
Alvin M. Liberman, Michael Studdert-Kennedy, Katherine S. Harris, and Franklin S. Cooper

Haskins Laboratories, New York City

A paper recently published in this JOURNAL by Cross, Lane, and Sheppard (1965) argued that speech perception can be accounted for without assuming any special perceptual mechanism, including the kind suggested by a motor theory (Cooper, Liberman, Harris, and Grubb, 1958; Liberman, Cooper, Harris, and MacNeilage, 1963; Liberman, Cooper, Harris, MacNeilage, and Studdert-Kennedy, 1964; Lisker, Cooper, and Liberman, 1962). More particularly, the paper tried to show that one of the characteristics of the perception of certain consonants -- a characteristic that we have called categorical perception -- can be produced with simple visual stimuli by applying a "general paradigm for discrimination training and testing." In reply we should say:

(1) Though the experiments by Cross, Lane, and Sheppard (CLS) were intended to show the effects of particular training procedures on visual perception, there are no controls to establish how the stimuli were responded to without training or with only simple instructions. There is, therefore, no baseline against which to examine the effect, if any, of the training procedures.

(2) Though the effects attributed to the training procedures are said to be equivalent to the effects we have observed for consonant perception -- i.e., categorical perception -- the
authors do not properly test their data to see whether, in fact, this is so. We have made the appropriate test and found that their data fall short of categorical perception.

(3) Though none of their training conditions permits a match between response and stimulus -- a salient feature of speech perception and a central consideration in the motor theory to which the authors refer -- they assert that the motor theory would predict a difference in the outcomes of two of their training procedures. The motor theory makes no such prediction, and is therefore untouched by the outcome of their experiment.

(4) In developing the implications of their response latency data, the authors impute to us an interpretation we have never made or even implied. Since that interpretation is used to advance their argument, we must emphasize that it is their invention, not ours.

We will deal with these four main points in detail in considering the two CLS experiments.

CLS Experiment I

In the first of their two experiments, CLS were concerned with the effects of discrimination training procedures on the identification of visual stimuli. The stimuli were nine discs from which sectors, ranging in size from 30 to 62 degrees, had been deleted. Training consisted in presenting the two extremes of the stimulus range and reinforcing different verbal responses to them. One group of subjects was trained to say "bAb" or "bib" (a vowel contrast), the other "bAb" or "gAg" (a consonant contrast). To test the effects of this training, the experimenters determined generalization gradients, or identification functions, by presenting all nine stimuli 10 times each, with instructions to the subjects to identify each stimulus as "bAb" or "bib" (or "gAg"). They interpreted their data as examples of what we have called categorical perception, under both the vowel contrast and
the consonant contrast conditions of training.

There are four main defects in this experiment and in the interpretations put upon it:

(1) The authors make much of having obtained their data by "appropriate conditioning techniques" that require "twenty minutes of discrimination training," following a general procedure for such training that is referred to in a later paper as the "Cross-Lane paradigm" (Lane, 1965b). In fact, they might have obtained their data with no special training and less obfuscation, by simply following the two-category method of single stimuli (Woodworth and Schlosberg, 1954, p. 217).

We ourselves have obtained identification functions with cardboard replicas of the CLS stimuli after reading to subjects the following instructions:

I am going to show you a set of circles which differ only with respect to the size of the wedge cut out of them. Here is the smallest wedge. (The experimenter displays the smallest one for three seconds.) Here is the largest. (The experimenter displays the largest for three seconds.) Now I shall show you these and a number of other circles with wedges cut out of them of sizes between the two extremes. I want you to look at each circle, and then record on your answer sheet an S if you think the stimulus wedge is more like the smallest wedge, an L if you think it is more like the largest. Once again, here is the smallest. (The experimenter displays it for three seconds.) And here is the largest. (The experimenter displays it for three seconds.)

Stimuli were then displayed by hand from behind a screen for somewhat less than a second, at roughly 10-second intervals. Figure 1 presents the data. (Subjects 2, 5, 6, 8, 9, 10 had previously taken an ABX test with stimuli 4, 5, 6, and 7 from the series, but had had no identification training. Their data do not differ sys-
tematically from those of subjects 1, 3, 4, 7 who had had no previous experience whatever with the stimuli.) All the data agree very well with those of CLS, though generated without recourse to the "Cross-Lane paradigm." We conclude that whatever is shown in this first CLS experiment is not due to any peculiarly "appropriate conditioning techniques," but may be produced by simple instructions and use of a conventional two-category method of single stimuli.

(2) What then is shown? The authors infer from a comparison of their data with those collected for certain consonants (see CLS Fig. 5) that they have demonstrated categorical perception. This seems to be due to their mistaken belief that categorical perception may be inferred from the abruptness of the discontinuities in the identification functions.\(^4\) However, as we have said in our papers on this subject, (Liberman, et al., 1957, 1961, 1961; Fry, Abramson, Eimas, and Liberman, 1962) categorical perception is, rather, evidenced by a relation between discrimination and absolute identification. In completely categorical perception, a subject can discriminate stimuli no better than he can identify them absolutely. This is a mode of perception that yields categories directly and without distinguishable sub-categories, not merely a process that leads to classification of stimuli despite discriminable differences among them.\(^5\) Thus, an adequate test for categorical perception requires a quantitative comparison of discrimination and absolute identification. We described such a test in our first paper on the near-categorical perception of the stops (Liberman, Harris, Hoffman, and Griffith, 1957). Our conclusions about tendencies toward such perception in the stops (and the quite different tendencies we find with the vowels and certain other phoneme distinctions) have been based on the application of that test. (Bastian, Harris, and Liberman, in preparation; Eimas, 1963; Griffith, 1957; Fry, Abramson, Eimas and Liberman, 1962; Liberman, Harris, Hoffman and Griffith, 1957; Liberman,
Harris, Kinney, Lane, 1961; Liberman, Harris, Lisker, Bastian, 1961;

It is, of course, true that categorically perceived stimuli will have abrupt and distinct identification functions. But there are other conditions that will produce such functions. In a simple two-choice situation like that of the CLS experiment, we should expect the identification functions to be abrupt if the stimuli were spread widely apart. That the CLS stimuli are, in fact, spaced widely apart is indicated by the results of their discrimination testing in the second experiment: we see there that discrimination of one-step differences is at a high level.

(3) CLS point out next that the same identification functions were obtained whether the subjects were trained to respond with "bAb" and "gAg" (consonant contrast) or with "bAb" and "bib" (vowel contrast). By way of explaining why they did this experiment, and what they make of the results, the authors say that, contrary to the implications of the motor theory, the "topography" of the response (that is, whether it was a consonant or a vowel contrast) made no difference. In assuming that it might be expected to make a difference, they have misunderstood the implications of the motor theory. A central fact of speech perception is that its stimuli are isomorphic with the perceiver's potential responses: the stimuli permit matching responses. This fact is essential to a motor theory of speech perception. Indeed, its significance was explicitly developed in one of the papers CLS cite (Liberman, 1957). Certainly, our version of a motor theory produces no expectation of what should happen if the responses bear only an arbitrary relation to the stimuli. To ask subjects, as CLS did, to respond to visual stimuli with a consonant contrast in one case and a vowel contrast in another, is to carry out an experiment that simply has no bearing on speech perception.

(4) CLS make much of one other finding of their first experiment: there was, in most cases, an increase in response la-
tency for the stimuli that lie midway between the labeled classes. This is, of course, to be expected since the subjects would presumably be most uncertain there. But the authors try to make this simple and obvious finding relevant to their argument, and in so doing they seriously misrepresent what we have done and said in our own work on reaction time. Referring to the results of their experiment, CLS say (p.72):

These findings for response latency in identification of non-speech visual forms accord with those reported for identification of synthetic speech stimuli (although the latter findings were interpreted as favoring a mediation hypothesis): "For both stops and vowels voice and button-press reaction times were lower near phoneme centers than in the vicinity of phoneme boundaries. [Studdert-Kennedy, Liberman and Stevens, 1963, p. 1009]." (Factual ours).

The quoted statement by Studdert-Kennedy et al. is from an abstract of a paper read before the Acoustical Society of America. It does, indeed, report that reaction time increases were found at phoneme boundaries, but there is in the abstract no word, phrase, or sentence that could be taken to imply that such increases are assumed to have any relation to a "mediation hypothesis" or a motor theory. In the paper itself, Studdert-Kennedy et al. considered the increases in reaction time in the light of Hick's law that reaction time is a linear function of uncertainty. They were interested to know whether Hick's law holds for the kind of uncertainty that arises at phoneme boundaries. (It does, approximately.)

A bit farther on CLS point out that latency peaks have been found at stimulus boundaries by many investigators, and then say: "...if it [that finding] is interpreted as implying that perception is categorical it is, nevertheless, not peculiar to speech perception." We fail to see any plausible basis for such an inter-
pretation; certainly, we have never made it, nor, so far as we know, has anyone else. In the present context, reaction time merely reflects uncertainty, which is hardly peculiar to speech.

CLS Experiment II

In the second experiment, CLS were concerned with the effects of their training procedures on the discrimination of the same visual stimuli they had used in the first experiment. Four of the original nine stimuli, taken from the middle part of the range, were used in this experiment. Subjects were first reinforced for saying "bab" in the presence of Stimuli 4 and 5 and "gag" in the presence of Stimuli 6 and 7. Identification testing was similar to that of the first experiment. In addition, the subjects' ability to discriminate the stimuli was tested by an ABX procedure. This was done after the discrimination training had been completed.

The authors point out that there are peaks in the discrimination functions. They say, in effect, that these peaks show that they have produced categorical perception by their training procedures.

There are three main objections to this experiment and to the interpretation the authors put upon it.

(1) The essential control is missing: there are no data to tell us how these stimuli were responded to without any training. Given this astonishing omission, one cannot possibly explain either the level or shape of the discrimination functions CLS obtained after training. Were there, in fact, any effects of the discrimination training? And if the peaks were produced by training, are we to suppose that they resulted from a sharpening of discrimination in the center of the stimulus range or a reduction of discrimination at the extremes?

In an attempt to answer these questions, we repeated the experiment with the appropriate control and the same training.
procedure, though under somewhat different viewing conditions. The discrimination functions we obtained before the training were essentially flat and at a high level. But they were also flat, and at the same level, after application of the CLS training procedures. In other words, we found no effect of the CLS procedures. We are, of course, reluctant to draw conclusions from this negative result, the more so because of the difference in viewing conditions. At the very least, however, we should suppose that the effects CLS show must be extremely labile and sensitive to small variations in procedure.

(2) The difficulty of interpreting the CLS results is further increased by the procedures they employed. Though they used nine stimuli in the first experiment, they cut the number to four in the second. As a result, their discrimination functions consist of only three points. These provide a rather meager basis for comparing the obtained discrimination functions with those that are predicted by the assumption of categorical perception, or for evaluating that fit in relation to the fit obtained in our various studies of speech.

(3) Indeed, CLS ignore the need for a comparison of discrimination and identification functions as a basis for claiming categorical perception. We said earlier that CLS were wrong to suppose, in their first experiment, that such perception is indicated by the steepness of the identification function; now, in the second experiment, they are wrong in pointing to the peaks in the discrimination functions as evidence that they have achieved it. It is, of course, true that were there is categorical perception there will be peaks in discrimination functions at the boundaries between categories. But peaks in discrimination, and especially relatively small ones, can presumably be produced by a variety of conditions. As we pointed out earlier, categorical perception of continuously variable stimuli means that the subject can dis-
criminate no better than he can identify absolutely. In evaluating the extent to which the perception of certain consonants approximates that extreme -- and to us interesting -- condition, we have regularly applied a test which permits a comparison of the obtained discrimination functions with those that would have occurred if, indeed, discrimination had been no better than absolute identification.

Such a test is needed here, also. Since the test was not made by CLS, we have applied it to the available data, despite the limitations we have already noted. So far as we can tell, their subjects discriminated the stimuli far better than they identified them absolutely -- in striking contrast to the results we obtained on the perception of stop consonants.

In this reply we have touched on a number of defects in the CLS experiments. We should say again that one can hardly interpret the data, even at the simplest level, because essential controls are missing. Beyond that, equally serious difficulties arise. The attempt to test the motor theory of speech perception by having subjects assign different phonemic labels to visual stimuli was based on erroneous assumptions and led to a pointless experiment. The more general attempt to duplicate, with these same visual stimuli, the near-categorical perception previously observed with stop consonants was inconclusive, to say the least. Though the authors make much of their response latencies, these data are irrelevant to categorical perception. And in claiming that the identification and discrimination functions they obtain with the visual stimuli match those obtained with speech, CLS fail to take into account that it is a relation between identification and discrimination that characterizes categorical perception and should be used to test for its occurrence.

There are broader issues as well. One that we have merely alluded to in this reply is that the evidence for special mecha-
nisms in speech perception goes well beyond the phenomenon of catego-

erical perception.\(^3\) Thus, the attempt by CLS to demonstrate that this phenomenon could be duplicated by discrimination train-
ing would have provided, at most, an incomplete basis for their sweeping conclusion that "the postulation of a special perceptual mechanism for speech perception is not warranted." In the event, even that basis has yet to be established; CLS have not demonstra-
ted that they can duplicate, by discrimination training procedures, the near-categorical perception that is so striking a feature of the response to some of the sounds of speech.

References

Bastian, J., Harris, K.S., and Liberman, A.M. Identification, disc-

rimination and mimicry of a phonemic contrast determined by si-

tent interval duration. (In preparation).


Cross, D.V., Lane, H.L. and Sheppard, W.C. Identification and discrimination functions for a visual continuum and their rela-
tion to the motor theory of speech perception. J. exp. Psychol., 1965, 70, 63-64.


Footnotes

1. Also at the University of Connecticut.
2. Also at Barnard College, Columbia University.
3. A thorough evaluation of the experiments by CLS, or more generally, of their argument that speech perception requires no special mechanism, would take account of omissions and misinterpretations of our work, which are not dealt with here. These points concern the full range of evidence on which the motor theory rests, the unique efficiency of speech sounds, the encoded character of the cues for speech perception, the psychologically important differences between categorical and continuous perception, and other general considerations having to do with perception and language. But the inadequacies of the CLS paper in these matters are repeated in a separate and longer paper by Lane (1965a) so we shall deal with them in a separate and longer reply now in preparation. Here we shall confine ourselves to the particular points that concern most directly and specifically to the CLS experiments.
4. In discussing Fig. 5, CLS point out the necessity of adjusting their data and the consonant data to permit a direct comparison of the degree of categorical perception obtained in the two cases. However, the procedure they use serves only to adjust for different numbers of sample points (ordinates); it does not adjust for differences in psychological distance between category centers.
5. We have more than an academic interest in categorical perception because it is, in our view, an important condition underlying the efficiency of some speech sounds as vehicles of information transmission.
6. That Lane fails to understand the significance of this feature of language and of the motor theory is clear in his recent critique (Lane, 1965a). There he presents in schematic form an in-
terpretation of our motor theory which is erroneous in that it contains no indication that the consequences of the mimicking response can in some sense be matched to the stimulus. Either Lane has failed to grasp the theory or else he assumes, wrongly, that this aspect of it is of no importance.

It was, to be sure, a stated objective of the CLS experiment to avoid stimuli which "are generated by their identification responses." In choosing to use visual stimuli and responses characterized by phonemic contrasts, CLS seem, however, to be making a most remarkable assumption, namely, that the intermediate visual stimuli will somehow evoke auditory or other intermediates which become the entities involved in the perceptual processing. Such an assumption seems inescapable, for, without it, the intermediate visual stimuli would merely generalize to one or the other of the visual anchors and be named accordingly, quite without regard for the "topography of the identification responses" or, for that matter, the objective of the experiment.

7. CLS say that this aspect of their experiment is also intended to make more meaningful certain comparisons between consonant and vowel perception. That assertion is based on two misinterpretations (in addition to the special assumption mentioned in Fn.6). One concerns the nature of the relation between mimicking response and stimulus that we have referred to above. The other has to do with their misunderstanding of the contrast between the nearly categorical perception of the stops and the nearly continuous perception of the vowels. They say that this contrast is meaningless because the stimulus scales are not comparable. Lane develops this misunderstanding at greater length in another paper (Lane 1965a); in order to save space here, we shall deal with his misunderstanding in a separate reply.

8. Some of these issues, and our reason for deferring a discussion of them, are noted briefly in Fn. 3.
NUMBER OF "small" RESPONSES

IDENTIFICATION FREQUENCY FUNCTIONS FOR INDIVIDUAL SUBJECTS. (COMPARE CLS FIG. 1)