Bulletin of the Psychonomic Society
1981, Vol. 18(1), 12-14

Perceptual equivalence of two kinds of ambiguous speech stimuli

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut 06510

Stimuli from two synthetic /da/-/ga/ continua were presented in a speeded labeling task. One continuum was generated by parameter interpolation; the other was generated by adding the waveforms of the endpoint stimuli in varying proportions. Both continua showed an increase in latencies at the category boundary, suggesting that the two procedures yield equally ambiguous stimuli.

Ambiguous stimuli play a central role in speech perception research. By virtue of their perceptual instability, they serve as indicators of a large variety of laboratory phenomena, including categorical perception, selective adaptation, phonetic trading relations, and all sorts of context effects. Traditionally, ambiguous stimuli have been constructed with the aid of speech synthesizers: Two unambiguous stimuli from different phonetic categories are selected, and a number of steps are interpolated between their parameter values, leading to a continuum that includes some ambiguous stimuli in the region of the phonetic category boundary. Until recently, this was the only method available. However, a new technique was applied in a recent doctoral thesis by Stevenson (1979). Instead of interpolating parameter values between two endpoint stimuli, he added the digitized waveforms of the endpoint stimuli in various proportions, increasing the amplitude of one component waveform while decreasing that of the other, and so producing a continuum. In fact, he was able to construct such continua from carefully aligned natural utterances of /ba/, /da/, and /ga/; but the technique can, of course, be used with synthetic speech as well.

Electronically mixed synthetic stimuli have been used previously, primarily to compare their perception with that of the same component stimuli presented dichotically (Halwes, 1969; Porter & Whittaker, 1980; Repp, 1976, 1980). However, Stevenson (1979) was apparently the first to construct whole stimulus continua that way. His technique is interesting, especially because it can be used with natural speech. However, are there any important perceptual differences between an ambiguous stimulus created by superimposing two unambiguous stimuli and one characterized by a single set of intermediate parameters? Stevenson used his stimuli in a variety of standard experimental tasks, including categorical perception, selective adaptation, and dichotic listening, and he obtained results very similar to those found with traditional stimulus continua, although he never performed any direct comparison.[1]

The present study explored one way in which the two types of ambiguous speech stimuli might differ in perception. When presented with an ambiguous stimulus of the traditional kind, which has acoustic properties that are truly intermediate, listeners experience uncertainty that increases the time needed to assign the stimulus to one of two categories (Pisoni & Tash, 1974; Studdert-Kennedy, Liberman, & Stevens, 1963). However, when listening to a stimulus from a Stevenson (1979) continuum, which contains two unambiguous sets of cues superimposed, there might be no uncertainty on a given trial; rather, perception might go with one or the other set of unambiguous cues on a probabilistic basis. The present study tested this hypothesis by examining whether the characteristic peak in identification latencies at the category boundary of traditional speech continua (Pisoni & Tash, 1974; Studdert-Kennedy et al., 1963) is present to the same extent on a continuum of electronically mixed stimuli.

METHOD

Subjects

Eight paid student volunteers participated. They had little or no experience in experiments of this kind.

Stimuli

The syllables /da/ and /ga/ were synthesized on the OVE IIIc synthesizer at Haskins Laboratories. They were distinguished only by the third-formant (F3) transition, whose onset frequency was 2,976 Hz in /da/ and 2,150 Hz in /ga/. All other characteristics were shared: fully periodic waveform, a duration of 250 msec, a fundamental frequency that fell linearly from 110 to 80 Hz, 50-msec linear formant transitions, F1 rising from 285 to 771 Hz, F2 falling from 1,770 to 1,233 Hz, and an F3 steady-state frequency of 2,520 Hz.

The mixed (Stevenson-style) continuum was constructed in the following way: The two syllables were digitized at 10 kHz, using the Haskins Laboratories PCM system. Nine intermediate stimuli were obtained by adding the /da/ and /ga/ waveforms point by point after reducing the amplitude of each by a certain amount. That amount was determined by translating the ratios 1:9, 2:8, . . ., 8:2, 9:1 into decibel values, under the constraint

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I am grateful to Dale Stevenson for bringing his work to my attention and for commenting on an earlier version of this paper.

that the amplitude of the combined waveforms remain constant. The resulting attenuation values were −1, −2, −3, −5, −6, −8, −10, −14, and −20 dB SPL for the /da/ component; they applied in inverse order to the /ga/ component.[2] Only these nine stimuli were used in the experiment.

The interpolated (traditional) continuum was constructed by synthesizing eight intermediate stimuli between /da/ and /ga/, changing the onset frequency of F3 in equal decrements. All 10 stimuli were digitized at 10 kHz. To control for any possible artifacts due to waveform addition on the other continuum, and to match the numbers of stimuli on the two continua, the 10 stimuli were reduced to 9 by adding the waveforms of neighbors on the continuum. Stimulus amplitudes were first reduced by 6 dB SPL, to match the amplitudes of the stimuli on the mixed continuum.

Randomized stimulus sequences were recorded on tape. The stimuli from both continua were randomized together to yield a basic unit of 18 stimuli. Five such units formed one continuous block of 90 stimuli, with interstimulus intervals of 2 sec. Four such blocks were recorded, with longer pauses in between. Each block was prefixed with four warm-up stimuli that were not scored. At the very beginning of the tape was a practice sequence of 40 stimuli containing only instances of the endpoint stimuli of the two continua. To the author, the stimuli from the two continua were phenomenally indistinguishable.

Procedure

Subjects were tested individually in a soundproof booth. They sat in front of a table and rested their index fingers on two telegraph keys labeled "dah" and "gah." The response-to-keys assignment was counterbalanced across subjects. The instructions stressed speed of response. The subjects were permitted to stop the tape recorder by remote control between blocks and take a rest, if they desired. The tape was played back on a Crown 800 tape recorder located in an adjacent room, and the subject listened over Telephonics TDH-39 earphones. Reaction times were measured by a Hewlett-Packard 5302A 50-MHz universal counter and printed out by a Hewlett-Packard 5150A thermal printer. The counter was triggered by a signal recorded on the second tape channel and synchronized with syllable onset.

RESULTS AND DISCUSSION

The results, averaged over subjects, are displayed in Figure 1. It can be seen that the labeling functions for the two continua were virtually identical, and so were the latency functions. The perhaps fortuitous coincidence of the category boundaries[3] is less important than the fact that both latency functions exhibited peaks of equal magnitude at the category boundary. Analysis of variance confirmed a significant effect of stimulus number [F(8,56) = 2.85, $p < .01$], but no significant effect involving type of continuum.

Thus, the two kinds of continua were perceptually equivalent in this speeded labeling task. In particular, stimuli from the centers of the two continua were equally ambiguous and created equal uncertainty in listeners. This tells us something about the perceptual processing of mixed stimuli. Apparently, it is not the case that the superimposed conflicting cues are accessed individually by some selective attention mechanism (as perhaps suggested by the concept of auditory "listening bands"; Divenyi, 1979) or subject to mutual lateral inhibition or masking. Rather, conflicting transitions of the same formant seem to engage in a "trading relation," just as transitions of different formants do (see Mattingly & Levitt, 1980, for a recent study). The outcome of this tradeoff appears to be perceptually equivalent to an acoustically intermediate specification, at least as far as phonetic perception is concerned. Stevenson's (1979) extensive data obtained with electronically mixed stimuli suggest that they are equivalent to traditional stimuli in many other respects. It seems unlikely, then, that the new technique of stimulus construction will lead to any new insights about the mechanisms of speech perception, although it deserves continued attention because of its applicability to natural speech.

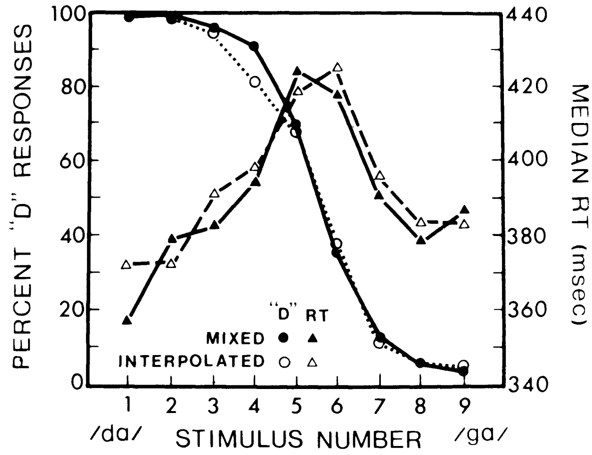

Figure 1. Identification ("d" responses) and latency functions for two kinds of /da/-/ga/ continua.

Several limitations of Stevenson's (1979) method should be pointed out, however. First, it can be used only with stimuli of similar temporal structure; that is, it is restricted primarily to variations in spectral cues (see also Footnote 1). Second, it does not work with stimuli that do not readily fuse into a single percept, such as vowels (Stevenson, 1979). The factors at work here seem to be very similar to those governing dichotic fusion (cf. Cutting, 1976). Third, mixed continua have the property that stimuli become increasingly less discriminable (on purely auditory grounds) the farther they are from the center of the continuum, which is undesirable in categorical-perception experiments, in which the detectability of within-category differences is of prime interest. Therefore, it appears that Stevenson's technique will be useful only under very special circumstances.[4]

REFERENCES

Cutting, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review,* 1976, **83**, 114-140.

Divenyi, P. L. Some psychoacoustic factors in phonetic analysis. *Proceedings of the Ninth International Congress of Phonetic Sciences,* 1979, **2**, 445-452.

Halwes, T. G. *Effects of dichotic fusion on the perception of speech.* Unpublished doctoral dissertation, University of Minnesota, 1969.

Mattingly, I. G., & Levitt, A. G. Perception of stop consonants before low unrounded vowels. *Haskins Laboratories Status Report on Speech Research,* 1980, **SR-61,** 167-173.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics,* 1974, **15,** 285-290.

Porter, R. J., Jr., & Whittaker, R. G. Dichotic and monotic masking of CV's by CV second formants with different transition starting values. *Journal of the Acoustical Society of America,* 1980, **67,** 1772-1780.

Repp, B. H. Identification of dichotic fusions. *Journal of the Acoustical Society of America,* 1976, **60,** 456-469.

Repp, B. H. Stimulus dominance in fused dichotic syllables: Trouble for the category goodness hypothesis. *Journal of the Acoustical Society of America,* 1980, **67,** 288-305.

Stevenson, D. C. *Categorical perception and selective adaptation phenomena in speech.* Unpublished doctoral dissertation, University of Alberta, 1979.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America,* 1963, **35,** 1900. (Abstract)

NOTES

1. Stevenson (1979) drew an analogy between his ambiguous stimuli and certain ambiguous visual figures, such as the Necker cube: A continuum can be constructed by beginning with an unambiguous drawing of Orientation A of the (opaque) cube and by then slowly increasing the intensity of the added line segments unique to Orientation B while decreasing the intensity of the line segments unique to A until only B remains. At the center of the continuum, where all lines are equally intense, we have the maximally ambiguous figure: the (transparent) Necker cube. It is interesting to note that this visual analogy is not appropriate for the traditional method of constructing speech continua; if applied to the cube drawings, that method would use spatial interpolation between lines unique to the two endpoint stimuli, resulting in curvilinear distortions that destroy the identity and three-dimensionality of the cube. However, the interpolation technique could be used to construct a continuum from, say, a circle to a square, whereas Stevenson's method would fail here because intermediate stages would be seen as a square superimposed on a circle, not as one or the other. Apparently, the endpoint stimuli must have a rather special relation to each other if both methods shall result in truly ambiguous stimuli. It appears that this condition is satisfied only by certain speech stimuli, such as stop-consonant-vowel syllables differing in (stop) place of articulation.

2. Since only integer decibel values could be used on the computer, overall amplitude varied over a range of .5 dB SPL. Also, the calculated values strictly apply only to perfectly correlated waveforms (cf. Stevenson, 1979). However, since the present stimuli differed only in F3, and only during the first 50 msec, the values used were quite adequate.

3. The author, as a pilot subject, had different boundaries on the two continua. No claim is being made here that the two continua constitute equivalent perceptual scales (i.e., that there is a one-to-one equivalence of stimuli).

4. This conclusion is not intended as a critique of Stevenson, whose careful and sophisticated (but, unfortunately, unpublished) work made a valuable methodological contribution.

(Received for publication June 9, 1981.)