Category Boundaries for Speech and Nonspeech Sounds*

J. R. Sawusch and D. B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47401

Recent experiments employing an adaptation paradigm have demonstrated the possibility of phonetic feature detectors in speech perception (Eimas and Corbit, 1973). These results could be explained by Adaptation Level Theory (ALT), assuming the reference for categorizing a speech sound is external. In the present experiment, identification functions were obtained for a series of synthetic speech sounds ranging perceptually from /ba/ to /pa/ and a series of tones varying in intensity from 60 to 84 dB SPL. The distribution of occurrences of each stimulus in a series was varied for both tones and speech stimuli. The category boundaries for the tones shifted as a function of the relative number of occurrences of each tone as predicted by ALT. However, the phonetic boundaries for the speech stimuli failed to show the analogous shift. These results suggest that the response criteria for phonetic boundaries may be mediated by an internally-generated reference. In contrast, the reference for the nonspeech category boundaries appears to be under external stimulus control.

*This is a draft of a paper to be presented at the 86th meeting of the Acoustical Society of America, November 1, 1973, Los Angeles, California. The research was supported in part by PHS grants SO5 RR 7031, MH 24027, and T01 MH 11219-04 to Indiana University. We wish to thank Dr. Franklin S. Cooper for making the facilities of Haskins Laboratories available to us for the preparation of the stimulus materials and Professor Frank Restle for his interest in this work.

Category Boundaries for Speech and Nonspeech Sounds

J. R. Sawusch and D. B. Pisoni

Indiana University

In recent years a large body of evidence has been accumulated to suggest that the perception of speech sounds may be quite different from the perception of other auditory stimuli. Although several theories of speech perception have been proposed, they are for the most part quite vague and general and it is relatively difficult to derive any specific predictions that are testable. In the present study we wish to consider the nature of category judgments and specifically the nature of the boundaries between categories for speech and nonspeech sounds. We have chosen this particular problem to study primarily because we feel that two of the current theories of speech perception, the Haskin's Motor Theory and Steven's Quantal Theory, would make specific predictions about the nature of category boundaries for certain classes of speech sounds.

We may think of two relatively broad views of the nature of category boundaries for speech sounds. One view is that the category boundaries between phonetic segments are arbitrary in the sense that they are simply the consequence of a psychophysical partitioning of a stimulus continuum into equivalent response categories. In contrast, an alternative view and one which could be predicted from either speech theory is that the boundaries between phonetic segments are not arbitrary. Rather, due to constraints on the articulatory mechanism and the resultant changes in the acoustic signal, the boundaries between segments may be relatively fixed.

If the perceptual boundaries between phonetic segments are arbitrary in the sense of a simple psychophysical partitioning of the stimulus continuum, then it should be relatively easy to produce systematic changes in the location of the boundary by manipulations of the probabilities of occurrence of different stimuli. For example, Adaptation Level Theory, which has been used extensively in psychology to account for changes in the judgment of brightness, hue, loudness, and pitch, could be applicable to the judgment of phonetic segments.

Slide 1 please

Let us consider such a prediction in detail. Slide 1 shows an idealized identification function for a two category absolute identification task. In the control condition, each stimulus occurs with an equal probability and the subject partitions the continuum into two equivalent categories. When the probabilities are unbalanced and, for example, stimulus number one occurs more often than any of the other stimuli, the boundary should shift toward the more frequently occurring stimulus, or anchor. The same effect should be obtained when stimulus seven occurs more often than any of the other stimuli; the category boundary should shift toward that stimulus. Adaptation Level Theory would predict these results for both speech and nonspeech continua since it is assumed that the standard or reference used to categorize a particular stimulus is for the most part under the control of external stimuli.

In the present study, we were concerned with the effect of unbalanced probabilities of occurrence of stimuli on the identification of speech and

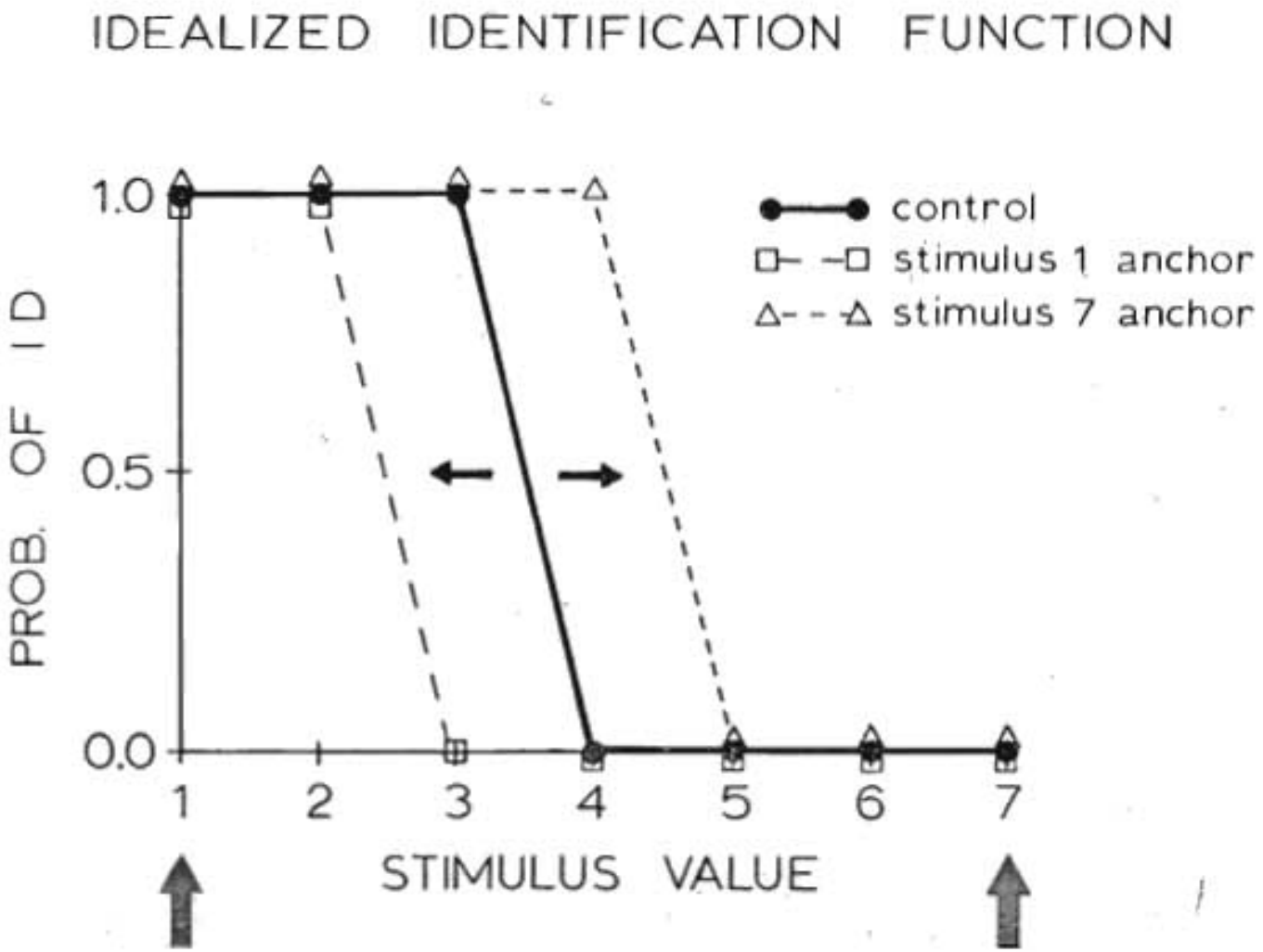
IDEALIZED IDENTIFICATION FUNCTION
1.0
0.5
0.0
PROB. OF I D
control
stimulus 1 anchor
stimulus 7 anchor
1
2
3
4
5
6
7
STIMULUS VALUE

Figure 1.

nonspeech continua. If the boundaries between phonetic segments are arbitrary and under the control of external stimuli we should expect to find a shift in the identification function as the relative frequency of occurrence of the end or anchor stimuli are changed. On the other hand, if we do not find a shift in the identification function for the speech continuum but we do find one for a nonspeech continuum, we may conclude, at the very least, that the boundaries between phonetic segments are not simply due to an arbitrary partitioning of a stimulus continuum.

Method

Two sets of stimuli were used in this experiment, a speech continuum and a nonspeech continuum. The speech stimuli consisted of a set of seven three-formant patterns ranging perceptually from /ba/ through /pa/ and were produced on the speech synthesizer at Haskins Laboratories. The seven stimuli varied in 10 msec steps along the voice onset time continuum from 0 msec VOT to +60 msec VOT. The stimuli were recorded on magnetic tape in random order to produce three identification tests. In the control tape, each stimulus occurred equally often. In the /ba/ anchor tape, the stimulus with 0 msec VOT occurred twice as often as each of the other six stimuli. In the /pa/ anchor tape, the stimulus with +60 msec VOT occurred twice as often as each of the other stimuli.

The nonspeech continuum consisted of a set of seven tones varying in intensity. The tones varied in 4 dB steps from 60 dB to 84 dB. These stimuli were recorded in random order on magnetic tape to produce three analogous identification tests: a control tape, a loud anchor tape, and a soft anchor tape.

Subjects listened to four different tapes: a speech control, a tone

control, a speech anchor, and a tone anchor. In the speech condition subjects were told to identify each stimulus as a /ba/ or a /pa/. In the nonspeech condition subjects were told to identify each tone as loud or soft.

Results and Discussion

Slide 2 please

Slide 2 shows the average identification functions for the nonspeech condition. In Group I shown on the left subjects heard the loud anchor tape. The identification function shows a consistent shift toward the loud stimulus, relative to the control identification function. In Group II shown on the right subjects heard the soft anchor tape. This identification function shows a shift but this time it is toward the soft stimulus, relative to the control function. Both shifts, which would be predicted by Adaptation Level Theory, reveal how the loud-soft judgment is arbitrary in the sense that it is under the control of the stimuli occurring during the test. Now let us turn to the speech data for comparison.

Slide 3 please

Slide 3 shows the average identification functions for the /ba/-/pa/ continuum for the same subjects. Group I heard the /ba/ anchor tape. The identification function shown on the left shows no shift relative to the control condition. The same is true for the /pa/ anchor group. There is no shift relative to the control tape. If the speech stimuli were evaluated

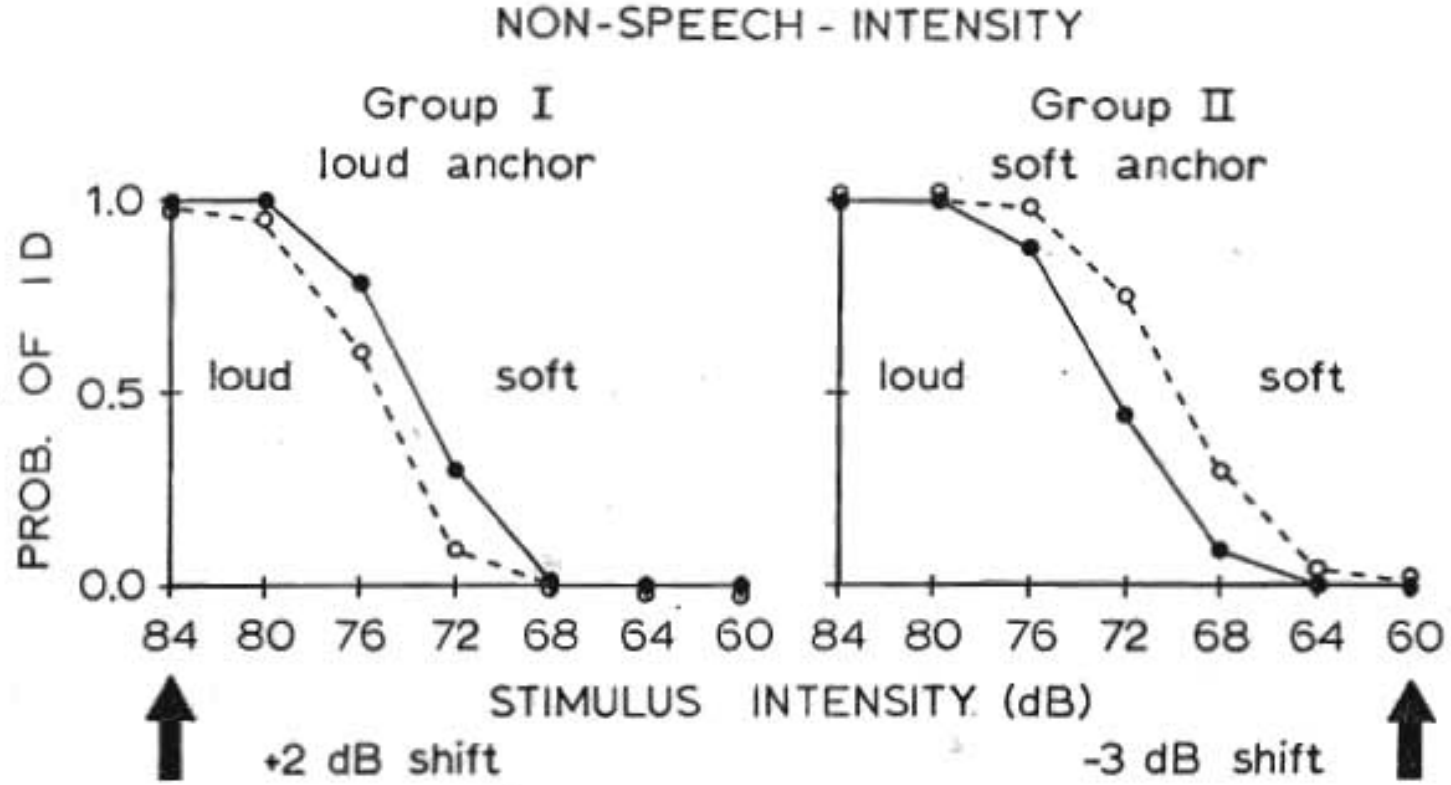
NON-SPEECH - INTENSITY
Group I
loud anchor
Group II
soft anchor
PROB. OF I D
1.0
0.5
0.0
loud
soft
loud
soft
84 80 76 72 68 64 60
84 80 76 72 68 64 60
STIMULUS INTENSITY (dB)
+2 dB shift
-3 dB shift

Figure 2.

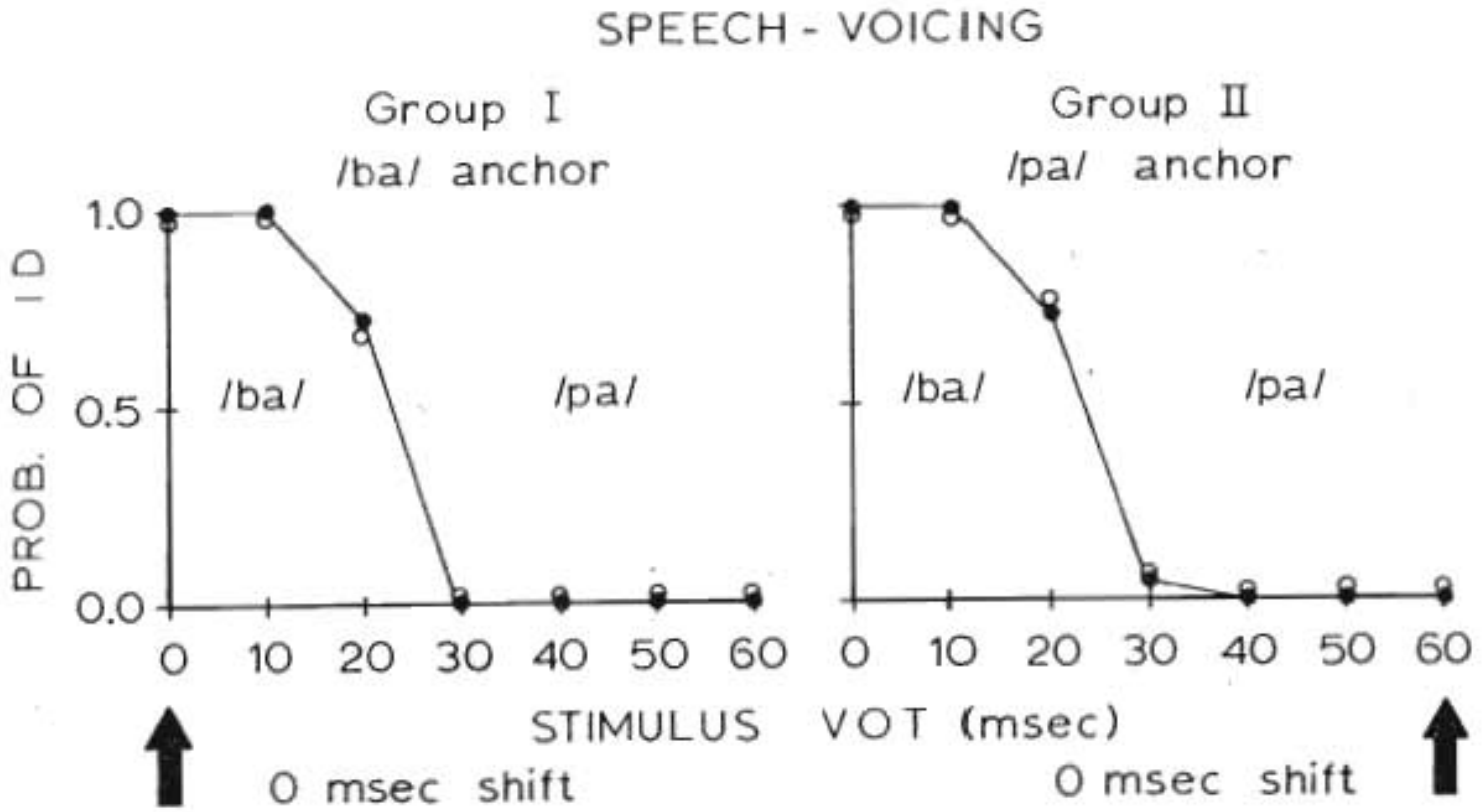
SPEECH - VOICING
Group I
/ba/ anchor
Group II
/pa/ anchor
PROB. OF I D
1.0
0.5
0.0
/ba/
/pa/
/ba/
/pa/
0 10 20 30 40 50 60
0 10 20 30 40 50 60
STIMULUS VOT (msec)
0 msec shift
0 msec shift

Figure 3.

according to the same criterion as the tones we would have expected a shift for speech anchor conditions relative to the control. Surprisingly, no shift occurred for the speech.

Experiments similar in format to these but using synthetic stop consonants varying in place of production and tones varying in frequency have also been run. The results indicate that the category boundary between hi and low frequency tones shifts as predicted by ALT. The category boundary for the place continuum between /bae/ and /dae/ also failed to show any shift.

The view that phonetic categories are arbitrary and simply the result of a psychophysical partitioning of a stimulus continuum seems to be an inadequate explanation of our results. If this were in fact the case we would have expected the speech anchor conditions to show a shift analogous to that found in the tone anchor conditions. It seems that the category boundaries at least for place and voicing in stop consonants result from the application of relatively stable and non-arbitrary criteria. We suggest that these criteria may be mediated more by an internally generated reference than an externally calculated standard.

To summarize, the effect of unbalanced probabilities of occurrence of stimuli produced a shift in the category boundary for a nonspeech continuum but failed to produce a parallel shift in a speech continuum. These results suggest an internal, highly stable and non-arbitrary criteria for the categorization of phonetic segments.