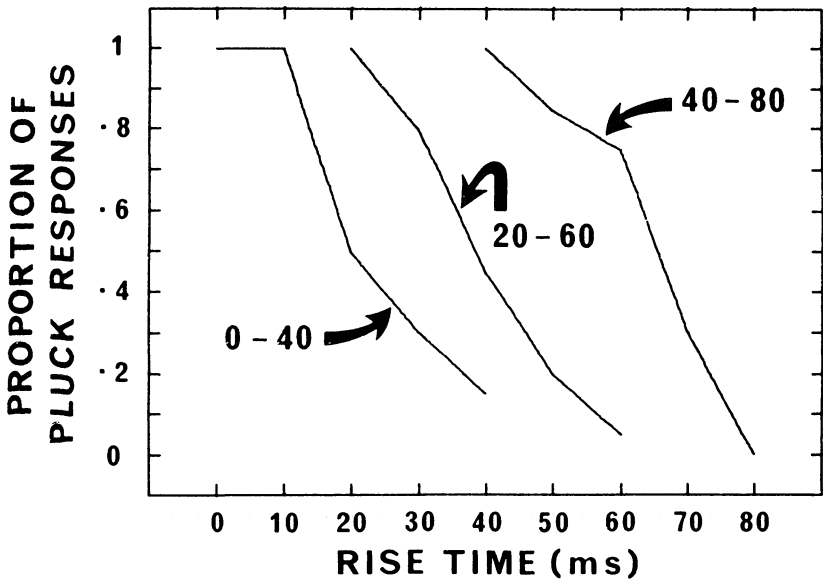


Figure 7. Identification functions of sounds drawn from various subranges of the pluck-bow continuum shown in Figure 6. For a range of stimuli drawn from a subrange with short rise times (0-40 ms), the subject changes categorizations over from 'pluck' to 'bow' at shorter values of rise time than for an intermediate range of rise time (20-60 ms), which in turn changes over categorizations at shorter values than at the longest range (40-80 ms)

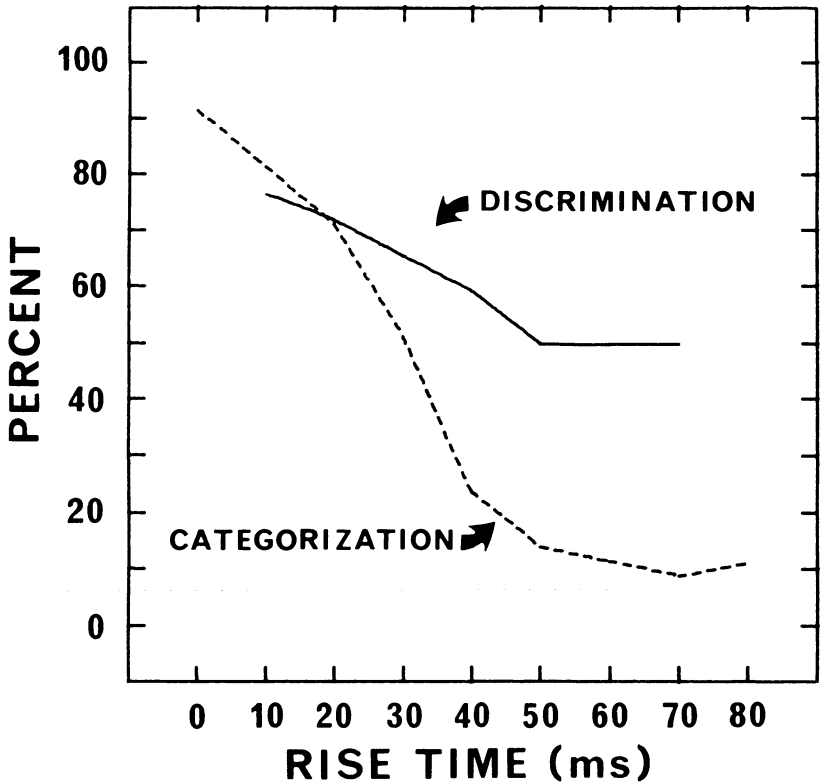


Figure 8. *Categorization and discrimination functions for a rise-time continuum (taken from Rosen and Howell 1981 with permission of Perception and Psychophysics)*

affect discrimination results. It was with this issue in mind that we set to work with Cutting and Rosner's (1974) rise-time continuum.

We (Rosen and Howell 1981) first attempted to replicate the results reported by Cutting and Rosner (1974) with the rise-time continuum of Figure 6. Eight subjects performed identification and discrimination tasks with the stimuli. The results obtained are shown in Figure 8.

The categorization function is sharp and there is a boundary at a rise time of about 30 ms. There is, however, no midcontinuum peak in the discrimination function. Discrimination is always best at the shortest rise times and decreases monotonically with increases in rise time. Our own results stood the test of several replications.

These results on the identification and discrimination of the pluck-bow continuum are consistent with the results of van Heuven and van den Broeke (1979), who found the variability of settings in a rise-time

Table 2. Rise times of Cutting and Rosner's stimuli, measured in milliseconds from oscillographic traces of the original experimental tape

Nominal									
	0	10	20	30	40	50	60	70	80
Measured	4	6	15	19	37	43	57	66	76

reproduction task to increase linearly with increasing rise time. In the light of our several failures to replicate Cutting and Rosner (1974), we considered that the serious discrepancy between our results and Cutting and Rosner's could not be attributed to differences in the way the experiments were performed. We were able to obtain the original tape used by Cutting and Rosner, and with it obtained results very similar to theirs. In particular, we found a clear discrimination peak at about 40 ms.

When we took oscillograph recordings of Cutting and Rosner's stimuli and measured the rise times, we found that the rise times were not as specified but had the rise times presented in Table 2. Oscillograms of Cutting and Rosner's stimuli are presented in Figure 9.

As seen clearly in the figure and the table, the measured rise times for Cutting and Rosner's stimuli are quite different to the nominal values. Furthermore, they are wrong in a way that explains Cutting and Rosner's

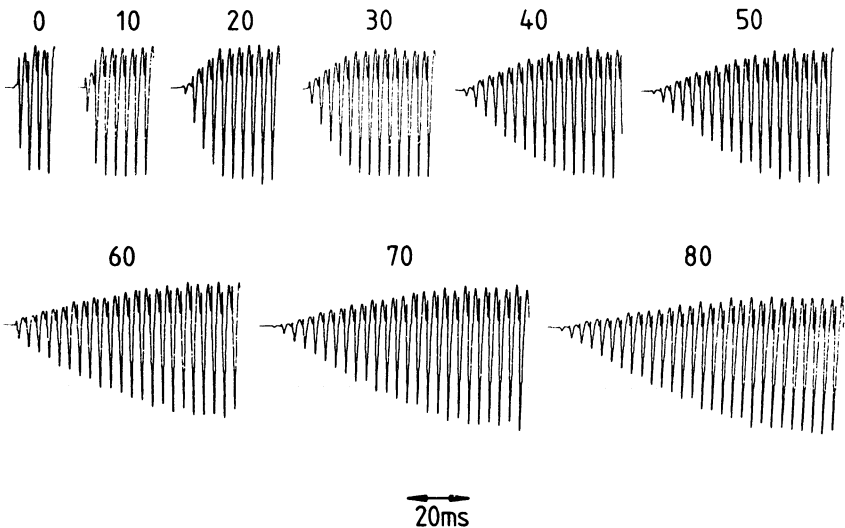


Figure 9. Oscillographic traces of the rise-time continuum used by Cutting and Rosner (1974). These are presented in the same manner as the rise-time continuum in Figure 6 (from Rosen and Howell 1981 with permission of Perception and Psychophysics)

Table 3. *Difference in rise times, measured in milliseconds between the seven two-step ABX pairs for nominal and measured values*

	Stimulus pair						
	0/20	10/30	20/40	30/50	40/60	50/70	60/80
Nominal	20	20	20	20	20	20	20
Actual	11	13	22	24	20	23	19

discrimination results, as seen in Table 3. Here are calculated both the nominal and the measured difference in rise time for the seven two-step ABX pairs. The physical difference between the stimuli is actually greatest at the point Cutting and Rosner found their discrimination peak. Hence, this would explain why Cutting and Rosner found a peak at this point. The nonlinear spread of the stimuli along the rise-time continuum would also sharpen up the categorization function.

We showed theoretically and empirically that a peak would occur in the discrimination function with rise times of these, rather than the nominal values. Empirically, we created a continuum with the rise times measured from Cutting and Rosner's stimuli and were able to obtain clear midcontinuum peaks in the discrimination function (Figure 10). Theoretically, we were able to fit a relatively simple Weber function to the ABX discrimination results found with our subjects and a rise-time continuum with rise times evenly spaced by 10 ms. We then showed, assuming that Cutting and Rosner's subjects behaved like ours, that a peak in the discrimination function at about 40 ms would be expected for the values of rise times we had measured for their stimuli (Rosen and Howell 1981).

Rise time on this nonspeech continuum is not, then, categorically perceived with respect to the criteria usually adopted. But Cutting and Rosner also reported that a /tʃa/-/ʃa/ continuum cued by rise time is perceived categorically. If THIS is correct, categorical perception would be peculiar to speech. That is in direct contradiction to Stevens's view. It would be better from Stevens's point of view if the analogous speech continuum is not categorically perceived.

Because of this and because of the problems encountered with discrimination of rise time on the nonspeech rise-time continuum, we decided to check discrimination of the speech sounds too. Before we could embark on this issue, we encountered a further problem. In our early attempts at synthesizing a /tʃa/-/ʃa/ continuum varying on the same range of rise times as those reported by Cutting and Rosner, we found that the stimuli with short rise times sounded unnatural. The range of rise times used by Cutting and Rosner seemed to be based on the early measurements made

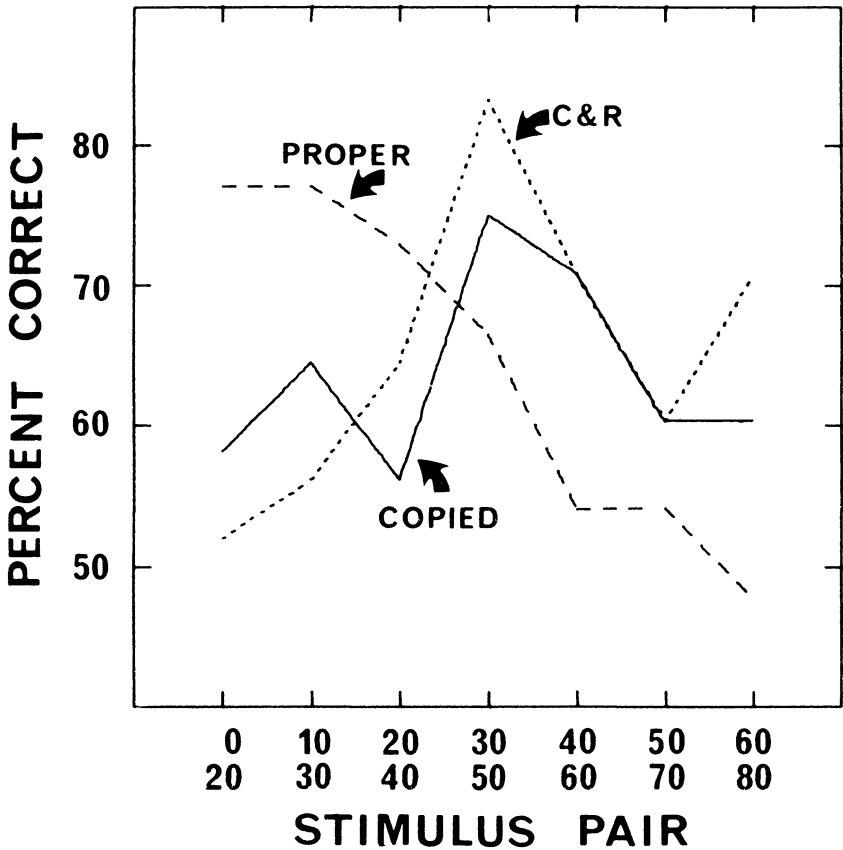


Figure 10. *ABX* discrimination results for three different rise-time continua. The 'proper' stimuli had the correct rise times, 'C & R' stimuli were from the original tapes of Cutting and Rosner (1974), and the 'copied' stimuli had the same rise times as measured from the original Cutting and Rosner (1974) stimuli as shown in Table 2 (taken from Rosen and Howell 1981 with permission of Perception and Psychophysics)

by Gerstman (1957). We measured rise times of the voiceless affricates and fricatives when in initial, medial, or final position in nonsense syllables, isolated words, and words in running speech from an oscillographic display (Howell and Rosen 1983). Oscillograms of representative stimuli are presented in Figure 11.

The mean rise times of affricates and fricatives for four subjects are shown in Table 4 for the three types of speech material. Two things are of note: first, there is a difference in rise-time values in the three speech contexts. Second, the rise times we measured are appreciably longer than

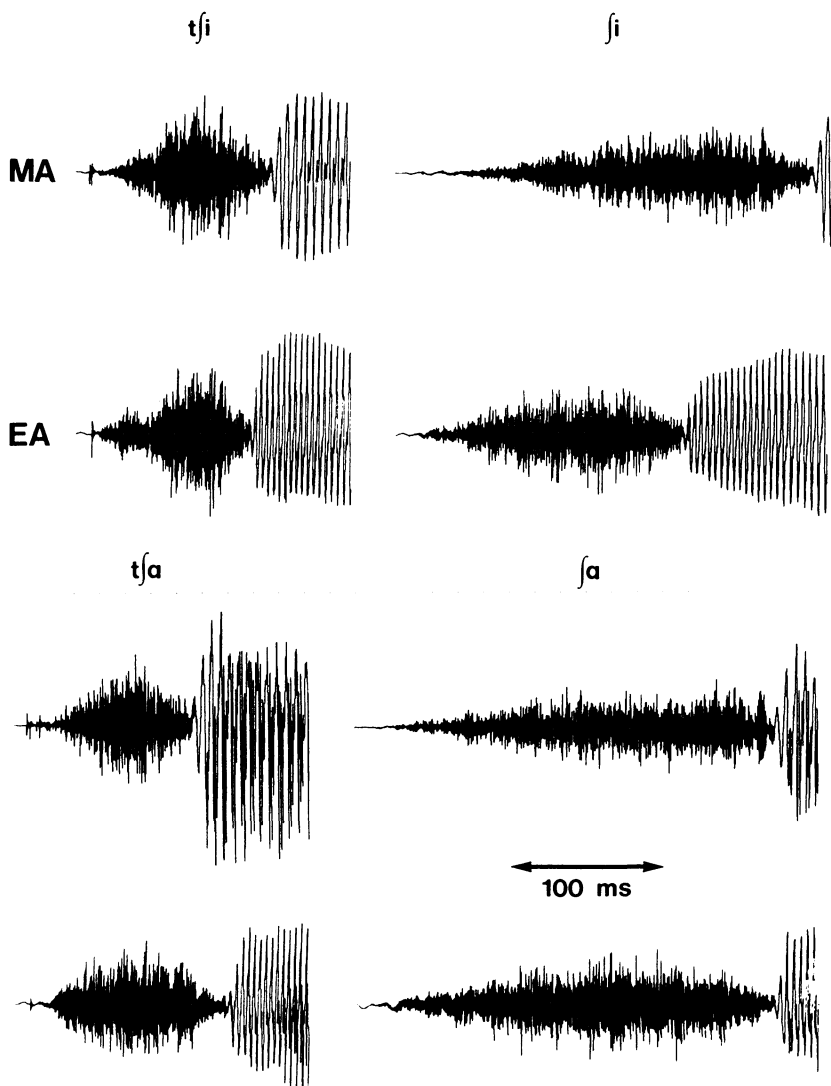


Figure 11. Oscillograms of /tʃi/, /ʃi/, /tʃa/, and /ʃa/ spoken by EA and MA

those measured by Gerstman (1957). The first finding indicates that for an optimal division of affricates from fricatives, a listener would have to vary his phoneme boundary depending on the type of speech material he is hearing. A fixed sensitivity would not be appropriate for this. The second finding indicates that if there was a sensitivity at 40 ms, this would not serve to distinguish natural affricates from fricatives, particularly for

Table 4. Mean rise times (in milliseconds) of affricates and fricatives drawn from three contexts: running speech, isolated words, and nonsense syllables

	Affricates	Fricatives
Running speech	33	76
Isolated words	49	123
Nonsense syllables	61	120

material most like the synthetic utterances Cutting and Rosner (1974) and Gerstman (1957) used (nonsense syllables in isolation). In fact, a value of 40 ms applied to our production measurements of affricates and fricatives on nonsense syllables would have all but one classified as fricatives.

Does a discrimination peak occur for the speech continuum? If so, is it at 40 ms or roughly in the middle of the range of rise times a /tʃɑ/–/ʃɑ/ continuum SHOULD vary on (i.e. at about 80 ms — the middle of the 40–120 ms range)? Standard identification and discrimination experiments were performed on the subranges 0–80 and 40–120 ms for the contoured speech continuum described in section 3.C. In the identification task listeners appear to place the boundary roughly at the midpoint of the range. Cutting and Rosner found a boundary at 40 ms for their speech continuum because of the subrange they employed and not because 40 ms represents a natural sensitivity. With the 0–80 ms range, discrimination decreased monotonically, which is reminiscent of the results we have reported earlier for the nonspeech rise-time continuum. For the 40–120 ms range, the discrimination function is flat across the entire range (Figure 12). In summary, identification boundaries are affected by the subrange of stimuli, discrimination is similar to that observed with the nonspeech continuum when the subranges correspond, and nowhere is a discrimination peak observed.

From the natural-sensitivities view, this outcome is not too bad. No one has claimed that all phonemic contrasts are mediated by natural sensitivities, and this particular contrast, despite earlier claims, is clearly one continuum that is not.

B. Categorical perception of relative onset-time continua

It has been reported that categorical perception occurs when stimuli consisting of a sound which has two components varying in their relative onset are presented to be identified and discriminated (Miller et al. 1976; Pisoni 1977). Miller et al. (1976) employed a noise–buzz continuum where

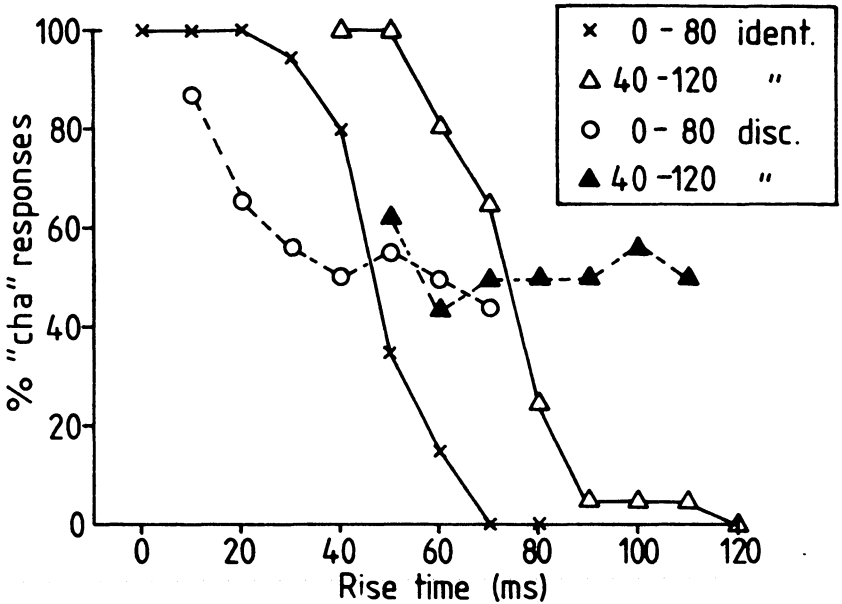


Figure 12. Identification and discrimination functions for two subranges of an affricate-fricative continuum created by contouring the rise of a naturally spoken [ʃa]

the noise onset started 80 ms before to 10 ms after the buzz (in 10 ms steps). They found a discrimination peak when the noise led the buzz by 16 ms. Pisoni (1977) used two-tone complexes consisting of tones at frequencies of 500 Hz and 1500 Hz. The level of the 500 Hz tone was 12 dB higher than the 1500 Hz tone. Pisoni found a discrimination peak when the tones were offset by about 20 ms, whichever tone led. He noted that the ability to discriminate the relative onsets of two sounds might be involved in decisions about whether a sound is voiced or voiceless.

It is of note that Stevens has not used these particular nonspeech findings in his natural-sensitivities account of speech perception. The reason for this may be that Stevens and Klatt (1976) consider that the invariant acoustic factor that underlies VOT is more complex than this involving the rapid spectral change at voicing onset and not solely relative onset time.

One problem in relating the onset-time data to speech contrasts is that the discrimination peaks found do not correspond with discrimination peaks found in their supposed analogues — VOT continua (Table 5). The peaks in discrimination performance occur at lower values than those found on VOT continua. Yet a simple auditory sensitivity underlying VOT discriminations would require that they correspond. It may be that

Table 5. *Nonspeech onset-time boundaries (and VOT in studies where they are reported too) in milliseconds*

Author	Type of nonspeech continuum	Estimate	Estimate for speech	
Miller et al. (1976)	noise-onset time	+ 16	–	
Pisoni (1977)	tone-onset time	+ 20	–	
Summerfield (1982)	noise-onset time	Exp 1 F1 = 200 + 27	Exp 1 F1 = 200 + 50	
		= 300 + 20	= 300 + 41	
		= 400 + 17	= 400 + 32	
		= 500 + 19	= 500 + 30	
		Exp 2 F1 = 200 + 21	Exp 2 F1 = 200 + 40	
			= 300 + 21	= 300 + 36
			= 400 + 21	= 400 + 30
			= 500 + 20	= 500 + 28
	tone-onset time	Exp 1 F1 = 200 + 20		
		= 300 + 14		
= 400 + 13				
= 500 + 12				
Exp 2 F1 = 200 + 20				
		= 300 + 23		
		= 400 + 19		
		= 500 + 22		

the location of the boundary is affected by secondary (learned) factors which are not invariantly related to the phoneme but are associated with it (cf. Klatt 1979).

If secondary cues affect the location of the discrimination peak, this would detract somewhat from the attractive simplicity of this type of theory. Even this retreat is not available because it is contradicted by some recent data. Summerfield (1982) reports that continua like Miller et al.'s (1976) and Pisoni's (1977) are not affected by the frequency of the low-frequency tone, unlike VOT continua whose VOT boundaries shift up as F1 shifts down. Thus, the discrimination peak of these analogues is not affected by 'analogous' secondary cues in the same way that speech is.

5. Evidence from infants

Investigations of the perception of speech by infants were initiated by Eimas et al. (1971) in a now-classic study using a bilabial VOT continuum (/b-/p/). Employing a dishabituation paradigm, they showed that one-

and four-month-old infants were better able to discriminate a pair of stimuli straddling the adult English voiced–voiceless boundary than a pair within a phoneme category, even though the physical difference in VOT between the two stimuli was always 20 ms. This effect is clearly similar to the discrimination peak found with adult listeners and was thus taken as evidence for some kind of innate basis for categorical perception. Although few would argue with this interpretation, Eimas et al. (1971) went further in claiming that this represented perception in ‘a linguistic mode’. Unfortunately, as we have already pointed out, the position of the phoneme boundary on the VOT continuum varies across languages. Yet Spanish infants also have their greatest sensitivity to VOT changes for stimuli across the English boundary, although their own language has its boundary elsewhere (Eilers et al. 1979). Furthermore, at six months of age, Spanish infants do about as well with stimuli straddling the Spanish as the English boundary. Thus it seems likely that the English boundary is an AUDITORILY determined one, in contrast to the LINGUISTIC determination claimed by Eimas et al. (1971), since all infants seem to show it. It is not, however, necessary for such an auditory sensitivity to exist for infants to develop appropriate perception of a contrast. The natural-sensitivities account, even if some sensitivities DO exist, loses much of its appeal in such light.

Discrimination peaks for adult boundary-straddling stimuli have also been shown in infants for an English voiced-stop place continuum by Eimas (1974) and Miller and Morse (1976). Nothing is known about the generality of such results across languages.

Two experiments have now been reported testing infant capabilities on the nonspeech continua that are meant to be analogues of categorically perceived speech continua. One group (Juszyk et al. 1977) had the misfortune to use the sawtooth rise-time continuum of Cutting and Rosner (1974). They found better discrimination for a 30–60 ms pair than for either 0–30 ms or 60–90 ms and claimed this as evidence for an innate basis for categorical perception. As we have already pointed out, these stimuli were improperly constructed. The continuum is not, in fact, categorically perceived, and the infants’ performance is predictable from a consideration of the actual rise times of the stimuli and a monotonically decreasing discrimination function.

Of more interest is work with the tone-onset time (TOT) continuum studied in adults by Pisoni (1977). Juszyk, Pisoni, Walley, and Murray (1980) presented 7–13-week-old infants with various pairs of stimuli from the TOT continuum in a dishabituation paradigm. Although the physical difference between the pairs was always 30 ms, there were differences in the relative discriminability of the pairs. The pairs the infants did best

on were not, however, the stimulus pairs adults discriminated best, but pairs with longer onset asynchronies. This seems somewhat disturbing from the natural-sensitivities point of view, although the authors did not find it so. Infants are meant to perceive with little (or none) of the long experience of phonetic and phonological processes of language. Therefore, with nonspeech stimuli, avoiding the adult's extensive knowledge of speech sounds, adults and infants should show similar regions of sensitivity. It could be argued that infants were in general less sensitive and therefore had their peak sensitivities at longer onset asynchronies. However, if a simultaneity/successiveness judgment is the basis of the voicing distinction for stops, infants should also show displaced boundaries for speech VOT stimuli.

6. Evidence from speech continua and animal listeners

The final evidence which we will consider in our evaluation of the natural-sensitivities account of phonemic perception comes from studies conducted on two species of mammals: chinchillas and monkeys. The data are of considerable interest since there is no possibility of a specifically linguistic component in the performance of animals. Since mammalian auditory systems, including those of humans, have much in common, if natural auditory sensitivities did exist, it should be possible to demonstrate them in nonhuman species. Although it is difficult to determine from present evidence which auditory factors are involved in the tasks performed by animals, the data are impressive enough to offer substantial support to a natural-auditory-sensitivities account.

A. VOT continua

Kuhl and Miller (1975, 1978) have explored chinchillas' ability to differentially 'label' stimuli from VOT continua. Using an instrumental avoidance technique, the animals were first trained to respond properly to the end points of a VOT continuum with appropriate rewards and punishments. Once proper behavior to the endpoint stimuli had been established, the animals' performance was evaluated for stimuli along the whole continuum in a generalization paradigm. Only for endpoint stimuli were the animals given feedback appropriate to their response. For intermediate stimuli, they were always rewarded no matter what their behavior. In this way one generates a function that is equivalent to an identification function derived from humans.

Kuhl and Miller (1975) showed that the identification function for an alveolar VOT continuum derived from chinchillas was very similar to that obtained from English-speaking adults with nearly identical phoneme boundaries. They then proceeded to argue that this boundary was psychoacoustically determined and thus that the English VOT boundary was a natural one. This was not particularly compelling evidence if one supposed that both the chinchillas and the humans were subject to range effects and were simply placing their boundary near the midpoint of the continuum. We have already discussed the existence of such effects for human subjects, and they have been shown to occur in animals as well by Waters and Wilson (1976). They used a similar technique to Kuhl and Miller (1975) to obtain identification functions on a VOT continuum from rhesus monkeys. When the endpoints of the continuum were -140 and $+140$ ms, the phoneme boundary was about $+22$ ms, but when the endpoints were changed to 0 and $+140$ ms, the boundary shifted to about $+66$ ms. The animals appear to place their boundary toward the midpoint of the range of sounds spanned by the VOT tokens. Kuhl (1981) implies that these range effects are negligible, but a shift of 44 ms cannot be disregarded. This shift is, in fact, much bigger than has been reported for range manipulations in humans and implies that some of the resistance of human phoneme boundaries to changes in range on VOT continua is due to linguistic experience.

Much more convincing evidence for natural auditory sensitivities is presented by Kuhl and Miller (1978), who looked at chinchilla and human identification functions for labial and velar, as well as alveolar, VOT continua. They found that the category boundaries for chinchillas shifted with place of articulation in the same way as they did for humans. Again, the correspondence between the values obtained for chinchillas and humans is striking, as can be seen in Table 6. Since the range of stimuli is constant for all places of articulation, the shift in boundary across places of articulation cannot be due to range effects.

There is at least one major conundrum in the interpretation of this data. Since boundary estimates for VOT continua vary across the languages of the world, why should chinchillas show such remarkably close correspondence with their masters' language and not some other?

Table 6. *Boundary estimates for the voiced-voiceless distinction for animals and humans (from Kuhl and Miller 1978)*

	Bilabial	Alveolar	Velar
Animals	+23.3	+33.3	+42.5
English-speaking humans	+26.8	+35.2	+42.3

Another way around the problem of range effects is to use a discrimination paradigm. With humans, as we have already pointed out, unusual discrimination is the distinguishing feature of categorical perception. This, of course, should apply to the chinchillas if they are responding to speech sounds similarly to humans. Specifically, since it had been shown that VOT in the chinchillas is categorized similarly to that in humans, do the animals also show a discrimination peak?

This is the question Kuhl (1981) examined. She used a same-different discrimination task instead of the usual ABX format, presumably because the latter is more difficult for the animals to learn. For standards varying along the entire VOT continuum, she estimated the stimulus that was just discriminable from the standard using an adaptive procedure. If the chinchillas discriminate the continuum in a categorical manner, the distance between the standard and just-different stimulus should be smallest for standards near the phoneme boundary. Also, for midcontinuum stimuli where it is possible to estimate the just-discriminable stimulus from either more positive or more negative values of VOT, smaller differences should be just discriminable with comparison stimuli that approach the standard from the direction of the phoneme boundary. Both these things were found.

Kuhl's (1981) study was well designed and executed, and the results offer clear support for heightened discriminability of the category boundary on a VOT continuum, and thereby for a natural-sensitivities account. It is, perhaps, worth observing that the psychophysical procedures Kuhl (1981) adopts are more satisfactory than those that have been typically employed with human subjects.

B. Place continua

Evidence for natural sensitivities is rather more ambiguous for place continua. Positive evidence has been reported by Morse and Snowdon (1975) for rhesus monkeys using a /bae/-/dae/-/gae/ continuum. The technique they used is more closely related to the dishabituation paradigm used in infants, but using a measurement of heart rate rather than a voluntary sucking response. When monkeys are presented with a sound, their heart rate decelerates. With continued presentation of the same sound, this response habituates, i.e. the heart continues to beat at a constant rate. If however, a novel stimulus is presented, the response dishabituates, which is to say, its rate decelerates once again. It is thus possible to obtain a measure of the ability of monkeys to discriminate between different sounds by presenting one stimulus 20 times followed by a different stimulus and then noting any changes in heart rate.

Using this technique, Morse and Snowdon (1975) showed that although the monkeys exhibited dishabituation even to stimuli that were within human phoneme classes, they showed a greater dishabituation to stimuli drawn from across human phoneme boundaries. Thus the monkeys showed greater sensitivity at the human phoneme boundaries and could therefore be said to have peaks in their discrimination function at places appropriate for this speech contrast.

Sinnot et al. (1976) also investigated a place contrast (/ba-/da/) with monkey listeners. They used an instrumentally conditioned discrimination task to obtain psychometric functions where the standard stimulus was always one of the endpoints of the continuum. They also ran human subjects with the same paradigm (with money instead of banana-flavored pellets as reward) for comparison. Unfortunately the data are not presented in a way which makes it easy to assess whether there is increased discrimination at the adult phoneme boundary or not. Only the psychometric functions are given, and no explicit analysis of discrimination measures as a function of stimulus position on the continuum is performed. Nevertheless, it is clearly seen that both monkeys and humans showed a smooth improvement in discrimination performance as the comparison stimulus moved away from the standard. There was no sudden jump at the phoneme boundary, as would be expected if a natural sensitivity existed.

7. Conclusions

There are two main aspects of the natural-sensitivities account of phonemic perception that need to be distinguished. First of all, we need to consider if any sensitivities do in fact exist. Once this is assured, we need to assess their role in speech perception and its development through infancy and childhood.

There is reasonably strong evidence for natural sensitivities from three sources. Nonspeech with at least some similarity to speech in its acoustic form shows increased discriminability in certain regions, although this has only been shown for VOT-like continua. Infants show increased discriminability at adult boundaries for place and VOT continua, as do animals.

Even if the existence of some natural auditory sensitivities is granted, the role of these sensitivities in speech perception and its development is still problematic. As we have already seen, the sensitivity can be in the wrong place on the continuum for a proper partitioning of incoming stimuli into phonemic classes. There is a further caveat in interpreting results from infants and animals. The tasks we assign to the animal and

infant are vastly different from the task an adult faces. The infant and animal are operating on an acoustic basis in what is essentially an arbitrarily defined task. Any difference between stimuli, even ones irrelevant for the phonemic contrast in general, will help. This is quite different from the task faced by the adult or child trying to extract a meaningful message from an acoustic signal. Here, irrelevant features must be ignored and only distinctive aspects attended to. This ability can only develop over time (Simon and Fourcin 1978). Simply because an animal or an infant finds it easier to discriminate stimuli differing in some additional way from another pair of stimuli does not mean that that cue will be attended to at the beginning of development.

In short, it seems that such natural sensitivities as do exist can only be of minor importance in speech perception and its development. Although they may play some role, they are neither necessary nor sufficient to explain the perception of any single phonemic contrast.

*Department of Psychology
University College London
Gower Street
London WC1E 6BT
England*

*Department of Phonetics
and Linguistics
University College London,
4 Stephenson Way
London NW1 2HE
England*

Note

1. Both authors are supported by the Medical Research Council of the United Kingdom. Thanks to Professor A. J. Fourcin for his suggestions on the manuscript.

References

- Brady, S. A., and Darwin, C. J. (1978). Range effects in the perception of voicing. *Journal of the Acoustical Society of America* 63, 1556–1558.
- Cutting, J. E., and Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception and Psychophysics* 16, 564–570.
- Diehl, R. L., Souther, A. F., and Convis, C. L. (1980). Conditions on rate normalisation in speech perception. *Perception and Psychophysics* 27, 435–453.
- Eilers, R., Gavin, W., and Wilson, W. (1979). Linguistic experience and phonemic perception in infancy. *Child Development* 50, 14–18.
- Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics* 16, 513–521.
- , Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306.
- Elman, J. L., Diehl, R. L., and Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62, 971–974.

- Gerstman, L. J. (1957). Perceptual dimension for the friction portion of certain speech sounds. Unpublished doctoral dissertation, New York University.
- Halle, M., and Stevens, K. N. (1980). Some reflections on the theoretical basis of phonetics. In B. Lindblom and S. E. G. Ohman (eds.), *Frontiers of Speech Communication Research*. London and New York: Academic Press.
- Howell, P., and Harvey, N. (1983). Perceptual equivalence and motor equivalence in speech. In B. Butterworth (ed.), *Language Production*, vol. 2. London: Academic Press.
- , and Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America* 73, 976–984.
- Jusczyk, P. W., Pisoni, D. B., Walley, and Murray (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America* 67, 262–270.
- , Rosner, B. S., Cutting, J. E., Foard, C. F., and Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception and Psychophysics* 21, 50–54.
- Klatt, D. H. (1975). Voice onset time, friction and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18, 686–706.
- (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7, 279–312.
- Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities to the perception of speech-sound categories. *Journal of the Acoustical Society of America* 70, 340–349.
- , and Miller, J. D. (1975). Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. *Science* 190, 69–72.
- , and Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63, 905–917.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74, 431–461.
- , Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.
- , and Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, and H.-L. Teuber (eds.), *Handbook of Sensory Physiology*, vol. 3: *Perception*. New York: Springer.
- Lisker, L., and Abramson, A. S. (1970). The voicing dimension: some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 1967. Prague: Academia.
- , and Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language* 47, 767–785.
- Miller, C., and Morse, P. (1976). The heart of categorical speech discrimination in young infants. *Journal of Speech and Hearing Research* 19, 578–589.
- Miller, G. A. (1956). The magical number seven plus or minus two, or, some limits on our capacity for processing information. *Psychological Review* 63, 81–96.
- Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (1976). Discrimination and labelling of noise–buzz sequences with varying noise-lead times: an example of categorical perception. *Journal of the Acoustical Society of America* 60, 410–417.
- Morse, P. A., and Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics* 17, 9–16.
- Parducci, A. (1965). Category judgment: a range-frequency model. *Psychological Review* 72, 407–418.

- Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., et al. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance* 3, 686–696.
- Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two-component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61, 1352–1361.
- Poulton, E. C. (1977). Quantitative subjective assessments are almost always biased, sometimes completely misleading. *British Journal of Psychology* 68, 409–425.
- Rosen, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics* 7, 393–402.
- , and Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception and Psychophysics* 30, 156–168.
- Simon, C. and Fourcin, A. J. (1978). A cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America* 63, 925–935.
- Sinnott, J. M., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America* 60, 687–695.
- Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant and M. A. A. Tatbam (eds.), *Auditory Analysis and Perception of Speech*. London: Academic Press.
- (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America* 68, 836–842.
- (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: data from phonology, acoustics and psychoacoustics. In T. F. Myers, J. Laver, and J. Anderson (eds.), *The Cognitive Representation of Speech*. Amsterdam: North-Holland.
- , and Klatt, D. H. (1976). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America* 55, 653–659.
- Strange, W., and Broen, P. A. (1980). Perception and production of approximant consonants by three-year-olds. In G. Yemi-Komshian, J. F. Kavanagh, and C. A. Ferguson (eds.), *Child Phonology: Perception, Production and Deviation*, vol. 2. New York: Academic Press.
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7, 1074–1095.
- (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America* 72, 51–61.
- van Heuven, V. J. J. P., and van den Broeke, M. P. R. (1979). Auditory discrimination of rise and decay times in tone and noise bursts. *Journal of the Acoustical Society of America* 66, 1308–1315.
- Waters, R. S., and Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: the voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics* 19, 285–289.

