ON THE IDENTIFICATION OF SINE-WAVE ANALOGUES OF CV SYLLABLES[1]

Michael F. Dorman,[2] Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510, United States of America

In order to answer the question - Do infants perceive speech phonetically? - the stimulus continuum presented to the subjects must have phonetic category boundaries which are clearly dissociated from auditory category boundaries. For, if the two boundaries coincide, then the subjects' basis for response can not be determined. This situation, in the view of several authors, characterizes the identification of categories along the voice-onset-time (VOT) continuum. For example, Pisoni (1977) suggests that the auditory categories of simultaneous and nonsimultaneous onset could underlie infants' discrimination along the VOT continuum.

In the present series of experiments our aim was to determine whether auditory categories may also underlie infants' ability to discriminate between stop consonants which differ in place of articulation. An examination of the stimuli used in Eimas' (1974) and Miller and Morse's (1976) studies of infant place discrimination suggests a possible psychoacoustic basis for the discrimination between [bae] and [dae] - i.e., the discrimination could be based on the difference between frequency change and no frequency change in the second and third formant transitions. While the outcomes of the two studies lend little support to this position, we felt, nevertheless, that it would be important for future research to assess whether auditory categories generally coincide with phonetic categories along a continuum of F_2 and F_3 change.

The procedure used in our experiments was to present adults with consonant-vowel (CV) syllables synthesized with formant structure and CV analogues synthesized with frequency and amplitude modulated sine waves. Our rationale for this approach was that if listeners placed category boundaries at the same place along both the speech and nonspeech continua, then we should believe that, for these stimuli at least, the phonetic category boundaries coincide with acoustic category boundaries. These stimuli, of course, would be inappropriate for use with infants. If, on the other hand,

---

(1) This research was a collaborative effort among Dr. Peter Bailey, Dr. Quentin Summerfield and myself.

(2) Also, Arizona State University, Tempe, Arizona 85281, United States of America.

the two boundaries did not coincide, then the speech stimuli could well prove probative in studies of infant speech perception.

The stimuli for our first experiment were a [bo-do] continuum and a [be-de] continuum (see Figure 1). The first and third formants in both continua were identical - only the second formant differed between the two. The parameter values were selected so that both continua would be physically symmetrical but phonetically asymmetrical. We intended the phoneme boundary along the [bo-do] continuum to be associated with a falling transition so that the majority of the stimuli would be heard as [bo]. In contrast, we intended the phoneme boundary along the [be-de] continuum to be associated with rising transitions so that the majority of the stimuli would be heard as [de]. In this way we intended to dissociate phonetic boundaries from auditory boundaries that may accompany flat as opposed to rising or falling transitions, or from auditory boundaries that might simply coincide with the center of the stimulus range.

To generate identification functions for these stimuli, we presented the stimuli to our listeners in an AXB format. On each trial three stimuli were presented; the first and third members of the triad were the end points of the continuum, the second member was a stimulus drawn randomly from that continuum. The task of the listeners was to indicate whether the second stimulus was more like the first or more like the last member of the triad. We chose this task to avoid a problem usually associated with the absolute identification of nonspeech stimuli - that listeners have more difficulty attaching category labels to the nonspeech stimuli than to the speech stimuli. By presenting the end points of the stimulus continuum on each trial in both the speech and nonspeech conditions, we hoped to make the identification task equally difficult in both conditions.

Turning now to the result of our first experiment, we see in Figure 2 the identification function for the speech signals. As predicted there were more [b]-like responses for the [bo-do] continuum than for the [be-de] continuum. However, the difference between the locations of the phoneme boundaries fell short of significance. In contrast to the largely asymmetric identification functions shown for the formant stimuli, the identification functions for the sine-wave analogues, shown in Figure 3, coincide throughout their range. We would conclude from this outcome that at least the [bo-do] boundary does not coincide with an acoustic category boundary. There are, however, two possible interpretations of this outcome: the asymmetrical categorization of the formant continua could either be correlated with the way the stimuli are heard - as speech or nonspeech - or may simply be correlated with the different spectral properties of the formant and sine-wave stimuli. To rule out the latter possibility we would like the same physical signal to be heard as speech-like in one context and as nonspeech in another. If the category boundaries differed in this instance then it certainly could not be argued that spectral differences account for the outcome. Fortunately, the sine-wave stimuli used in our experiment fit this requirement nicely. After we instructed our listeners as to the nature of the sine-wave stimuli, they readily agreed that the stimuli could be heard as stop initiated.

The outcome of this experiment (when the sine waves were heard as speech) is shown in Figure 4. The pattern of results is clearly very different from that when the sine waves were heard as nonspeech. Here, the two functions no longer overlap. As with the formant stimuli, the majority of the [bo-do] analogues were heard as [b]-like. Moreover, the category boundaries along the two continua differed significantly. It is clear that the pattern of results obtained when the sine waves were heard as speech-like is more akin to that obtained for the formant stimuli than for that obtained when the sine waves were heard as non-speech. It appears, then, that the difference between the speech and nonspeech conditions was not due to the spectral differences as such, but, rather, was due to the way in which the stimuli were heard.

To assess the reliability of our first experiment we conducted a second experiment. For this experiment we synthesized a single [ba-da] continuum and a corresponding nonspeech analogue with sine waves. The speech continuum was more natural sounding than either of those used in our first experiment and, perhaps as a consequence, many listeners heard the sine-wave analogues as speech-like without prompting. Thus, we were able to divide our subjects into two groups on the basis of their perception of the sine-wave stimuli.

The identification function for the subjects who heard the sine waves as speech is shown in Figure 5 along with the identification function for the formant stimuli. The two functions are quite similar and two phoneme boundaries fall to the right of the mid-point of the stimulus continuum. In contrast, the identification function for the subjects who heard the sine waves as nonspeech appears quite different from that generated in response to the formant stimuli (see Figure 6). This difference is reflected in the significantly flatter slope of the nonspeech function and fewer [b]-like responses to the nonspeech stimuli.

Summary

Earlier in this paper we raised the question of whether auditory category boundaries generally coincide with phonetic category boundaries along F_2-F_3 continua. Unfortunately, our results provide an equivocal answer; the [bo-do] boundary in our first experiment did not seem to coincide with the auditory boundary, but the phonetic and auditory boundaries in the second experiment were uncomfortably close. Nevertheless, we have gained a significant purchase on a methodology that will allow us to dissociate auditory and phonetic boundaries. We see, then, an opportunity to construct continua which will be of use in the study of the ontogeny of phonetic perception.

Moreover, we see quite clearly that the perceptual system categorizes sine-wave stimuli as a function of how they are heard: when heard as speech they are categorized like formant stimuli; when heard as nonspeech they are categorized differently. We should wonder then what mechanisms underlie this changing percept of an unchanging stimulus. The nature of those mechanisms will be, I believe, the topic of Dr. Bailey's and Dr. Summerfield's paper.

References

Eimas, P.D. (1974): "Auditory and linguistic processing of cues for place of articulation by infants", Perc.Psych. 16, 513-521.

Miller, C.L. and P.A. Morse (1976): "The 'Heart' of categorical speech discrimination in young infants", JSHR 19, 578-589.

Pisoni, D.B. (1977): "Identification and discrimination of the relative onset times of two component tones: Implications for voicing perception in stops", JASA 61, 1352-1361.

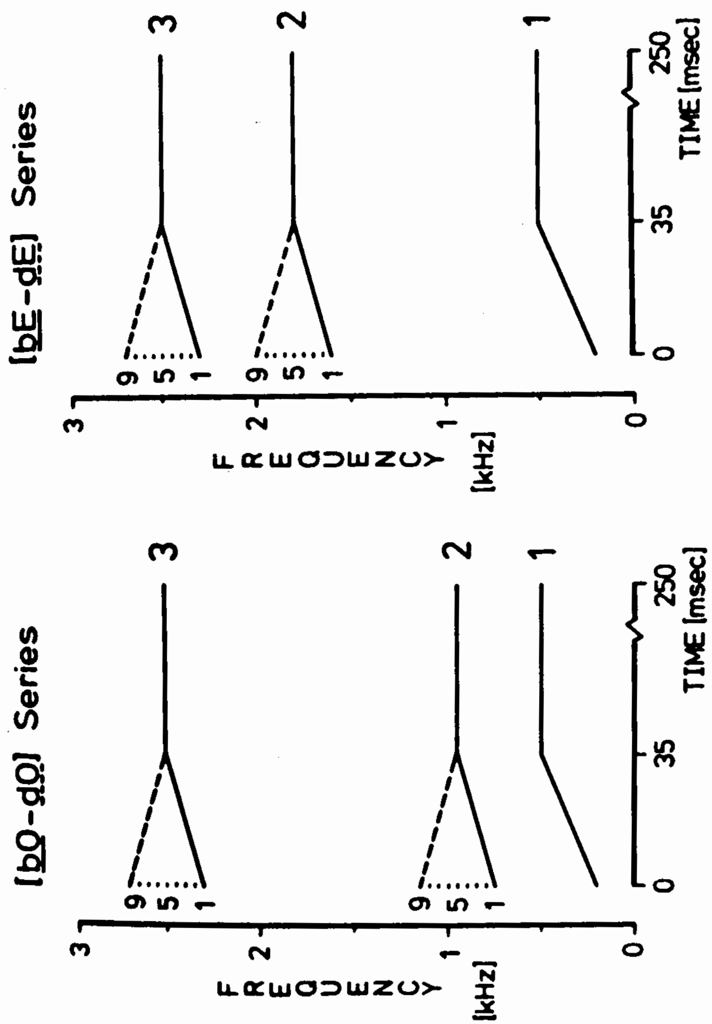

Figure 1

Stimuli for first experiment

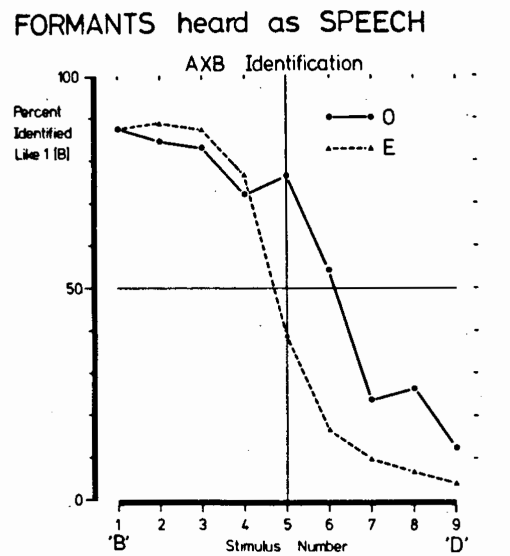

Figure 2

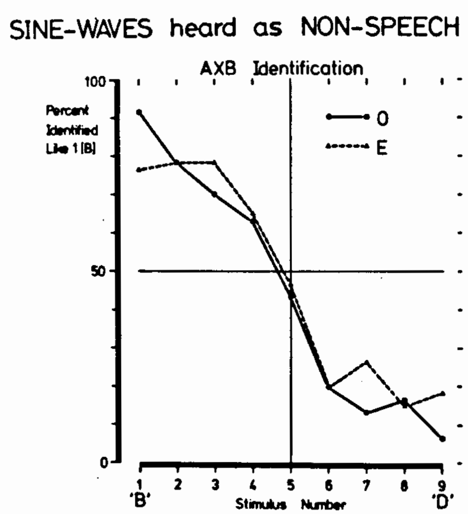

Figure 3

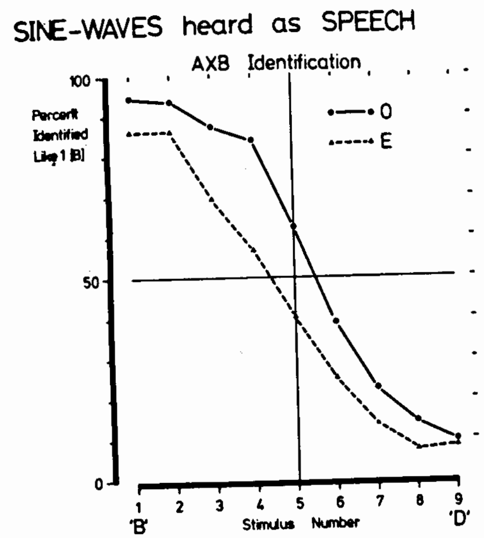

Figure 4

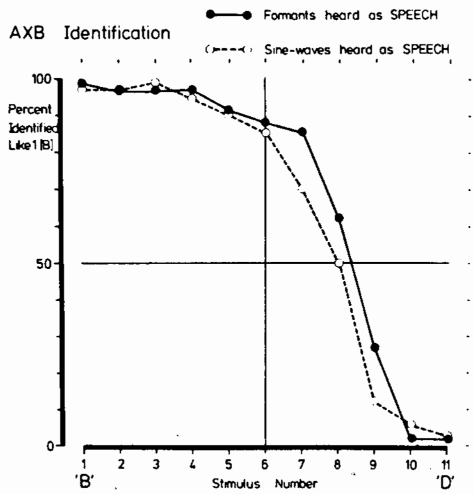

Figure 5

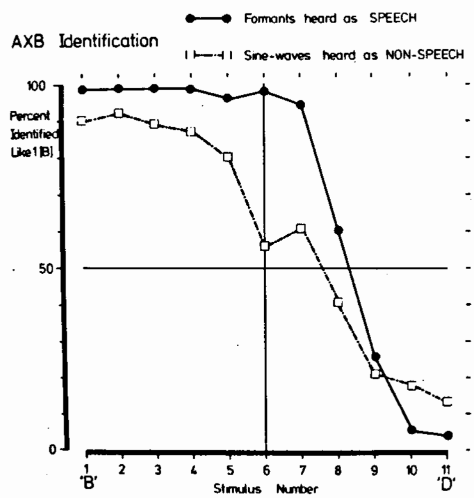

Figure 6