Psychological Review
1970, Vol. 77, No. 3, 234–249

THEORETICAL NOTES

MOTOR THEORY OF SPEECH PERCEPTION: A REPLY TO LANE'S CRITICAL REVIEW[1]

MICHAEL STUDDERT-KENNEDY,[2] ALVIN M. LIBERMAN,[3] KATHERINE S. HARRIS[4]
AND FRANKLIN S. COOPER

Haskins Laboratory, New Haven, Connecticut

In reviewing data that bear on a motor theory of speech perception, Lane criticized experiments that had demonstrated contrasting tendencies toward "categorical" perception of stop consonants and "continuous" perception of vowels and nonspeech sounds. He also undertook to demonstrate that categorical perception of nonspeech sounds can be produced by the ordinary procedures of discrimination training, and so to refute the claim that such perception is an interesting characteristic of the speech mode. The purpose of the reply is to correct the serious misunderstandings which Lane's criticisms reveal, and to show, contrary to his claim, that discrimination training is not sufficient to produce categorical perception.

Lane (1965) reviewed data that may be interpreted as supporting a motor theory of speech perception, and challenged that interpretation. His own conclusion and that of his colleagues was that "the postulation of a special perceptual mechanism for speech perception is not warranted" (Cross, Lane, & Sheppard, 1965).

First, we should make clear that the evidence for a special perceptual mechanism goes well beyond the data reviewed by Lane (1965). We have presented that evidence elsewhere (Liberman, Cooper, Harris, MacNeilage, & Studdert-Kennedy, 1968; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Mattingly & Liberman, 1970; Studdert-Kennedy & Shankweiler, 1970). In general, it shows speech perception to be markedly different from other auditory perception and suggests that special speech processors may exist. Furthermore, it suggests that these processors may have access to mechanisms that are in some sense motor. Were every objection raised by Lane accepted, this other evidence would still remain, and a special theory of speech perception—in our view, a motor theory—would still be called for. The purpose here, however, is not to review that evidence, nor to defend the general hypothesis that the processes of speech perception differ in interesting ways from those of nonspeech perception. We are concerned, rather, to set the record straight about a mode of perception—we have called it "categorical"—that characterizes the way some phonetic segments are heard. Lane devoted a large part of his review to criticisms of experiments that deal with this type of perception. The purpose of this paper is to answer that criticism and to deal with certain misunderstandings on which it was based.

CATEGORICAL PERCEPTION

As we have used the term, categorical perception refers to a mode by which stimuli are responded to, and *can only be responded to, in absolute terms.* Successive stimuli drawn from a physical continuum are not perceived as forming a continuum, but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur. Subjects asked to discriminate between pairs of such "categorical" stimuli are able to discriminate between stimuli drawn from different categories, but not between stimuli drawn from the same category. In other words, discrimination is limited by identification: subjects can only discriminate between stimuli that they identify differently.

---

[1] Preparation of this paper was supported by grants from the National Institute of Child Health and Human Development and the National Institute of Dental Research, and under contract from the Office of Naval Research.

[2] Also at Queen's College, City University of New York. Requests for reprints should be sent to Michael Studdert-Kennedy, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

[3] Also at the University of Connecticut and Yale University.

[4] Also at City University of New York.

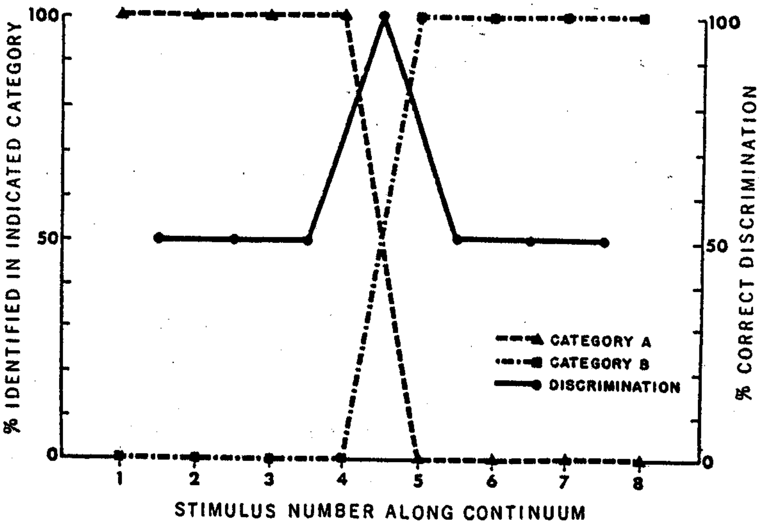

FIG. 1. Idealized identification functions (left ordinate) and discrimination function (right ordinate) to illustrate categorical perception of eight stimuli distributed at equal intervals along a physical continuum.

Figure 1 illustrates an ideal case. Eight stimuli are spaced at equal physical intervals along a continuum. Stimuli 1–4 are identified as members of Category A, Stimuli 5–8 as members of Category B (left ordinate). Identifications are completely consistent (i.e., there is no context effect), so that the boundary between categories is clear-cut. If the stimuli are presented in successive order, subjects perceive no change at all between Stimuli 1 and 4; between Stimuli 4 and 5 they perceive an abrupt change from Category A to Category B; and they again perceive no change at all between Stimuli 5 and 8.[5]

A further behavioral measure of this perceptual classification is provided by subjects' performance when asked to discriminate between these stimuli on any basis whatever. If we were to measure discrimination by an ABX procedure, for example, we should obtain results like those shown in the figure. On the right ordinate is plotted the percentage of correct discrimination in ABX judgments of neighboring stimuli: performance is at chance level for all pairs drawn from the same identification category, but is perfect for the pair (Stimuli 4 and 5) drawn from different categories. We have a peak in the discrimination function at the boundary between categories and (we wish to emphasize this) a trough in the function within categories. Discriminative performance may be perfectly predicted from identification performance: the subject, asked to discriminate, is able, in fact, only to identify and to base his supposed discriminative judgments on these identifications.

We may contrast this ideal case of categorical perception with the ideal case of "continuous" perception, typically approximated in psychophysical studies: successive stimuli drawn from a physical continuum are, in fact, perceived as drawn from a continuum. Subjects may learn to group stimuli into discrete categories, although they are likely to experience some difficulty in doing this, particularly with stimuli that lie on a border between categories and may be shifted across category borders by context effects. But if stimuli are widely enough spaced along the continuum, subjects may learn to identify them as consistently as those illustrated in Figure 1. However, if the stimuli are presented in order, subjects perceive progressive changes corresponding to progressive changes in the stimuli. The sharp discontinuities of categorical perception are not encountered. Called on to discriminate between pairs of these "continuous" stimuli, subjects discriminate equally well across the en-

---

[5] Something close to this is perceived if we make physically equal changes in the acoustic cue that underlies the perception of the voiced stop consonants /b,d,g/: the listener hears an almost quantal change from one phoneme to the next, not step-by-step changes to match the step-by-step progression in the acoustic stimulus.

tire continuum: that is to say, they discriminate just as well between stimuli drawn from the same category as between stimuli drawn from different categories. Discriminative performance cannot be perfectly predicted from identification performance: the process of discrimination is independent of the process of identification.

In sum, categorically perceived stimuli *can only be perceived absolutely.* To identify them differently is to discriminate between them and to discriminate between them is to identify them differently. Continuously perceived stimuli, on the other hand, are perceived relationally, and even if placed in classes (categories), may be perceived as different. To discriminate between them is to detect a difference whether or not they are identified as belonging in the same class.

From the foregoing it should be clear that the degree to which perception is categorical or continuous can only be unequivocally determined from an examination of both identification and discrimination data. Identification functions alone are not enough: clear-cut categories are characteristic of categorically perceived stimuli, but may also be yielded by any continuum of which the stimuli selected for testing lie far enough apart. Discriminative peaks and troughs alone are not enough: they must be appropriately correlated in level and position along the continuum with identification categories, the peaks occurring at category boundaries, the troughs within categories.

What is needed, therefore, is a procedure for comparing identification and discrimination functions and determining the degree to which they are related. This may be done succinctly in a single test. On the hypothesis that the stimuli are categorical, the probability that two stimuli will be correctly discriminated is equal to the probability that they will be identified as members of different classes, and the discrimination function may be predicted from the identification function. The predicted function may then be compared with the obtained and a decision made as to how well the hypothesis is supported. In practice, the hypothesis is seldom fully supported: the obtained function almost always lies somewhat above the predicted, indicating that there remains some basis for discrimination, however marginal, between stimuli that are placed in the same category.

Nonetheless, the deviation of the obtained from the predicted is so much greater for some continua than for others that we are justified in maintaining a distinction. In particular, we have found that certain speech sounds (stop consonants) tend to be perceived categorically, while certain other speech (vowels) and nonspeech sounds tend to be perceived continuously. This is a fact that, taken together with a variety of other evidence, we believe to be of some importance to a theory of speech perception.

LANE'S OBJECTIONS

On pages 291–305 of his review, Lane considers alleged methodological limitations of the Haskins work and presents counterevidence intended to show that the

> reported [by Haskins] relations between identification and discrimination functions are not at all unique to the perception of consonants but describe as well the perception of vowels and of entirely nonlinguistic stimuli . . . [p. 292].

Methodological Limitations (pp. 292–294)

Subject selection. Lane cites three Haskins studies (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961) in which subjects were selected, and presents three objections to them: (*a*) the criteria for selection were "neither explicit, quantitative, nor uniform"; (*b*) the procedure "restricts the generality of inferences based on the findings"; (*c*) the procedure "vitiates comparisons among different experiments."

The purpose of these experiments was not to collect normative data on how synthetic speech stimuli are perceived by the population in general, but to see how subjects who categorized members of a set of stimuli discriminated between them. It was necessary, therefore, to use subjects who did categorize the stimuli. Since synthetic speech stimuli are often noisy and not well categorized, subjects had, in some studies, to be selected.[6] We then chose those subjects

---

[6] Difficulties of this kind are familiar to Lane. In an experiment with Cross (Cross & Lane, 1962) on the discriminative control of vocal responses, he selected 3 out of 12 subjects; in an experiment with Schneider (Lane & Schneider, 1963), he selected 3 out of 8; in another experiment with Cross (Cross & Lane, 1964), he based his conclusions on the "only one [of 5 subjects] who yielded identification functions that were even nearly categorical." His subject rejection rate in these three experiments is more than 70%—as compared with the 50% that he finds offensive in the three Haskins studies. We cite these examples not to suggest that Lane's selection of subjects "re-

whose behavior was most consistent. The explicitly stated criterion was that subjects' identification functions be clear-cut: what this means may be seen, in two of the three studies to which Lane objects, by referring to the individual data. It is true that we did not define "clear-cut" quantitatively. But had we done so, we would still not have been able to hold to the criterion across experiments, since synthetic speech stimuli vary considerably in quality. The best that could be done was to select, for a given group of stimuli, the most consistent subjects from a reasonably large sample. In this limited sense—limited, it should be remarked, by the nature of the experiments—the criterion was uniform across experiments and permits comparisons between them. For fuller discussion of this point, see below under Lane's criticisms of the prediction test.

> *Averaging data.* The procedure of averaging individual identification and discrimination functions and presenting group data, employed in all the synthetic speech studies cited earlier but two, may be placed alongside preselection of subjects in vitiating comparisons between the findings of experiments when these comparisons are based directly or indirectly on the labeling gradients [Lane, 1965, p. 293].

Of the 15 Haskins studies that Lane has cited up to this point in his review, 12 do not present identification and discrimination functions, two present individual data, and one does indeed present averaged functions. This last is presumably Lane's "all the . . . studies cited earlier but two. . . ." Apart from this, averaging data certainly does not invalidate (if this is what Lane means by "vitiate") comparisons between phonemic boundaries and discrimination peaks, or between predicted and obtained discrimination functions. These are the purposes to which the data of this study were put.[7] We also discuss this question more fully below under Lane's criticisms of the prediction test.

Comparisons of identification functions. Lane objects to comparisons between identification functions ("labeling gradients") for different stimulus continua, such as consonants and vowels, since "their abscissae are not commensurate." He is right: the abscissae are not physically commensurate, and no Haskins paper has ever suggested that they were. What has been said is that certain consonant and vowel continua are linguistically comparable. For example, Fry, Abramson, Eimas, and Liberman (1962) say,

> the comparison between the vowel data and the (consonant) results . . . is a reasonable one. In both cases there were three phoneme classes: these three classes were divided into fourteen steps in the case of the stops and thirteen in the case of the vowels [p. 182].

In other words, the two continua were *behaviorally* equivalent: they were defined, as all linguistic continua necessarily are defined, by listeners' judgments. Within these limits, Haskins papers have commented that consonant identification functions tend to be sharper (i.e., more consistent, less liable to context effects) than vowel identification functions.

Evidence that Opposes the Motor Theory of Speech Perception

Identification functions. Lane reports two sets of identification functions for nonspeech continua to show that such functions may be as sharp as the corresponding functions for consonant continua. On pages 294–296 he gives the results of Lane and Schneider (1963) with the /do/–/to/ control stimuli; on page 299 he gives the results of Cross et al. (1965) with visual patterns (sectored circles). He also refers to several Haskins studies of vowel identification (Bastian & Abramson, 1962; Fry et al., 1962; Stevens, Ohman, & Liberman, 1963; Studdert-Kennedy, Liberman, & Stevens, 1963), in which clearly defined phoneme boundaries were observed. He concludes that the "form of labeling functions for consonant continua does not differ from that for vowel and nonlinguistic continua [p. 294]."

If we substitute "may not differ" for "does not differ" in this statement, it is one with which we fully agree. We cannot agree, however, with the implication that the degree of categorical perception does not differ for these several types of continua, since a sharp identification function is not a sufficient condition of categorical perception. As we remarked earlier in this reply, subjects may learn to identify stimuli consistently if the stimuli are spaced widely enough along the continuum. That the stimuli of Cross et al. (1965) were, in fact, spaced widely apart is evident from the high

---

stricts the generality" of his inferences or "vitiates comparisons" between his work and others, but to remind the reader that even the most rigorous experimenter may be forced to select subjects who meet the requirements of his experiment.

[7] We take it that Lane would approve such comparisons, since he himself has several times presented pooled discrimination data, both predicted and obtained (e.g., Cross & Lane, 1964, Experiments I and III; Lane & Schneider, 1963).

level of discrimination in their one-step discrimination test (Lane, 1965, Figure 12, p. 299). Similarly, the wide spacing of the synthetic vowel stimuli in the Haskins studies cited by Lane is shown by a generally high level of discrimination. For the /do/–/to/ control stimuli (Lane & Schneider, 1963; Liberman, Harris, Kinney, & Lane, 1961), spacing does not appear to have been wide, and we return to discussion of this experiment below.

Discrimination functions (pp. 296–300). Lane points out that peaks in discrimination at boundaries between identification classes have been observed for a vowel continuum (Stevens et al., 1963; see also Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969), a degraded vowel continuum (Cross & Lane, 1964), two nonspeech auditory continua (Cross et al., 1965; Lane & Schneider, 1963). and wavelength (Beare, 1963; Ekman, 1963). He concludes,

> Since the correlation between discrimination and labeling is observed for nonspeech as well as for speech stimuli, the correlation per se does not seem to warrant the postulation of a special perceptual mechanism for the discrimination of speech stimuli [p. 300].

We have four general comments, followed by some specific criticisms of Lane's experiments. First, we do not deny that discriminative peaks may be observed at phoneme boundaries along vowel continua—indeed, it was in a study with which we were associated that such peaks were first reported (Stevens et al., 1963). We do deny that troughs *within* vowel phoneme categories are as deep as would be expected if vowel perception were truly categorical: typically, discriminative performance within categories is well above the level (close to chance) predicted from identification performance. Lane, focusing his attention on the peaks that are present, seems not to have noticed that the equally important troughs are absent.

Second, we would no longer wish to argue for an absolute distinction between vowel and consonant perception as in some of the authors' earlier papers (e.g., see Lane's quotation from Liberman, Harris, Eimas, Lisker, & Bastian, 1961, p. 297). We would maintain rather that there is a difference in the *degree* to which consonants and vowels are categorically or continuously perceived, a difference that may be shown by application of the predictive test.

Third, we cannot agree that peaks in the hue discrimination function represent a significant approximation to categorical perception. We know that many hues can be discriminated within each of the color-name classes, so that even though discriminative peaks may be present, the required troughs are clearly absent.

Fourth, a specialized perceptual mechanism for speech is not postulated on the basis of this "correlation per se," any more than is a specialized perceptual mechanism for color. For speech, as for color, a good deal of other evidence suggests that perception may be a specialized process. We cannot accept Lane's implication that discriminative peaks at color-class boundaries along the wavelength continuum and at phoneme boundaries along some acoustic continuum are adequate evidence for the operation of the same processes in color and speech perception.

We turn now to detailed criticism of Lane's two nonspeech studies: the /do/–/to/ control study (Lane & Schneider, 1963) and the sectored circle study (Cross et al., 1965).

The /do/–/to/ control study. The procedure and results of this experiment are reported on pages 294–295 and pages 298–299 of Lane's review. The stimuli (described in detail by Liberman. Harris, Kinney, & Lane, 1961) were seven auditory patterns obtained by inverting the /do/–/to/ spectrograms before converting them to sound. The inversion preserves the stimulus variable of first-formant cutback (ranging from 0 to 60 milliseconds (msec.) in 10-msec. steps) by which /do/ may be discriminated from /to/, but destroys the resemblance of the sounds to speech. The purposes of the experiment were to show that (*a*) subjects could learn to identify the extremes of the nonspeech continuum as /do/ and /to/, and to give clear-cut identification functions; (*b*) subjects, displaying chance levels of discriminative performance before training, would develop discriminative peaks at their "phoneme boundary" after training; (*c*) discriminative performance after training could be accurately predicted from identification functions. If this program were fulfilled, it would show that a stimulus variable categorically perceived in speech may, after brief training, be categorically perceived in nonspeech. and the grounds for supposing categorical perception to be anything more than a general result of discrimination training would be destroyed.

How far, then, were the purposes of the experiment achieved? 1. Three out of eight subjects reached the experimental criterion of 30 successive correct identification responses in less than 400 training trials with the two extreme stimuli (0- and 60-msec. cutback), and

TABLE 1

TWO-STEP AND THREE-STEP COMPARISONS

Subject	Stimuli grouped by "phoneme classes"	Same class	N	% correct	Different classes	N	% correct	Different minus same % correct
				Two-step comparisons				
1	01; 23456	2–4, 3–5, 4–6	60[a]	75[b]	0–2, 1–3,	53.5[a]	79[b]	4
2	012; 3456	0–2, 3–5, 4–6	72	61[c]	1–3, 2–4,	48	84[c]	23
3	0123; 456	0–2, 1–3, 4–6	60[a]	75[b]	2–4, 3–5	53.5[a]	79[b]	4
				Three-step comparisons				
1	01; 23456	2–5, 3–6	44.7[a]	63[b]	0–3, 1–4	58.8[a]	85[b]	22
2	012; 3456	3–6	24	88[c]	0–3, 1–4, 2–5	72	87[c]	−1
3	0123; 456	0–3	22.3[a]	63[b]	1–4, 2–5, 3–6	88.2[a]	85[b]	22

[a] The Ns are the weighted average numbers of ABX pairs for Subjects 1 and 3 after the N for Subject 2 (24 comparisons per pair) has been stubracted from the total N given in Lane's Table 1.

[b] Estimated and averaged from data of Lane's Table 1 after subtraction of Subject 2's performance. Equal weight is given to Subjects 1 and 3.

[c] Read and computed from Lane, 1965, Figure 11.

were presented with the generalization test series of all seven stimuli in random order. The resulting identification functions (generalization gradients) are clear-cut (i.e., show little or no context effect) and are displayed in Lane's Figure 9. Thus, the first condition of categorical perception was established for three subjects.

2. "Following training, these (ABX) discrimination functions showed peaks at the phoneme boundary in the one-, two-, and three-step comparisons for all subjects [p. 298]." Average results for the three subjects are shown in Lane's Table 1 (p. 298): a significantly higher level of discrimination "between stimuli drawn from within a 'phoneme category' [p. 299]." However, the reported results are more ambiguous than this statement implies. Individual ABX data are given for only one subject, "the subject who provided the most orderly discrimination data" (Figure 11, p. 298). But, from Figure 9 (p. 294), Figure 11, and Table 1 (p. 298), we can reconstruct some of the results for the other subjects. From Figure 9, we can read the "phoneme classes" of each subject; from Figure 11 we can read the ABX performance of one subject; from the caption of Figure 11 (see the present Figure 2), we learn that for Subject 2, each ABX pair was presented 24 times; with this information we can infer from Lane's Table 1 how many times each ABX pair was presented on the average to the other two subjects and what their average ABX performance was. The results of these inferences are given in the present Table 1.[8]

We see immediately that the peaks at the "phoneme boundaries" (i.e., the increases in discriminative performance for stimuli drawn from different classes as against stimuli drawn from the same class) are highly irregular. Subject 2, whose discrimination data are the "most orderly," shows an obviously significant increase of 23% on the two-step function, and a presumably nonsignificant decrease of 1% on the three-step function. Subjects 1 and 3, on the other hand, show an average performance

---

[8] A curious fact emerges from this table: Each ABX pair was apparently not presented 24 times to Subjects 1 and 3, as it was to Subject 2, but some number less than 24 times for stimuli drawn from the same class, greater than 24 times for stimuli drawn from different classes. It would seem, therefore, that the ABX test presented to Subjects 1 and 3 was somehow differently weighted than that presented to Subject 2. Lane neither explains nor even mentions this oddity.

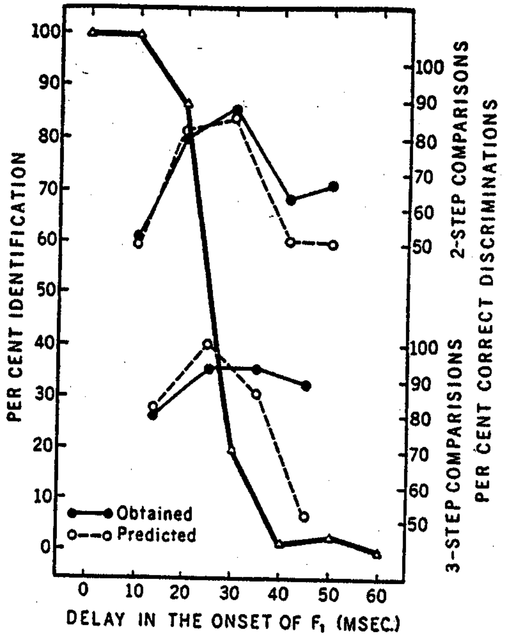

FIG. 2. "Discrimination of nonspeech (control) stimuli when presented in ABX triads to one listener. (Each point is the relative frequency of correct discriminations in 24 presentations of pairs of stimuli separated by two or three steps in the series of seven stimuli. This subject's identification gradient . . . is also plotted to reveal that peaks in discrimination occur at the 'phoneme boundary' of the nonspeech stimuli. The discrimination functions predicted from the identification gradient are shown by the dotted lines.)" (Reprinted from an article by Harlan Lane published in the July 1965 *Psychological Review*. Copyrighted by the American Psychological Association, Inc., 1965.)

that is just the reverse—a trivial peak of 4% on the two-step function, an obviously significant peak of 22% on the three-step function. Exactly how these peak advantages (averaged over several data points and two subjects) were made up, we cannot tell. But it is clear that (*a*) Subject 2, whose data were selected for display, shows a boundary peak only on the two-step comparisons; (*b*) on the average, Subjects 1 and 3 display boundary peaks only on the three-step comparisons; and (*c*) the discrimination functions are far more variable across subjects than Lane's summary statement of the results implies. The second condition of categorical perception has scarcely been reliably established for these stimuli.

3. We turn finally to the third purpose of the experiment, the prediction of discriminative performance from identification functions.

Figure 2 (Lane's Figure 11) compares two-step and three-step predicted and obtained discrimination for the satisfactory subject (S2). Figure 3 shows one-step predicted and obtained values for the same subject: the obtained data were drawn from an earlier version of Lane's critique (Lane, 1963, Figure 12); the predicted values were computed from the identification function of Figure 2.

For the two-step functions (Figure 2, top), the fit of obtained to predicted is good, but for the one- and three-step functions, it is not. First, on the three-step functions (Figure 2, bottom), obtained discrimination is superior to predicted by nearly 40% for Stimuli 3–6, which belong in the same "phoneme class" for this subject: a result typical of continuously perceived stimuli. Second, for two of the three-step pairs drawn from different phoneme classes (0–3, 1–4), obtained performance falls *below* predicted. The difference, though small, is

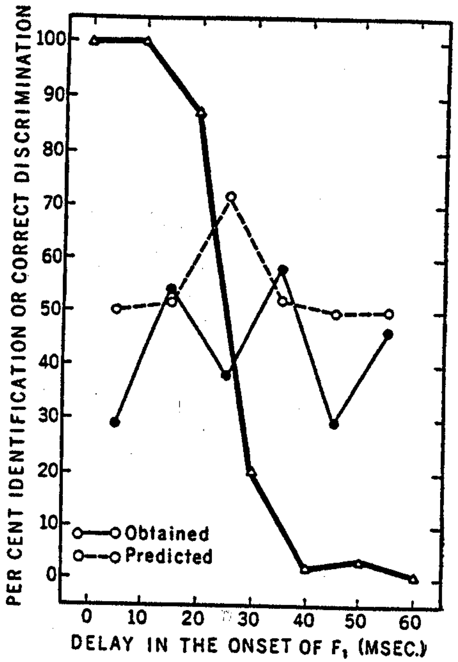

FIG. 3. One-step ABX discrimination of nonspeech (control) stimuli by the subject whose two- and three-step discrimination of these stimuli is graphed in Figure 2. (Obtained data redrawn, with permission, from Lane, 1963, Figure 12; predicted values computed from the identification function of Figure 2.)

arresting, since it means that on at least some trials, this subject failed to discriminate between stimuli that he had learned to identify, with some consistency, as different. This peculiar behavior is even more striking in the one-step function (Figure 3); four out of six obtained points (0–1, 2–3, 4–5, 5–6) fall *below* their predicted values. For pairs drawn from the same "phoneme class" (0–1, 4–5, 5–6), we may dismiss the defaults as perhaps no more than rather large chance deviations from predicted chance performance. But for the pair 2–3, members of different "phoneme classes," this will not do: a peak performance of 72% is predicted, a trough of 38% is obtained.

This failure of discrimination shows up in both the two- and three-step data of the other subjects. Table 2 presents the predicted and obtained discriminations for the averaged data of S1 and S3 (above) and for all three subjects (below). Predicted values were computed from Lane's Figure 9; obtained values for S1 and S3 (averaged) were drawn from the present Table 1, for the three subjects combined from Lane's Table 1. Here again obtained discrimination for stimuli drawn from the same "phoneme classes" falls well above the predicted discrimination, as might be expected for nonspeech stimuli. But even more important, on both two-step and three-step comparisons of stimuli belonging in different "phoneme classes," obtained performance falls *below* predicted. The interpretation of this outcome is clear: the acquired categorization was not binding. In fact, it was so unstable that subjects could not reliably make use of it to aid them in discrimination. If we recall that categorical stimuli can only be perceived absolutely, it is clear that these stimuli, so unreliably identified that they could not be reliably discriminated, do not qualify for inclusion in the class.

Given the highly variable discrimination data revealed in Table 1, and the anomalous outcome of the prediction tests, one cannot help feeling that Lane's summary account of the experiment failed to do justice to the richness of his data. Certainly the third condition of categorical perception has not been established for these stimuli.

Sectored circles study (*p. 299*). The purpose of this experiment was

> to determine whether an *O* who has learned to identify categorically stimuli drawn from a nonspeech continuum, discriminates more accurately between stimuli that lie on opposite sides of the identification boundary than he does between stimuli (separated by the same number of physical units) drawn from the same identification category [Cross et al., 1965, p. 69].

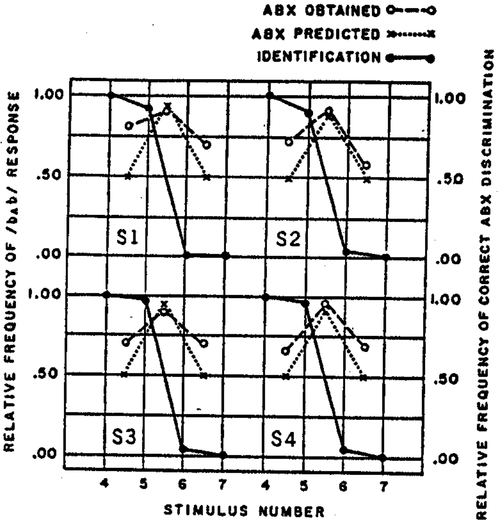

FIG. 4. "Identification and discrimination of visual stimuli by individual subjects. (Each point on the identification functions, filled circles, solid lines, shows the relative frequency of the /b ∧ b/ response in 24 presentations of the corresponding stimulus. Each point on the discrimination function, open circles, dashed lines, shows the relative frequency of correct discriminations in 36 presentations of each pair of stimuli.)" The predicted discrimination functions, dotted lines, crosses, were computed from the identification functions and added to Lane's figure by the present authors. (Reprinted from an article by Harlan Lane published in the July 1965 *Psychological Review*. Copyrighted by the American Psychological Association, Inc., 1965.)

The stimuli (a subset of stimuli used in a previous experiment and so numbered 4, 5, 6, 7) were four black discs, from which sectors of 42°, 46°, 50°, and 54° had been deleted. Training consisted in projecting the discs on a screen in random order and reinforcing (with a nickel) the response /b ∧ b/ to Stimuli 4 and 5, /g ∧ g/ to Stimuli 6 and 7, until 50 correct responses had been made. To test the effects of this training, the experimenters determined identification functions (generalization gradients) by presenting the four stimuli in random permutations, with instructions to identify each stimulus as either /b ∧ b/ or /g ∧ g/. Finally, they determined one-step ABX functions by presenting each possible stimulus pair 36 times. The results are reproduced in Figure 4. We have added the ABX function predicted from each individual's identification performance to the data shown by Lane (Figure 12, p. 299).

Was categorical perception demonstrated for these stimuli? (*a*) Identification functions are clear-cut. (*b*) Peaks appear at the class boundary (between Stimuli 5 and 6). This result is not unequivocal, however, since no control data are reported or were even collected. Were the peaks present before training? If not, do they represent a sharpening of discrimination at the boundary or a reduction within categories? It was to remedy this defect that we repeated the experiment (see below).

(*c*) The predicted and obtained functions agree well in their forms, but not in their levels within classes. Subjects generally discriminate much better than chance between stimuli drawn from the same class: even though they consistently give two stimuli (4 and 5, or 6 and 7) the same name in identifying them, they are still able to discriminate between them, if asked to do so. In other words, discrimination between these stimuli is clearly not dependent only on how they are identified. The categorical peaks are present, but the categorical troughs are absent. The procedures of this experiment appear to have established a degree of categorical perception typical of vowels.

Replications of the /do/–/to/ control and the sectored circles studies. The ambiguous results of these two attempts to establish categorical perception by identification training suggested the need for replications. Therefore, we repeated both studies; the sectored-circles study was also independently repeated by Parks, Wall, and Bastian (1969). Since space is limited, we will summarize the results of the replications and refer the interested reader to fuller accounts in the Haskins Laboratories Status Reports (Liberman, Studdert-Kennedy, Harris, & Cooper, 1965; Studdert-Kennedy & Liberman[9]).

Of the five subjects in the present /do/–/to/ control replication, only one achieved better than 75% accuracy in identifying the two extreme stimuli. The general effect of training on this subject's two-step discriminations was to raise performance: of five possible comparisons, only one showed a performance change in the direction predicted from identification. For three-step discrimination, training had no systematic effect: the fit of predicted to obtained was good for both before-training and after-training obtained discrimination. In short, we were no more successful than Lane and Schneider in establishing categorical perception of the /do/–/to/ control stimuli.

The present replication of the sectored-circles study also gives little support to Lane's contention. Identification training had no systematic effect on discrimination, and predicted functions showed no signs of agreement with obtained functions for any of the four subjects. The results were, in fact, remarkably similar to those of Parks et al. (1969), who reported that "individual discrimination functions were not related in any apparent, systematic way to the level of identification training [p. 245]." These investigators conclude, after a discussion of individual differences among their six subjects,

> it seems more probable that the consistent effect they [Cross et al., 1965] found was as much a matter of *S* [subject] selection as of identification training itself [p. 245].

From these five attempts (two by Lane, three by others) to establish categorical perception of nonspeech stimuli by identification training, we conclude that while training may, for some subjects, increase distinctiveness between categories and produce a degree of categorical perception such as that found with vowels, we have no evidence for an approach through training to decreased distinctiveness within categories and the degree of categorical perception found with consonants.

Correspondence between identification and discrimination functions (pp. 300–302). "Nor does the *degree* of correspondence between labeling and discrimination provide a basis for distinguishing among the perception of consonants, vowels, and nonspeech stimuli . . . [Lane, 1965, p. 300]." We disagree with this statement. The degree to which identification and discrimination functions are related may be estimated by application of the predictive test. Lane's criticisms of this test are so seriously mistaken that we will begin this section by spelling out its assumptions and procedure in detail.

The Predictive Test

Assumption 1: That stimuli are perceived categorically. (Note, that, as we have already stated, stimuli so perceived can *only* be identified absolutely: Therefore, perception is independent of context and stimuli can be discriminated only if they are identified as belonging to different categories.)

---

[9] M. Studdert-Kennedy and A. M. Liberman. Some repeated attempts to establish categorical perception of nonspeech stimuli by discrimination training. Haskins Laboratories Status Reports (forthcoming).

Assumption 2: That subjects use the same categories in discrimination as they use in identification. (This assumption permits us to estimate the probability of a given stimulus being (covertly) identified as belonging in a given category during the discrimination test from the probability that it was (overtly) so identified in the identification test.)

The procedure is to determine the relative frequency with which each stimulus was identified as belonging in each category during the identification test and to predict from this the relative frequency with which each pair of stimuli would be correctly discriminated (because they were in different categories) during the discrimination test. The predicted discrimination performance is then compared with the obtained performance. If agreement is perfect, we may say that the subject's behavior is consistent with the assumptions. If agreement is less than perfect, we must conclude that one or the other of the assumptions was not met. Since Assumption 2 is usually validated by the appearance of peaks in the discrimination data at phoneme boundaries, we can conclude from imperfect agreement that Assumption 1 did not hold. The necessary prediction formulae are derived in Liberman et al. (1957). There is, however, one danger point in the procedure: the determination of the identification probabilities. There are two possible ways of doing this, which we will call context-out and context-in. They lead to identical results if the stimuli, being categorically perceived, show no effects of context. But, if the stimuli are continuously perceived and do show effects of context, the context-in procedure (recommended by Lane, 1965, p. 301) will lead to a spuriously high degree of agreement between predicted and obtained discrimination curves.

The context-out procedure estimates identification probabilities directly from the overall relative frequencies with which each stimulus was identified as belonging in each category. The same probability for a given stimulus is then used to predict its discrimination from every stimulus with which it was paired. The context-in procedure, on the other hand, estimates identification probabilities for a given stimulus separately according to its neighboring stimulus. As many different identification probabilities are then used to predict its discrimination as there are stimuli with which it was paired.

The outcome of the context-in procedure is a much closer fit between predicted and obtained discrimination functions—*provided that there are context effects.* The reason for this is that if contextual effects occur in identification, the subject is not absolutely identifying the stimuli; rather, he is making, in some degree, a comparative or discriminative judgment and using that as a partial basis for his supposed identifications. Consequently, the identification probabilities for two stimuli—estimated from the relative frequency with which they were identified in the several categories *when they occurred as neighbors*—already reflect the processes of discrimination. When discrimination is predicted, we are then, in effect, predicting discrimination from discrimination.

An example may make this clear. Suppose that we have three continuously perceived pitches, 1, 2, 3, each 100% discriminable from its neighbor. We present them to a subject repeatedly in random order, asking him to identify them as High or Low. We find that he always calls 1 Low and 3 High; but 2 he calls High whenever it follows 1, Low whenever it follows 3. If we use the *context-out* procedure, we shall estimate the identification probabilities for Pitch 2 as .50 for Low and .50 for High: we shall then predict 50% discriminability between Pitches 1 and 2, and between 2 and 3. Since the pitches are all perfectly discriminable, the predicted curve will fall well below the obtained—as it should, since these stimuli display context effects and are therefore not categorical. On the other hand, if we use the *context-in* procedure, we shall estimate the identification probabilities for Pitch 2 as 1.00 for High when presented with Pitch 1, and 1.00 for Low when presented with Pitch 3: we shall then predict 100% discriminability between 1 and 2, and between 2 and 3. The predicted curve will perfectly match the obtained—as it should *not* since these stimuli are not categorical. The match is spurious and is due to an improper application of the test.

We may push this example further to illustrate other possible types of continua. Suppose that the three pitches were again continuously perceived and perfectly discriminable, but displayed no context effects, so that 1 and 2 were always called Low, 3 High. The *context-out* and *context-in* procedures would here yield identical results: we would predict 50% discriminability between 1 and 2, 100% discriminability between 2 and 3. But the predicted point for the discrimination of Pitches 1 and 2 would be 50% below the obtained point—as it should be, since these pitches, even though they yield sharp identification curves (i.e., no context effect), are all equally discriminable.

Suppose, finally, that the three pitches were categorically perceived, so that 1 and 2 were always called Low, 3 High, but could only be discriminated on the basis of their labels. The *context-in* and *context-out* procedures would again yield identical results: 50% discriminability between 1 and 2, 100% discriminability between 2 and 3. But the predicted and obtained curves would now agree perfectly—as, indeed, they should, since these pitches are categorically perceived.

In practice, predicted and obtained curves never agree perfectly, but some continua approach perfect agreement more closely than others. Correctly applied, the test permits a clear comparison between different sets of stimuli.

Lane is aware of this. The two methods of prediction were, in fact, compared in a study by Cross and Lane (1964, Experiment III) of the identification and discrimination of 1,000-Hertz tones differing in intensity. As might be expected of stimuli which, like vowels, are liable to strong context effects, Cross and Lane found that

> predictions based on the assumption that ABX discriminations are mediated by comparative responses [i.e., the context-in assumption] . . . are more in line with obtained discriminations than are the predictions based on the assumption that categorical judgments . . . mediate the discrimination [i.e., the context-out assumption] [p. 21].

Given this explicit recognition of the function of the test and of the difference between the two methods of prediction, some of Lane's criticisms in the review are rather surprising.

Lane's criticisms of the test. "1. The predicted discrimination function is derived from the averaged labeling gradients of preselected subjects . . . the criterion for preselection . . . has not been . . . comparable in experiments with vowels and consonants. Therefore . . . the predicted discrimination functions are also not comparable . . . [Lane, 1965, p. 300]."

The criticism makes two points: (*a*) the criteria for subject selection were not the same in different experiments and (*b*) the data were averaged. We answer them in this order. The difficulty of using the same selection criteria in all experiments arises because within-subject variance for vowel identification is almost always greater than for consonant identification: vowels are subject to context-effects, as consonants are not, so that sets of identifiers equally consistent under the two conditions are virtually impossible to find.

There seem to be two possible solutions to the difficulty. First, we could reduce the number of steps along the vowel continuum, while maintaining the same range, so that stimuli were more widely spaced: vowel identification functions would then be as free of context effects and as clear-cut as consonant functions. Subjects could be matched across conditions on the basis of their identification functions and Lane's criticism would be met.

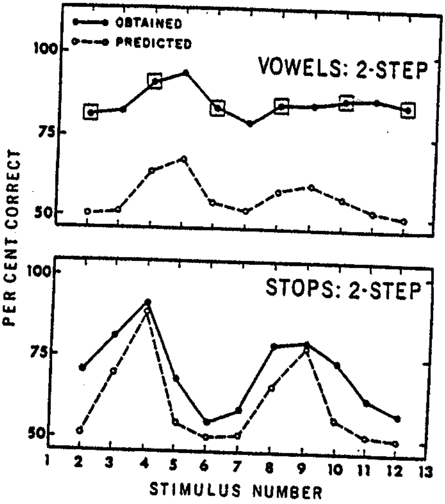

FIG. 5. A comparison of two-step obtained and predicted discrimination functions for a series of three stop consonants and three unrounded vowels. (Reprinted from an article by K. N. Stevens, A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Ohman published in *Language and Speech*, 1969, Vol. 12. Copyrighted by Robert Draper Ltd., Teddington, England, 1969).

Fortunately, we do not need to do this experiment since we already have the data from which to infer its outcome. Figure 5 presents data from a comparative study of vowels and consonants (Stevens et al., 1969). The stimuli were a series of vowels ranging from /i/ through /I/ to /ɛ/ in 12 logarithmically equal steps and a series of stop consonants ranging from /b/ through /d/ to /g/ in 12 equal steps. They were prepared on OVE II, the speech synthesizer at the Royal Institute of Technology in Stockholm. There were 11 subjects in the vowel experiment, eight in the consonant experiment. Subjects provided identification data (from which the predicted functions of Figure 5 were computed) and ABX discrimination data for all pairs of stimuli separated by one, two, or three steps. Figure 5 shows the two-step obtained and predicted discrimination for the two classes of stimuli.[10]

---

[10] Michael Vinegrad (personal communication, September 1969) has recently carried out a scaling study with these same stimuli. Perceived distances between physically adjacent stops were smaller within than between phoneme classes. In the case of the vowels, the perceived distances were more nearly equal.

TABLE 2

COMPARISON OF PREDICTED AND OBTAINED PERCENTAGE CORRECT DISCRIMINATION FOR LANE'S THREE SUBJECTS ON THE DO–TO CONTROL STIMULI

Comparison	Subjects	Same "phoneme class"			Different "phoneme class"		
		Obtained	Predicted	O−P	Obtained	Predicted	O−P
Two-step	S1, S3 (average)	75	50	25	79	93	−14[a]
Three-step	S1, S3 (average)	63	50	13	85	95	−10
Two-step	S1, S2, S3 (average)	67	50	17	81	90	−9
Three-step	S1, S2, S3 (average)	71	51	20	86	94	−8

[a] The negative values of obtained minus predicted in this column indicate that subjects were sometimes unable to discriminate between stimuli that they had learned to identify as different.

Suppose now that the vowel series had consisted of only seven stimuli instead of 13: say, Stimuli 1, 3, 5, 7, 9, 11, 13. Suppose, further, that identification functions for the seven vowels had been as clear-cut as those for the 13 consonants, so that the predicted function for the vowels had shown the same high peaks and low troughs (though not necessarily in the same places along the continuum) as the consonants. Would then the match between predicted and obtained vowel discrimination have been improved? The answer is given by an examination of the obtained vowel data of Figure 7. The obtained discrimination performances for comparisons between Stimuli 1–3, 3–5, 5–7, 7–9, 9–11, 11–13 are marked with squares. It is evident that even if the predicted peaks had fallen at the obtained second (3–5) and fifth (9–11) points above the continuum (thus matching the obtained as well as possible), for example, the predicted trough would not have been well matched. The very best agreement possible between predicted and obtained vowel discrimination would still be much inferior to the corresponding consonant agreement. Nor would the vowel match have been improved by reducing the vowel series to Stimuli 1, 4, 7, 10, 13, for example. On the contrary, whatever was gained in consistency of identification (and so in quantal leap from predicted trough to predicted peak) would only have served to increase the discrepancy between predicted and obtained discrimination, since three-step discrimination yields an overall higher level of performance than two step.

In short, even if subjects had been matched on their identification performances, as Lane might wish, the most that could have emerged is an improved match between predicted and obtained for the vowels at discrimination peaks; the discrepancy at the troughs would have been at least as great as that which we have observed. And we would, furthermore, still be faced with the question of why we were obliged to use fewer steps along a continuum encompassing three vowels than along a consonant continuum encompassing three consonants in order to obtain similar identification functions. In other words, we would still be confronted with the fact—central to the concepts of categorical and continuous perception in speech—that vowels are more readily confused with one another than are consonants.

The second solution to the problem of matching consonant and vowel continua is the one that we chose: to divide the continua into equal numbers of steps and to select for study those subjects whose identifications of a given set of stimuli were most consistent. In some studies it was not necessary to select. The procedure was essentially that of conditioning studies in which the experimenter uses only those organisms that he is able to condition: we rejected only those subjects whose identification performance was so poor that their data were unintelligible (compare Cross & Lane, 1964, e.g.) Since some subjects almost always have difficulty with synthetic speech stimuli, some such selection procedure seems to be inevitable.

Lane's second point is that the data of the selected subjects were averaged in two studies (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Fry et al., 1962). He points out correctly that predicted functions vary with the consistency of the identifiers: if an inconsistent identifier is added to a group, the predicted curve will be lowered, while if an inconsistent identifier is dropped, the predicted curve will be raised. By judicious dropping and adding, Lane suggests, one may manipulate the match between predicted and obtained.

The fallacy in this argument is that it assumes identification and discrimination to be unrelated—the negative form of the very assumption that we wish to test. For if they are related (i.e., if perception is categorical, so that discrimination is a function of identification), every movement up or down of the predicted curve will be matched by a corresponding movement up or down of the obtained curve, and no amount of dropping or adding subjects will improve the match. Lane's argument reduces to the statement: this test for categorical perception is illegitimate because perception is not categorical. This is not an acceptable argument.

2. Lane's next step is to state three assumptions for the test when, in fact, there are only two. He splits Assumption 1 into two parts to yield his Assumptions 1 (p. 300) and 3 (p. 301) (see also p. 283 of his review). He gives these assumptions as: (1) "that speech stimuli are discriminated to the extent that they are labeled differently" and (3) that "the various stimuli within each triad are perceived independently of each other [quoted by Lane (p. 301) from Liberman et al., 1957, p. 363]." This division is unnecessary, since if stimuli are discriminated only when they are differently identified, they must be perceived independently. They *cannot* be subject to contextual effects, since contextual effects arise from comparative, or discriminative, judgment, which Assumption 1 has already precluded by positing that the stimuli can only be identified. This is not a trivial point, since it is Lane's failure to see that categorical perception entails context-free perception that leads him astray. We should emphasize once again that context-free perception is a necessary but not a sufficient condition of categorical perception.

Having stated Assumption 1, Lane goes on to say that it "has been disconfirmed by every experiment." He then cites in evidence Fry et al.'s (1962) vowel study in which predicted and obtained discrimination gave a poor match. But this, of course, is precisely what Fry et al. were at pains to point out. The disconfirmation of Assumption 1 for the vowels constitutes the rejection of the hypothesis of categorical perception for vowels. Why Lane sees this rejection of the hypothesis as somehow invalidating the test is difficult to understand. Does he object to the rejection of the hypothesis (assumption) of equal means in an analysis of variance on the grounds that it invalidates the *F* test?

Be that as it may, we fully agree with Lane that Fry's study shows that vowels are not categorically perceived and may be readily discriminated. Lane goes on to suggest that the high discriminability of the vowels may be largely due to their long duration (as compared with consonant duration) and their marked variations in "intrinsic" amplitudes. We agree that these are crucial differences between consonants and vowels. We do not understand why Lane implies that these differences are artifacts that should have been eliminated from the experiments. Had they been eliminated, we would no longer have been comparing consonants and vowels.

3. Lane's next criticism is of Assumption 2. He cites data to show that subjects may be able to use more categories in their identifications of vowels than they are permitted to use by the experimental instructions. Fry et al. (1962) permitted subjects to use only three categories: /I/, /ε/ or /æ/ in their identifications; a subject in Lane's laboratory, permitted to use nine categories (/i, I, ε, æ, a/ and four intermediate values) for the same stimuli, used all nine equally often. Lane argues that Fry et al. were wrong to assume that during discrimination, subjects used the same categories as they had used during identification: in fact, they may have used many more and this may account for the very high level of discriminative performance observed in this study—and for the large discrepancy between predicted and obtained discrimination.

We cannot refute this argument by reference to Fry et al.'s data, since there are no peaks in the discrimination function to indicate where subjects placed category boundaries. However, Lane himself has provided the relevant data. Cross and Lane (1964) (see p. 297 of Lane's review) "degraded" the Fry et al. vowels so as to reduce the overall level of discrimination: "marked peaks in the discrimination functions were observed at the phoneme boundaries" of the only subject whose "identification functions were even nearly categorical." How many peaks were there? If the subject had been using nine categories, there should have been eight. In fact, there were only two. Since these fell neatly at the /I/–/ε/ and /ε/–/æ/ boundaries, this subject was evidently using only three categories during discrimination. Thus the only data that we have, relevant to the validity of Assumption *2* for these vowels, clearly support the assumption.

4. Lane's final criticism of the test is that "the predicted discrimination functions for the vowels violate a third assumption of the prediction formula." As we have seen, this "third assumption" is entailed, in fact, by the first assumption: that "the various stimuli within each triad are perceived independently of each other" is entailed by the categorical assumption. Once again, it is not a matter of "violating" the assumption, but of rejecting the hypothesis of categorical perception for the vowels.

In making this point, Lane reveals a curious misunderstanding of the predictive test. He writes,

> had Fry et al. (1962) acknowledged context effects in their predictive formula, the apparent disparity between predicted and obtained discrimination functions would have been reduced 10 to 15% [p. 301].

Lane is right. But had Fry et al. done this, they would also have been misapplying the test, for, as we saw earlier, it is precisely with stimuli that are subject to context effects that the context-in procedure yields a spuriously high degree of agreement between predicted and obtained discrimination. The agreement is spurious because by "acknowledging context," we predict discrimination from discrimination. If one recalls that Lane himself conducted an excellent experiment (Cross & Lane, 1964, Experiment III) explicitly directed toward a comparison between the consequences of making the context-in (or comparative) assumption and the consequences of making the context-out (or categorical) assumption, he is really puzzled. Why does Lane expect Fry et al. to have made the comparative assumption when it was the categorical assumption that they wished to test?

CONCLUSION

We have now answered Lane's major criticisms. We disagree with many other interpretations scattered throughout his review, but we have said enough for the main outlines of the present position to be clear. What can be said in summary?

Lane does not deny that perception may be categorical. On the contrary, he sees the phenomenon as quite general. He believes that the relation between the identification and discrimination of consonants (from which the concept of categorical perception was derived) is also found with vowels and nonspeech stimuli, and that "these relations are attributable to the general paradigm for discriminative training and testing [p. 307]." As far as nonspeech stimuli are concerned, we do not believe that he has made his case. His /do/–/to/ control study is an inconclusive tissue of variability and anomaly. His sectored circle study shows that (like vowels) nonspeech stimuli may be grouped into categories with some enhancement of discrimination at category boundaries, but that (unlike consonants) they remain highly discriminable within categories. Replications of these studies by the present authors and others give even less support to his position. For the vowels, Lane does not deny the large discrepancy between predicted and obtained discrimination when predictions are based on the categorical assumption. But he objects to this assumption (which the prediction procedure is designed to test): he proposes that we "acknowledge contextual effects" and make the test with a comparative assumption. This reflects a serious misunderstanding, since with a comparative assumption the test would be, as far as we can judge, otiose. The misunderstanding stems from Lane's failure to see that categorical perception is necessarily independent of context.

Once this misunderstanding has been cleared away, Lane's view and the present view of the differences between consonant and vowel perception seem to be quite similar. Elsewhere (Cross & Lane, 1964, pp. 23 and 24), he has written of consonants:

> It is understandable . . . that comparative responses would play no role in ABX discriminations, since *S* has learned only to identify these sounds categorically.

of vowels:

> Comparative responding . . . as well as the categorical response labels . . . may both play a role in the discrimination of these sounds . . . the acoustic properties of the vowels make them prey to comparative as well as categorical identification and discrimination.

There remains, however, an important disagreement between Lane and the present authors in the interpretation of categorical perception. Lane believes that it is merely a result of discrimination training. We, on the other hand, are willing to entertain the hypothesis that it reflects some structurally determined process, adapted to the complex code that links the sounds of speech to the phonetic message they convey. On this view, the basic difference between stops and steady-state vowels is that stops are more complexly encoded in the sound stream and, therefore, more in need of a special speech-sound processor (see Liberman et al., 1967). Categorical perception is only one consequence of the operation of the speech-sound processor (Liberman, in press). Moreover, the different processing of the more complexly encoded stops and the less complexly encoded vowels is reflected in at least two other striking differences in perception: stops and certain other consonants are more strongly lateralized in the cerebral hemispheres than steady-state vowels (Darwin, 1969; Kirstein & Shankweiler, 1969; Rupf, Hughes, & House, 1969; Shankweiler, in press; Shankweiler & Studdert-Kennedy, 1967a, 1967b; Studdert-Kennedy & Shankweiler, 1970); and these two kinds of speech sounds are affected in opposite ways by binaural time differences (Porter, Shankweiler, & Liberman, 1969).

We accept that experience may affect the speech-sound processor, or even be a necessary condition for its proper operation, since we see no reason to suppose that the processor is different in this respect from other biological devices, including those that underlie highly instinctual and species-specific kinds of behavior. In the particular case of categorical perception, we know from cross-language studies by Lisker and Abramson (1964, 1965, in press) and Abramson and Lisker (in press) that linguistic experience has a considerable effect. The present authors' disagreement with Lane is in our belief that experience and training are not sufficient conditions for categorical perception or for other characteristics of perception in the speech mode.

Lane correctly attributes to us the view that speech perception is somehow carried out in articulatory terms. But we would prefer now to apply this hypothesis only to the encoded sounds, not necessarily to those aspects of the speech signal that are not highly encoded. The most important basis for the motor hypothesis is that the apparently arbitrary code linking sound and phonetic message may be rationalized by an articulatory model. Such a model is plainly available to man—even, perhaps, to one who has never spoken; it is hard to believe that man does not use a key to which he has such ready access.

Lane is also correct in saying that we have used the facts of categorical perception to support a motor theory. But theories about the nature of the speech-sound processor ought, we think, to be separated from questions about its existence and about the perceptual phenomena it produces. That distinction is important in the context of this reply because the authors' disagreement with Lane extends to a broader and more interesting question: Is phonetic perception to be explained by the principles of auditory psychophysics and discrimination learning? Lane would answer, "Yes." We say, "No." In our view, some of phonetic perception may be accomplished by a special decoding device available to man as part of his species-specific capacity for language. Categorical perception would then be one result of the operation of that device.

REFERENCES

ABRAMSON, A. S., & LISKER, L. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences, Prague, 1967*. Prague, in press.

BASTIAN, J., & ABRAMSON, A. S. Identification and discrimination of phonemic vowel duration. *Journal of the Acoustical Society of America*, 1962, 34, 743–744. (Abstract)

BEARE, A. C. Color name as a function of wavelength. *American Journal of Psychology*, 1963, 76, 248–256.

CROSS, D. V., & LANE, H. L. On the discriminative control of concurrent responses: The relations among response frequency, latency and topography in auditory generalization. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 487–496.

CROSS, D. V., & LANE, H. L. *An analysis of the relations between identification and discrimination functions for speech and nonspeech continua.* (Report No. 05613-3-P) Ann Arbor: University of Michigan, Behavior Analysis Laboratory, 1964.

CROSS, D. V., LANE, H. L., & SHEPPARD, W. C. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *Journal of Experimental Psychology*, 1965, 70, 63–74.

DARWIN, C. J. Laterality effects in the recall of steady-state and transient speech sounds. *Journal of the Acoustical Society of America*, 1969, 46, 114. (Abstract)

EKMAN, G. Contributions to the psychophysics of color vision. *Studium Generale*, 1963, 16, 54–64.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D., & LIBERMAN, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, 5, 171–189.

KIRSTEIN, E., & SHANKWEILER, D. Selective listening for dichotically presented stop consonants and vowels. Paper presented at the meetings of the Eastern Psychological Association, Philadelphia, April 1969.

LANE, H. L. *The motor theory of speech perception: A critical review.* (Progress Report No. 4) Ann Arbor: University of Michigan, Behavior Analysis Laboratory, 1963.

LANE, H. L. The motor theory of speech perception: A critical review. *Psychological Review,* 1965, 72, 275–309.

LANE, H. L., & SCHNEIDER, B. A. *Discriminative control of concurrent responses by the intensity, duration and relative onset time of auditory stimuli.* (Report) Ann Arbor: University of Michigan, Behavior Analysis Laboratory, 1963.

LIBERMAN, A. M. Some characteristics of perception in the speech mode. In, *Proceedings of the Association for Research in Nervous and Mental Disease* (*1968*). Baltimore: Williams & Wilkins, in press.

LIBERMAN, A. M., COOPER, F. S., HARRIS, K. S., MACNEILAGE, P. F., & STUDDERT-KENNEDY, M. Some observations on a model for speech perception. In, *Proceedings of the AFCRL symposium on models for the perception of speech and visual form* (*Boston, November 1964*). Cambridge: M.I.T. Press, 1968.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. S., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review,* 1967, 74, 431–461.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L., & BASTIAN, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech,* 1961, 4, 175–195.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology,* 1957, 54, 358–368.

LIBERMAN, A. M., HARRIS, K. S., KINNEY, J., & LANE, H. L. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology,* 1961, 61, 379–388.

LIBERMAN, A. M., STUDDERT-KENNEDY, M., HARRIS, K. S., & COPPER, F. S. *A reply to "Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception" by Cross, Lane, and Sheppard.* (SR-3) New York: Haskins Laboratories, 1965.

LISKER, L., & ABRAMSON, A. S. Cross-language study of voicing in initial stops: Acoustical measurements. *Word,* 1964, 20, 384–422.

LISKER, L., & ABRAMSON, A. S. Stop categories and voice onset time. In, *Proceedings of the fifth international congress of phonetic sciences, Münster, 1964.* Basel: S. Karger, 1965.

LISKER, L., & ABRAMSON, A. S. The voicing dimension: Some experiments in comparative phonetics. In, *Proceedings of the sixth international congress of phonetic sciences, Prague, 1967.* Prague, in press.

MATTINGLY, I., & LIBERMAN, A. M. The speech code and the physiology of language. In K. N. Leibovic (Ed.), *Information processing in the nervous system.* New York: Springer Verlag, 1970.

PARKS, T., WALL, C., & BASTIAN, J. Intercategory and intracategory discrimination for one visual continuum. *Journal of Experimental Psychology,* 1969, 81, 241–245.

PORTER, R., SHANKWEILER, D., & LIBERMAN, A. Differential effects of binaural time differences on perception of stop consonants and vowels. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September 1969.

RUPF, J. A., HUGHES, G. W., & HOUSE, A. S. Interaural switching effects on the recognition of consonants. *Journal of the Acoustical Society of America,* 1969, 46, 114. (Abstract)

SHANKWEILER, D. An analysis of laterality effects in speech perception. In P. Kjeldergaard (Ed.), *Perception of language.* Pittsburgh: University of Pittsburgh Press, in press.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology,* 1967, 19, 59–63. (a)

SHANKWEILER, D., & STUDDERT-KENNEDY, M. An analysis of perceptual confusions in identifications of dichotically presented CVC syllables. *Journal of the Acoustical Society of America,* 1967, 41, 1581. (Abstract) (b)

STEVENS, K. N., LIBERMAN, A. M., STUDDERT-KENNEDY, M., & OHMAN, S. E. G. Cross-language study of vowel perception. *Language and Speech,* 1969, 12, 1–23.

STEVENS, K. N., OHMAN, S. E. G., & LIBERMAN, A. M. Identification and discrimination of rounded and unrounded vowels. *Journal of the Acoustical Society of America,* 1963, 35, 1900. (Abstract)

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

STUDDERT-KENNEDY, M., & SHANKWEILER, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America,* 1970, in press.

(Received June 23, 1969)