AN EFFECT OF LEARNING ON SPEECH PERCEPTION: THE DISCRIMINATION OF DURATIONS OF SILENCE WITH AND WITHOUT PHONEMIC SIGNIFICANCE

ALVIN LIBERMAN, KATHERINE SAFFORD HARRIS, PETER EIMAS, LEIGH LISKER, and JARVIS BASTIAN
Haskins Laboratories, New York

Reprinted from
LANGUAGE AND SPEECH,
Vol. 4, Part 4, October-December 1961, pp. 175 · 195

AN EFFECT OF LEARNING ON SPEECH PERCEPTION: THE DISCRIMINATION OF DURATIONS OF SILENCE WITH AND WITHOUT PHONEMIC SIGNIFICANCE*

ALVIN LIBERMAN,** KATHERINE SAFFORD HARRIS, PETER EIMAS,**
LEIGH LISKER,*** and JARVIS BASTIAN****
Haskins Laboratories, New York

Discrimination of an acoustic variable (various durations of silence) was measured when, as part of a synthetic speech pattern, that variable cued a phonemic distinction and when the same variable appeared in a non-speech context. In the speech case the durations of silence separated the two syllables of a synthesized word, causing it to be heard as *rabid* when the intersyllabic silence was of short duration and as *rapid* when it was long. With acoustic differences equal, discrimination proved to be more acute across the /b,p/ phoneme boundary than within either phoneme category. This effect approximated what one would expect on the extreme assumption that the listeners could hear these sounds only as phonemes, and could discriminate no other differences among them; however, the approximation was not so close as for certain other consonant distinctions.

In the case of the non-speech sounds the same durations of silence separated two bursts of noise tailored to match the onset, duration, and offset characteristics of the speech signals. There was, with these stimuli, no appreciable increase in discrimination in the region corresponding to the location of the phoneme boundary. If we assume that the functions obtained with the non-speech patterns represent the basic discriminability of the durations of silence, free of the influence of linguistic training, we may conclude that the discrimination peaks in the speech functions reflect an effect of learning on perception. It was found, too, that discrimination of the non-speech patterns was, in general, poorer than that of the speech. From this we conclude that the effect of learning must have been to increase discrimination across the phoneme boundary; there was no evidence of a reduction in discrimination within the phoneme category.

In studies of the perception of /b,d,g/, /d,t/, and /sl,spl/[1] (Liberman, Harris, Hoffman and Griffith, 1957; Griffith, 1958; Bastian, Eimas and Liberman, 1961; Liberman, Harris, Kinney and Lane, 1961) we have found that discrimination of a

* *This research was supported by the Operational Applications Office of the Air Force Electronic Systems Division in connection with Contract AF 19(604)-2285.*
** *University of Connecticut.*
*** *University of Pennsylvania.*
**** *Now at Center for Advanced Studies in the Behavioral Sciences, Stanford, California.*

[1] *These phonemes were presented in contexts as follows: /b,d,g/ were in absolute initial position before /a/; /d,t/ were in absolute initial position before /o/; /sl, spl/ were in the words* slit *and* split.

given acoustic difference is considerably more acute across phoneme boundaries than in the middle of the phoneme categories. To make the appropriate measurements we had first to identify acoustic variables which are sufficient cues for the perceived phonemic distinctions. For that we were able to fall back on the results of earlier studies (Liberman, Delattre, Cooper and Gerstman, 1954; Delattre, Liberman and Cooper, 1955; Liberman, Delattre and Cooper, 1958; Harris, Hoffman, Liberman, Delattre and Cooper, 1958; Bastian, Delattre and Liberman, 1959). Having thus selected for each phonemic distinction an appropriate acoustic cue, we prepared a series of synthetic patterns in which that cue was varied along a single continuum through a range large enough to encompass the phonemes being investigated. To measure discriminability, we arranged the patterns into ABX triads and asked the listeners to decide, on the basis of any similarities or differences they could hear, whether X was identical with A or with B. (A and B were, in fact, always different, and X was always identical with the one or the other.) To find the phoneme boundary, we presented the patterns with instructions to identify each one as /b/, /d/, or /g/ in the first experiment, as /d/ or /t/ in the second, and as /sl/ or /spl/ in the third.

Discrimination was so much better across the phoneme boundary than within the category as to suggest that the listeners could only hear these consonants categorically (i.e., as phonemes), and could discriminate no other differences among them. This suggestion was tested by finding the extent to which the discriminability of the patterns could be predicted from the way in which the listeners had assigned the stimuli to the various phoneme categories. Discrimination functions that were derived on this basis were found to fit rather closely those that had been obtained in the experiments.[3]

From a psychological point of view these results are quite unusual. With stimuli that vary along a single dimension (of frequency, intensity, or duration, for example) one typically finds that subjects discriminate many times more stimuli than they can identify absolutely (Pollack, 1952, 1953; Garner, 1953; Chapanis and Halsey, 1956; Miller, 1956). In everyday experience this is illustrated by the contrast between the ease with which we normally distinguish two tones as being of different pitch and, on the other hand, the great difficulty we have in absolutely specifying the pitch of either one. The very different result in the perception of /b,d,g/, /d,t/, and /sl, spl/ was that discrimination was little better than absolute identification. It is as if our listeners were able to distinguish only as many pitches as they could correctly name.

Viewed from a linguistic standpoint, these results might not appear surprising. Apparently the linguist is prepared to find, with some phonemes at least, that variations in a speech sound will be heard by phonetically naive listeners only when these variations are phonemic. More generally, he might see the extent to which this happens as a precise expression of the degree to which linguistic categories are also categorical in perception.

[3] *In the case of /b,d,g/ the fit was better in the Griffith (1958) study than in the experiment by Liberman* et al. *(1957). This was attributed to the fact that Griffith's synthetic speech stimuli were more realistic, and to certain procedural improvements he was able to make.*

Within either a psychological or linguistic framework, the peaks in discrimination should be of interest, we think, because their existence may be an important condition underlying the distinctiveness of speech sounds. Thus, an incoming stimulus which falls ever so slightly to one side of the peak becomes indistinguishable from, and no harder to identify than, a stimulus which lies in the precise centre of the phoneme region. The effect of this should be to reduce the area of uncertainty between phonemes, thereby increasing the accuracy and speed with which the listener sorts the various sounds of speech into the appropriate phoneme bins.

We should note that there appears to be considerable variation among phoneme classes in the size and sharpness of the discrimination peaks, and, correspondingly, in the extent to which the perception is categorical. For some phoneme distinctions there are no discrimination peaks at the phoneme boundaries, and the level of discrimination is far better than would be predicted from the extreme assumption that the listeners can hear the sounds only as phonemes. This kind of result has so far been found in the perception of vowels (Fry, Abramson, Eimas and Liberman, in preparation), and of several prosodic features (tones and vowel duration) which are phonemic (Abramson, 1961 ; Abramson and Bastian, in preparation).

To determine why the discrimination peaks develop at some phoneme boundaries and not at others, we should have to inquire quite deeply into the nature of the perceptual mechanism. For the purposes of this paper it is appropriate only to indicate the broad outline of our hypothesis. We believe that in the course of his long experience with language, a speaker (and listener) learns to connect speech sounds with their appropriate articulations. In time, these articulatory movements and their sensory feedback (or, more likely, the corresponding neurological processes) become part of the perceiving process, mediating between the acoustic stimulus and its ultimate perception. When significant acoustic cues that occupy different positions along a single continuum are produced by essentially discontinuous articulations (as, for example, in the case of second-formant transitions produced for /b/ by a movement of the lips and for /d/ by a movement of the tongue), the perception becomes discontinuous (i.e., categorical), and discrimination peaks develop at the phoneme boundary. When, on the other hand, acoustic cues are produced by movements that vary continuously from one articulatory position to another (as, for example, the frequency positions of first and second formants produced by various vowel articulations), perception tends to change continuously and there are no peaks at the phoneme boundaries.. Various aspects of our view have been described elsewhere (Cooper, Delattre, Liberman, Borst and Gerstman, 1952 ; Liberman, Delattre and Cooper, 1952 ; Liberman, 1957 ; Cooper, Liberman, Harris and Grubb, 1961 ; Bastian, Eimas and Liberman, 1961 ; Harris, Bastian and Liberman, 1961), and it will be developed further in papers now in preparation. A theory which is in certain ways related to ours has been put forward in an interesting paper by Ladefoged (1959).

Basic to our speculation about the mechanism which accounts for the discrimination peaks is the simple assumption that they are learned. The primary purpose of the experiments to be reported here is to provide data relevant to that assumption.

Whether the peaks are, in fact, acquired in experience, or whether they are somehow a part of our innately given sensitivity to the acoustic stimuli, is a question of broader scope than is a consideration of any particular mechanism as such.

We cannot dismiss, out of hand, the possibility that the discrimination peaks are innately given. If they are, we should suppose that the earliest speakers of the language wisely chose to locate the phoneme boundaries in the regions of highest discriminability. Assuming, alternatively, that the peaks reflect the learning that has occurred during each listener's long experience with the language, we must then answer a further question concerning the direction the learning has taken. Thus, it is possible that the peak is an increase in discrimination, acquired as a result of the listener's having had for many years to distinguish sounds that lie on opposite sides of the phoneme boundary. Such an effect might prove to be similar to what has been called "acquired distinctiveness"; for convenience, we will use that term to describe it. The contrary, and equally likely, possibility is that the peak is what remains after discrimination has been reduced by long training in responding identically to sounds that belong in the same phoneme class. This would likely be counted an example of "acquired similarity". It is also possible, of course, that the observed effect is the sum of both processes: acquired distinctiveness across the boundary and acquired similarity within the category.

As between these two assumptions—that the peaks are part of the listener's innately given sensory equipment, or, alternatively, that they are the result of learning—the latter interpretation is the more likely. One relevant consideration is that languages other than English have apparently located their phonemes differently on the acoustic continua with which we are here concerned. Although there are no data yet available to show that the inflections in the discrimination function are displaced to correspond with the different positions of the phoneme boundaries, the mere fact that the boundaries are differently located is, in itself, presumptive evidence that the highs and lows of the discrimination function are not innately given.

A learning interpretation is also favoured by the fact that the discriminations among synthetic /b,d,g/, /d,t/, and /sl,spl/ were so largely controlled and limited by the phoneme labels. As was pointed out above, this relatively close correspondence between differential sensitivity and absolute identification is in striking contrast to the usual psychophysical result. One may suspect that it has come about because the original or raw discriminations have been radically altered by long experience.

The experiment by Griffith (1958) on /b,d,g/, which we referred to earlier, has provided additional relevant evidence. Using essentially the same second-formant transitions employed by Liberman *et al.* (1957), he added one or another of two constant third-formant transitions which had the effect of changing the positions of the phoneme boundaries. The result was that the peaks and valleys of the discrimination functions shifted accordingly. Though not critical, this evidence strongly supports a learning interpretation.

The experiment with /d,t/ that was referred to earlier was also of a type designed to find out whether the observed peak in discrimination is a result of learning, and,

if so, whether it is a case of acquired distinctiveness, acquired similarity, or both. The point of this kind of experiment is to measure the discriminability of an acoustic variable which cues a phonemic distinction, and then to measure the discriminability of essentially the same variable in a non-speech context. For /d,t/ the acoustic variable was the time of onset of the first formant relative to the second and third. It was found, as has already been pointed out, that discrimination of this variable was better across the phoneme boundary than within the phoneme category. To produce appropriate non-speech controls the experimenters simply inverted the speech patterns on the frequency scale, thus producing sounds which could not be perceived as speech while yet preserving quite exactly the acoustic variations that had, in the speech stimuli, cued the perceived linguistic distinction. In the discrimination of the control stimuli no peak in discrimination was found, either in the region corresponding to the location of the phoneme boundary or, indeed, in any other part of the stimulus continuum. This would indicate that the discrimination peak found with the speech stimuli is to be attributed to learning. It was apparent, further, that the discriminability of the non-speech controls was, at all points, below that of the speech. From this one would conclude that the learning effect consisted entirely of acquired distinctiveness.

The inverted patterns were not a perfect control. Nor was it possible that they could have been, since the ideal condition would have required that the controls be identical with the speech stimuli and yet not be perceived as speech. The control stimuli that were used in the experiment on /d,t/ had the salient shortcoming that the frequency of the formant whose time of onset varied was below the other two formants in the speech stimuli and above them in the controls. Since masking effects are greater from low frequencies to high, it is possible that the variations in onset were to some extent masked out in the control.

The specific purpose of the present experiment is to obtain data from an experiment analogous to the one just described, but with a more appropriate control. The linguistic distinction to be investigated is that between /b/ and /p/ in intervocalic position, specifically in the words *rabid* and *rapid*. In studying the perception of this distinction Lisker (1957; in preparation) has found that a sufficient cue—not the most important one, perhaps, but one that is nonetheless adequate—is the duration of the silent interval between the first and second syllables. When that interval is relatively short the listener hears *rabid*; increasing the interval causes the perception to change to *rapid*. The discriminability of such patterns, differing only in duration of the intervocalic silent interval, can, of course, be measured and then compared with the discriminability of the same durations of silence enclosed between two bursts of noise. These latter stimuli are particularly appropriate as non-speech controls, since they can be identical with the speech sounds, not only in the values of the stimulus variables (i.e., the duration of the silent interval), but also in regard to certain important constant aspects of the stimuli, such as the durations and amplitude envelopes of the sounds that bound the silent intervals. Comparing the discriminability of these speech and non-speech stimuli should help greatly to determine whether the

discrimination of the speech sounds reflects the effects of learning, and, if so, to discover which direction the learning has taken.

Procedure

One set of stimuli was generated from a hand-painted spectrogram like that shown in Fig. 1. When converted to sound by the Pattern Playback, this spectrogram produces a reasonable approximation of the word *rabid*. From Lisker's research (in preparation), we know that if the interval of silence between first and second syllables is made longer than that shown, the listener will hear *rapid*. To vary the duration of the silent interval and thus produce a series of sounds which would be perceived as *rabid* at one end and as *rapid* at the other, we made numerous magnetic tape recordings of the sound produced from the spectrogram shown in the figure, cut the magnetic tape in each case so as to separate the two syllables, and then inserted appropriate lengths of blank tape. In this way we created a series of 12 stimuli in which the silent interval varied between 20 and 130 msec. in steps of 10 msec. For convenience, we will refer to this set of sounds as the "speech stimuli" and designate each member of the set by the duration of the silent interval. Thus, Speech Stimulus 40 is that pattern which has 40 msec. of silent interval between the first and second syllables.

Our own listening convinced us that these particular stimuli did, indeed, sound like *rabid* or *rapid*, and that the shift from the one to the other occurred in the vicinity of 70 msec. Not unexpectedly, it appeared further that the longest silent intervals produced stimuli that would surely sound odd and unrealistic to speakers of English. We nevertheless included these extreme durations because we wanted to make certain that complete psychophysical functions would be obtained with the possibly less discriminable control stimuli to be described below.

It should be noted here that it is possible to begin with a recording of *rapid* as spoken by a human being and then, by reducing the duration of the interval between syllables, to convert it to *rabid*. The conversion is not wholly convincing, however, because there are several other cues to voicing besides the duration of the silent interval, and these are not changed as the interval is lengthened or shortened. Synthetic speech has the advantage here that it is possible to neutralize all the cues except the duration of the silent interval, and thus to produce a more satisfactory set of stimuli.

It was indicated in the introduction that we wanted as control stimuli a set of sounds as similar as possible to the speech series, and yet not perceivable as speech. To obtain such controls we used the speech stimuli to modulate noise signals and thus produce patterns consisting of noise-silent interval-noise in which the durations and rates of turn-on and turn-off would be the same as in the speech stimuli. The equipment and procedure for producing the control stimuli were as follows:[3]

The original speech stimuli were modulated by a 10 kc. carrier in a balanced

[3] *We gratefully acknowledge our debt to Dr. Carl Brandauer for devising the method of generating the control stimuli.*

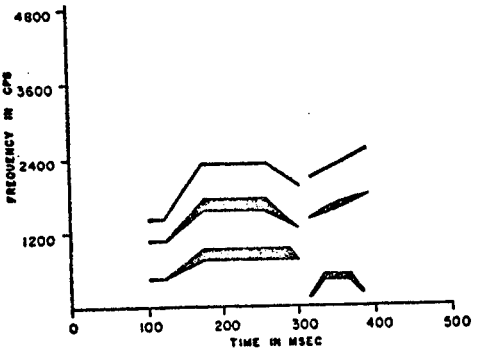

Fig. 1. Hand-painted spectrogram from which the stimuli of the experiment were produced.

modulator. The modulated signal was half-wave rectified, put through a low-pass filter (150 cps. cut-off), and this envelope waveform was then used to modulate a band-limited white noise (d.c. to 1500 cps.) in another balanced modulator. The circuit parameters were adjusted to give the best possible match to the envelopes of the original set of speech stimuli.

The extent to which we succeeded in matching speech and control stimuli was measured in two ways. First, we made a detailed visual comparison of the oscillograms of several pairs of stimuli. The envelopes of the speech and control stimuli were found to be very similar. As a second check, we made spectrograms on a Kay Sonograph of 36 pairs of speech and control stimuli and measured the duration of each silent interval. The averages for the two kinds of stimuli proved to be almost exactly the same. The variability of the control stimuli was somewhat greater, as we might expect, but the difference was not significant by an F test.

Subjects

All subjects in the experiment were undergraduate or graduate students at the University of Connecticut. They were paid volunteers with no special training in phonetics, and they were naive with respect to the purposes of the experiment.

There were 12 subjects in all, chosen from a group of 31 on the basis of a pre-test. The purpose of the selection was to insure that all subjects ultimately serving in the experiment would have a sharp and clear phoneme boundary.[4] In the pre-test, the

[4] *Since the experiment was intended to yield information on the relative discriminability of sounds within and across phoneme boundaries, only subjects with sharp boundaries were suitable.*

group of 31 subjects listened to the stimuli in various orders (approximately 28 presentations per stimulus) under instructions to identify each stimulus as *rabid* or *rapid*. On the basis of the data so obtained, we selected for service in the experiment the 12 subjects who had the sharpest phoneme boundaries. The results from this pre-test session were not otherwise used, and the data are not presented in the Results section. It should be noted, however, that all of the original group of 31 did reasonably well—so much so that the difference between the selected and rejected groups was very small.

Stimulus Presentation

As in the previous experiments in this series, the discriminability of both speech and control stimuli was measured by an ABX procedure—that is, the stimuli were presented in groups of three, and subjects were asked to determine, by whatever cues they could perceive, whether the third stimulus, X, was identical with the first stimulus, A, or the second stimulus, B. (In fact, X was always identical either with A or with B.) The measure of the discriminability of any pair of stimuli was, then, the proportion of the presentations on which the subject matched X correctly to A or B.

The A and B stimuli to be discriminated differed in silent interval by 20, 30, 40, 50, 60, 70, 80 and 90 msec. For example, Stimulus 20 was compared with Stimulus 40, 50, 60, 70, 80, 90, 100 and 110 ; Stimulus 30 was paired with Stimulus 50, 60, 70, 80, 90, 100, 110 and 120. The 10 stimulus comparisons in which pairs differ by 20 msec. will be called the 2-step series, the series of 9 pairs that differed by 30 msec. will be called the 3-step series, and so forth.

The total number of stimulus pairs is 52. Each stimulus comparison appeared in ABX triads of four forms—ABA, ABB, BAA, and BAB. For example, the four two-step comparisons for the 40-msec. stimulus were 40-60-40, 40-60-60, 60-40-40, and 60-40-60.

These triads were spliced together so as to form four test orders. Each stimulus comparison occurred once in each test order in one of its four forms. One second separated the members of a triad, while the triads were separated by four seconds. After the first four orders had been completed, a second set of four was made by shuffling each of the original orders. There were, then, eight test orders for the measurement of speech discrimination.

The non-speech stimuli were made into eight tapes in exactly the same fashion. That is, for each speech tape we made a control tape with a non-speech stimulus substituted for the speech stimulus having the same time separation between syllables.

The selected subjects listened to each of the eight speech and control tapes five times under discrimination instructions. Since a given stimulus comparison is presented once on each tape, the measure of discriminability at any point on each stimulus continuum is based on 40 determinations for each subject.

The purposes of the experiment required a comparison of the speech stimuli within and across the *rabid-rapid* phoneme boundary. Accordingly, we needed an accurate determination of the location of the boundary for the 12 subjects of the experiment.

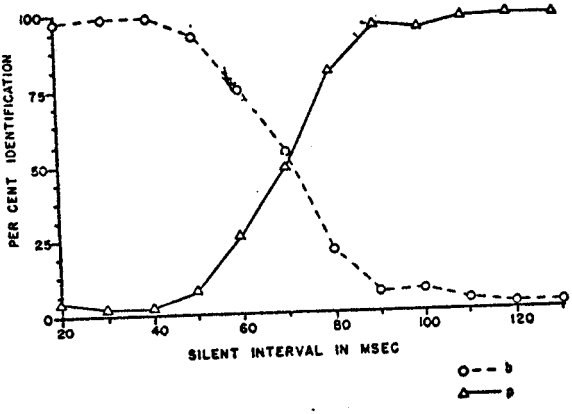

Fig. 2. Identification of the synthetic speech stimuli as /b/ or /p/, plotted against the duration of the silent interval between first and second syllables. The data are from the pooled responses of all 12 subjects.

To this end, the speech tapes were presented to each subject a total of 32 times with instructions to label each stimulus as *rapid* or *rabid*.

The whole experimental design, then, was set up so that each subject would perform three tasks: speech discrimination, noise discrimination, and phoneme labelling. A schedule was arranged for each subject such that the three tasks were distributed through all experimental sessions. Working in test sessions of about 20 minutes, each subject took about four months to complete the experiment.

Results

Phoneme Identification and the Location of the Boundary

Fig. 2 shows how the listeners assigned the phoneme labels /b/ or /p/ to the various stimuli. These functions which represent the pooled responses of all subjects, indicate that the phoneme identifications were made with fair consistency, and that the boundary between /b/ and /p/ lay at about 70 msec. of silent interval. It is also apparent from these data that the phoneme boundary is reasonably sharp.

Discrimination of the Speech Stimuli

The solid lines of Fig. 3 connect points which represent the percent correct discrimination of various pairs of speech stimuli at all values of the stimulus variable. As in the labelling functions of Fig. 2, the data from all subjects have been pooled.

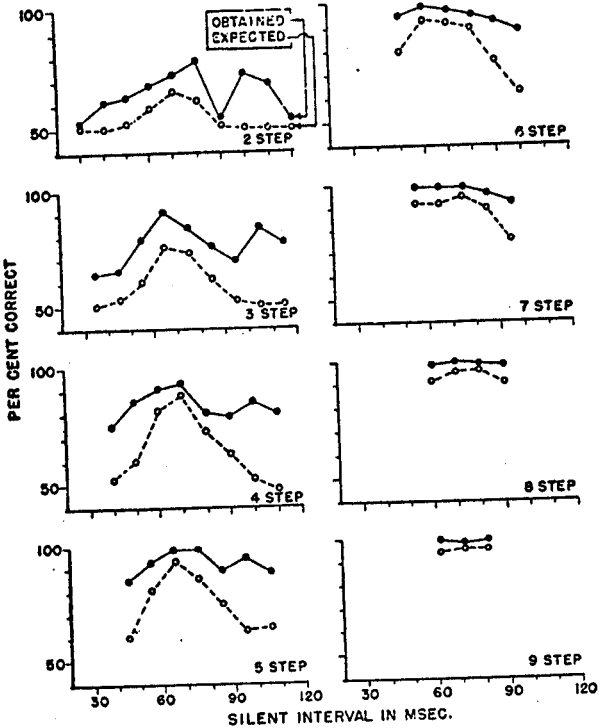

Fig. 3. Obtained and expected discrimination for the 1- through 9-step differences among the synthetic speech stimuli. The data are from the pooled responses of all 12 subjects.

For greater ease in reading the data, the graphs have been separated according to the amount of difference in silent interval by which the two stimuli are separated. Thus, in the first graph at the upper left, labelled "2 step," the stimuli which were paired for discrimination (in ABX triads) always differed by 20 msec. (two steps on the stimulus scale) of silent interval. The first point on this graph indicates, then, that the subjects discriminated with 53% accuracy when the stimuli in the ABX triads had 20 and 40 msec. of silent interval. The next point shows that discrimination rose to 61% for the stimuli with 30 and 50 msec. of silent interval. Data points on

the graphs for the other stimulus comparisons, in which the differences between the stimuli ranged from three to nine steps, are to be read in similar fashion.

It is apparent, especially in the 2-, 3-, and 4-step graphs, that there are two peaks in the discrimination functions, a relatively large one near the centre of the stimulus continuum and a somewhat smaller one farther to the right. For the moment we will confine our attention to the larger peak.

Reference back to the phoneme identification data in Fig. 2 reminds us that the phoneme boundary is in the vicinity of 70 msec. and we see in Fig. 3 that the larger peak in the discrimination function occurs in this same region. This is to say, of course, that discrimination is better across the phoneme boundary than in the middle of the phoneme category. But instead of developing this comparison stimulus by stimulus, we will turn to a simple model developed in an earlier study (Liberman *et al.*, 1957) of this same problem, and evaluate the data with regard to the extent to which they fit that model.

Make the extreme assumption that the listeners can only hear these stimuli phonemically—that is, as /b/ or /p/—and can detect no other differences among them. Using the phoneme labelling data as a basis, one then predicts the accuracy with which the listener can be expected to discriminate all stimulus pairs. Thus, if the subject had always identified two stimuli as being members of the same phoneme class, he would be expected to discriminate the stimuli at a chance level. To the extent that he identifies two stimuli as belonging in different phoneme classes, he would, to precisely that extent, correctly discriminate them. In general, this assumption will predict peaks in the discrimination function wherever there are abrupt changes or inflections in the phoneme labelling curves, the height of the peak being a function of the abruptness and extent of the shift in phoneme labels. A more detailed description of the model and the derivation of the technique for predicting the discrimination functions are to be found in the earlier article (Liberman *et al.*, 1957).[5]

The discrimination functions that are predicted from the assumption of categorical perception are shown in Fig. 3 as the dashed lines. A comparison of these "expected" curves with the discrimination data that were actually obtained, and described earlier, leads to several conclusions. First, it will be noted that the left-hand portions of the two curves are reasonably similar in shape. This means that the variations in discriminability follow the change in phoneme labels. More specifically, it means that a discrimination peak does, indeed, occur at the phoneme boundary. The second conclusion is that the obtained functions tend in general to lie above the expected functions. To that extent the listeners are able in discriminating these stimuli to extract some information in addition to that which is revealed by the way in which they label the stimuli as phonemes.

[5] *In the earlier experiment (Liberman et al., 1957) the labelling data were obtained by presenting the stimuli one at a time; in the present experiment the stimuli were presented for labelling in triads (see Procedure). Although the labelling functions were obtained by these different procedures, the calculation of the predicted discrimination values were made from the labelling data in exactly the same way in the two studies.*

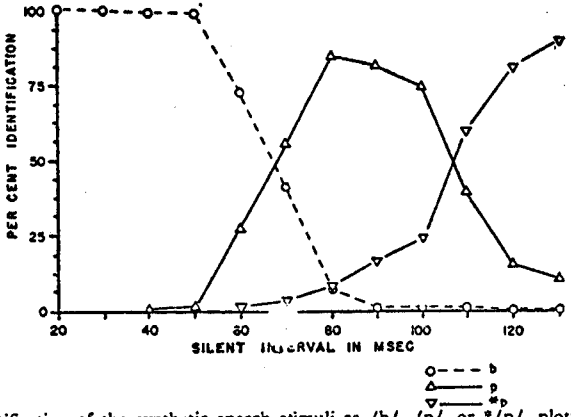

Fig. 4. Identification of the synthetic speech stimuli as /b/, /p/, or */p/, plotted against the duration of the silent interval. The data were obtained from the pooled responses of the seven subjects who served in this part of the experiment.

The Second Peak and a Third Category

Attention has been called to the fact that the obtained discrimination functions show a second peak at values of the stimulus variable greater than those at which boundary between /b/ and /p/ is located. On the assumption that this might imply a third category, we listened again to the stimuli and found that we had, indeed, carried the values of the silent interval to such extreme lengths as to have created, perhaps, an additional class of sounds.[6] This was a strange and unnatural /p/ to American ears, but we thought that it might nevertheless be heard and articulated by our listeners almost as if it were a different speech entity. We therefore recalled as many of the subjects as were still available (the number was seven), discussed the stimuli with them, and discovered that they, too, thought that some of them belonged in a separate class. To obtain more information about this third category we presented the stimuli (in random order, as before) to these subjects with instructions to identify each one as /b/, /p/, or */p/, the last named being the designation we chose for what we, and our subjects, had heard as the unnatural /p/. That the third category, */p/, did exist for these subjects is indicated by the graphs of Fig. 4. A comparison of these graphs with those that describe the results of the two-category judgments (Fig. 2) indicates that the distinction between */p/ and /p/ is not so clear as that between /b/ and /p/ in the original judgments (for which the subjects were allowed only the /b/ and /p/ categories). This is to be inferred from the fact that the curves

[6] *As was pointed out under Procedure, we were aware when we produced these speech sounds that curious effects could be heard at durations of silent interval greater than 100 msec., but we had decided to include these extreme stimulus values in order to be certain of obtaining complete psychophysical functions with the control stimuli.*

representing the /p/ and */p/ judgments (in Fig. 4) do not rise to 100% as the /b/ curve does, and as both the /b/ and /p/ curves do in the two-category situation (Fig. 2). Nevertheless, the labelling data show that a */p/ category does exist, and they provide a basis for predicting a new set of discrimination results from the assumption of categorical perception. These predictions are shown as the dotted lines in Fig. 5, together with the discrimination data that were actually obtained with the seven subjects who made the three-category judgments. There is a second peak in the expected discrimination function corresponding to the boundary between the second and third categories of Fig. 4. Moreover, this second peak fits moderately well the second peak in the obtained discrimination functions. Clearly, the expected and obtained functions now agree somewhat better than before, but there remains a constant difference between the functions in the direction of better discrimination than is to be expected on the extreme assumption of categorical perception. We will not try to specify the magnitude of the discrepancy in terms of some single meaningful quantity, because we are not prepared at this stage to decide which of several possible measures is best. We will only say that the discrepancy is somewhat greater here than it was in the studies of /b,d,g/, /d,t/, and /sl,spl/, where the perception was more nearly categorical; also, that it is less than in the vowels and prosodic features, where perception was essentially non-categorical, i.e., continuously changing with progressive changes in the stimulus.

In terms of the theory outlined in the introduction, perception of speech becomes linked to the feedback from the articulatory movements the listener makes in speaking. We should expect, then, that perception would be completely categorical (i.e., that the discrimination functions would show a peak at the phoneme boundary and be perfectly predictable from the phoneme labelling data) if the listener makes exactly the same articulatory response to the various stimuli to which he attaches the same phoneme label, and very different articulatory responses to sounds he calls by different phoneme names. At the other extreme, speech perception would be expected to be perfectly continuous (i.e., the discrimination function would show no peak at the phoneme boundary and might lie at a level far higher than that which is predicted from the phoneme labelling data) if the listener's mimicking articulations change in linear fashion with variations in the acoustic stimuli, both within and across phoneme boundaries.

One might expect something like the results we found in this experiment—perception which is almost categorical, but not quite—if it be the case that the articulatory response changes most rapidly at the phoneme boundary, but that there is, nevertheless, some small variation within the phoneme class. In an attempt to find out whether or not this was so, we had several of the listeners try to mimic the various stimuli, and then undertook to measure the duration of the silent interval between the two syllables. It proved to be difficult to obtain highly reliable measurements, chiefly because the subjects produced variations in other acoustic features, such as the first-formant transition and presence or absence of voicing, which are more important than

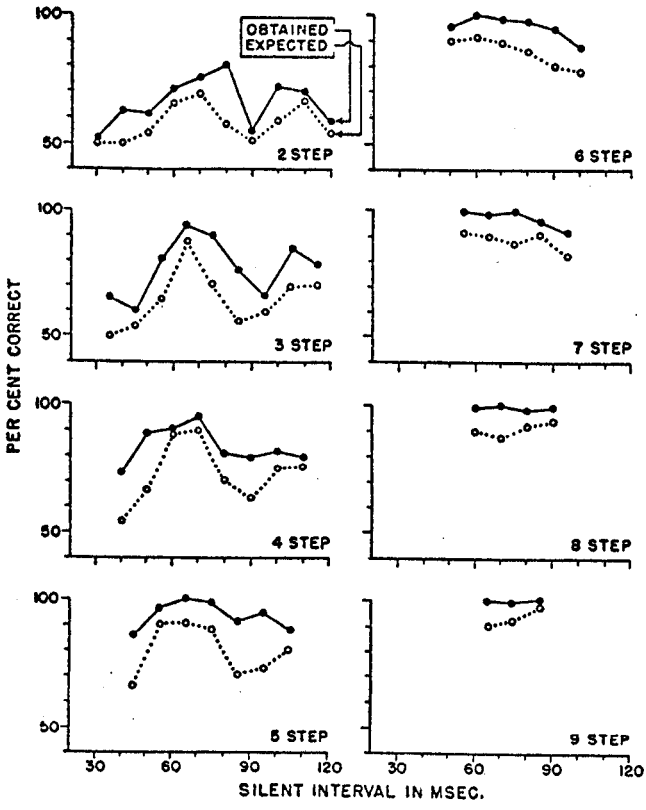

Fig. 5. Expected discrimination functions corrected to take account of the subject's identification of the third category, */p/, together with the obtained discrimination functions. The data are from the pooled responses of the same seven subjects whose three-category identification functions are shown in Fig. 4.

duration of silent interval as such, and which tended to obscure it. When, in the course of this work, it became apparent that we would be able to get far more precise mimicry data in the study of other phonemic distinctions,[1] we abandoned the attempt to measure the mimicry of *rabid-rapid*.

[1] *One such study has been completed since this paper was written and has been published as an abstract, see Harris, Bastian and Liberman (1961). It is now being prepared for regular publication.*

In some respects, then, this experiment yielded less than we might have wished. We should remember, however, that we undertook it because, being interested in the discrimination peaks which sometimes occur at phoneme boundaries, we wanted a fair comparison between the discrimination of an acoustic variable when it cues a phonemic contrast and when, in a non-speech context, it does not. Fortunately for that purpose, the discrimination of the speech stimuli does have a peak (or actually two) sufficiently high to make the comparison with the non-speech control an interesting one.

Discrimination of Noise Control

In the stimuli used as controls, bursts of noise served to bound silent intervals that duplicated those of the speech stimuli. It is also relevant to recall that the noise bursts were matched with the speech syllables in regard to such constant features as amplitude envelope and duration.

The discrimination results obtained with the noise control are shown in Fig. 6. For comparison the results obtained with the speech stimuli and previously shown in Fig. 3 are also presented.

One sees immediately that the discrimination peaks of the speech stimuli are much higher and sharper than any peaks which appear in the control data. More generally it is clear that the discriminability of the speech sounds is considerably greater than the control at most points. At a few values of the stimulus variable the two are equal, or very nearly so. Out of a total of 52 points at which the two sets of curves can be compared there is only one at which the speech discrimination dips below the control. In this one case the difference is small, and probably not at all reliable. If the noise stimuli are truly an appropriate control—that is, if they fairly represent the discriminability of the speech stimuli prior to linguistic training—we may conclude that the results obtained with the speech stimuli reflect the effects of a very considerable amount of learning. It is clear, further, that the entire learning effect consists of a sharpening of discrimination in the vicinity of the phoneme boundary. There is no indication that discrimination of the speech has been reduced (below the control) within the phoneme category. In terms of the psychological processes discussed in the introduction, we should say that there is here a very considerable amount of acquired distinctiveness, but no acquired similarity.

Reference has been made to the earlier study of /d,t/, in which discrimination of variations in a cue for a phonemic distinction were compared with discrimination of the same variable in a non-speech context. It was pointed out that the acoustic variations might have been masked in the control stimuli, and the conclusion drawn from a comparison of speech and control discrimination was, therefore, thought to be open to question. We should note that the results of the present experiment, with its more adequate controls, agree with those of the earlier study in that there was evidence of a large amount of acquired distinctiveness and no acquired similarity.

We ought, perhaps, to remark on the fact that this experiment and the earlier one on /d,t/ differ somewhat in regard to the finding of no reduction in discrimination within the phoneme category (acquired similarity). In the earlier experiment the

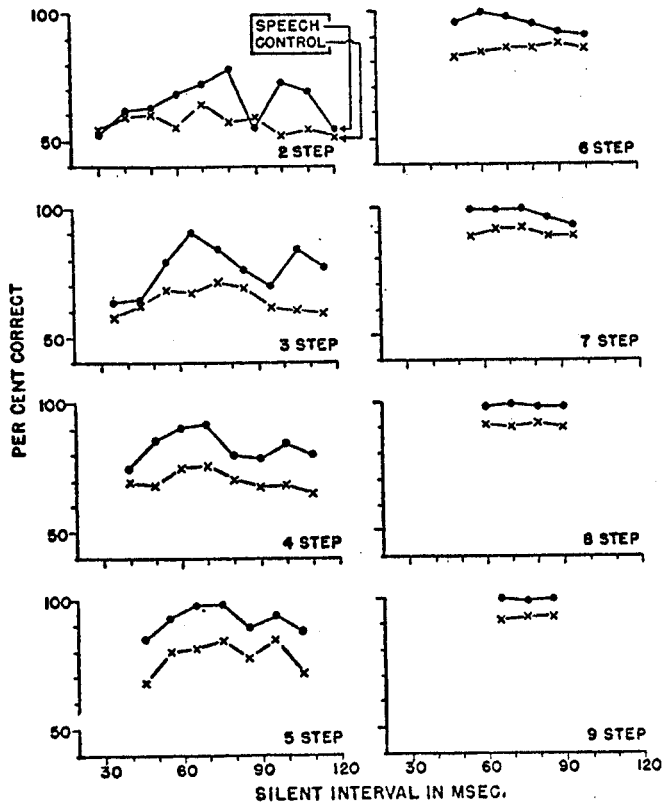

Fig. 6. Discrimination functions for the 1- through 9- step differences among the non-speech control stimuli. The obtained speech discrimination functions previously shown in Fig. 3 are reproduced here to facilitate comparison. For both sets of functions, the data are from the pooled responses of all 12 subjects.

discriminability of the non-speech stimuli was very poor, lying generally at or just slightly above chance. There was, then, no room for a process like acquired similarity to show itself, for no amount of training could possibly have reduced discrimination much further. In the present experiment the discrimination of the noise control stimuli rose to higher levels. To determine just how much room this provided for acquired similarity, we must compare the three discrimination functions previously presented: the obtained discrimination of the speech sounds, the expected discrimination of the speech sounds, and the obtained discrimination of the noise control. For that purpose

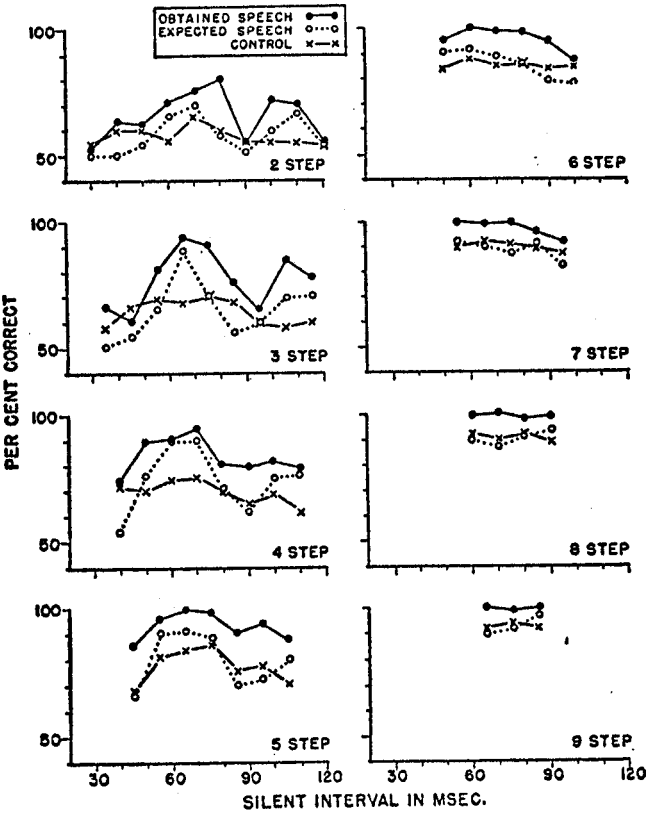

Fig. 7. Obtained and expected speech discrimination functions, together with the discrimination functions for the non-speech control stimuli. The obtained and expected speech functions are the same as those shown in Fig. 5. All data are from the pooled responses of the seven subjects who provided the results shown in that figure.

the three functions are shown together in Fig. 7. (We here use the data for the seven subjects who made the three-category judgments, because these data most adequately depict the relationship between discrimination and phoneme labelling.)

Within the /b/ category—that is, on the left-hand side of the graphs—we see in the 2-, 3-, and 4-step data that the discriminability of the noise does lie somewhat above the predicted discriminability of the speech. This means that the original discriminability of speech may be presumed to have been greater than it needed to be

to meet the requirements of the linguistic situation. It also means that if the listener makes the same articulatory response to these stimuli, we should expect, according to theory, that the discrimination of the speech would have been reduced below the noise control, down to the predicted level. We see that this has not happened, and we conclude that we have here a case in which acquired similarity did not occur, though there was room for it. Beyond the four-step comparisons practically all stimulus pairs fall across a phoneme boundary; discrimination is therefore predicted in general to be at fairly high levels, well above the noise control at all points, and we cannot make the kind of test we are here considering. Between the second and third categories (/p/ and */p/) the noise discrimination again lies above the expected values, and we find, as we did in the similar situation within the /b/ category, that the obtained speech discrimination has not been reduced to the expected level.

While the relevant data are not very clear or compelling, there is a certain amount of evidence that acquired similarity might have occurred but did not. Whether it *should* have occurred, according to theory, is a separate question, and one that is not readily answered because the critical mimicry data are missing. To see why these data are critical, let us imagine that reliable mimicry measurements have been obtained, and then consider the implications of different kinds of results. Suppose first, that the subjects are found to make the same articulatory response in mimicking the sounds within the phoneme category; the theory as it now stands demands that they be unable to discriminate these sounds. Given this outcome of the mimicry experiment, and given the fact that acquired similarity did not occur though there was room for it, we should have to modify the theory. It would appear then that while the articulatory responses become involved in the perception of speech, the listener still has some choice remaining to him: if falling back on the feedback from articulation serves to sharpen discrimination (as it apparently does at the phoneme boundary), the listener takes advantage of this possibility and discriminates better than he would otherwise have done; if the articulatory feedback has the effect of dulling discrimination, the listener effectively ignores it, responds directly to the acoustic stimulus, and suffers no loss in acuity. This would say, in general, that the acquired similarity paradigm describes an event which does not occur, at least not in speech perception; we should assume then that while we may, on occasion, find it necessary to disregard clearly perceptible differences, and for very practical reasons to call distinguishable stimuli by the same phoneme name, we do not as a consequence really lose our ability to discriminate.

The alternative result of the mimicry experiment would be to show that the listener can and does mimic some of the stimulus changes within the phoneme category —that while the articulatory response changes most rapidly at the phoneme boundary, there is nevertheless some variation in mimicking sounds the subject calls by the same phoneme name. In that case, we should not expect discrimination within the phoneme class to be reduced to chance, and, depending on the particular nature of the mimicry results, the theory in its present form might be rather precisely confirmed. We hope

that more light will be shed on this question in research on other phoneme distinctions where mimicry can be more easily measured.

It will be recalled that the noise control stimuli were produced in such a way as to make their amplitude envelopes and durations approximate the speech signals as closely as possible. This is surely the most appropriate way to obtain control signals for use with the speech stimuli of this experiment, but it may raise a question about the extent to which the results represent what would be obtained with the simpler and more regular stimuli that are usual in psychophysical studies. To answer this question, we prepared a new set of control signals which had intervals of silence like those of the original controls, bounded by segments of noise which differed from the original in having abrupt onsets and offsets (produced by cutting the magnetic tape at a 90-degree angle) and in being of equal duration (300 msec.). On testing the discriminability of these stimuli with the same subjects who had served in the earlier parts of the experiment, we obtained results very similar to those found with the original set of controls.

As a further test of generality we undertook to obtain some indication of the effect, if any, of the particular psychophysical procedure (ABX) that had been used. For that purpose we measured the discriminability of the new noise stimuli by the ABX method and also by an adaptation of the forced-choice temporal interval method developed by Blackwell (1952) for measuring visual thresholds. In the latter procedure, as in ABX, the stimuli are presented in triads composed of two stimuli which are identical and one which is different, but the "different" stimulus can appear in any of the three positions (first, second, or third) of the triad, and the subject's task is to tell where, i.e., in which position, it is. When the data obtained by the two methods were adjusted to take account of the different levels of chance performance (50% in ABX and 33⅓% in the new method), the levels of discrimination proved to be the same.

We feel reasonably certain that the original noise stimuli are appropriate controls. The two pulses of noise were carefully matched in envelope and duration with the two syllables of the speech stimuli, and the interval of silence, which was in both speech and control the variable part of the pattern, is out in the open, as it were, where it is not likely to be masked or otherwise interfered with. Moreover, as we have seen, the data obtained with these matched controls would appear to have some generality, since very similar results were found with more standard patterns and, indeed, with a different psychophysical method. We will assume, then, that the discrimination functions obtained with the noise controls approximate the discrimination of the speech patterns as they would have been without linguistic experience. The fact that the peaks of the speech discrimination functions rose above the control may then be taken to indicate a learned increase in discrimination across phoneme boundaries. The speech functions nowhere fell below the control, from which we conclude that there was, for whatever reason, no loss in discrimination within phoneme categories.

References

Abramson, A. S. (1961). Identification and discrimination of phonemic tones. *J. acoust. Soc. Amer.*, 33, 842 (Abstract).

Abramson, A. S. and Bastian, J. (in preparation). Identification and discrimination of phonemic vowel duration.

Bastian, J., Delattre, P. and Liberman, A. M. (1959). Silent interval as a cue for the distinction between stops and semivowels in medial position. *J. acoust. Soc. Amer.*, 31, 1568 (Abstract).

Bastian, J., Eimas, P. D. and Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *J. acoust. Soc. Amer.*, 33, 842 (Abstract).

Blackwell, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *J. opt. Soc. Amer.*, 42, 606.

Chapanis, A. S. and Halsey, R. M. (1956). Absolute judgments of spectrum colours. *J. Psychol.*, 42, 99.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. and Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, 24, 597.

Cooper, F. S., Liberman, A. M., Harris, K. S. and Grubb, P. M. (1961). Some input-output relations observed in experiments on the perception of speech. *Proceedings of the Second International Congress on Cybernetics.* Namur, Belgium.

Delattre, P. C., Liberman, A. M. and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, 27, 769.

Fry, D. B., Abramson, A. S., Eimas, P. D. and Liberman, A. M. (in preparation). The identification and discrimination of synthetic vowels.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 46, 373.

Griffith, B. C. (1958). A study of the relation between phoneme labelling and discriminability in the perception of synthetic stop consonants. *Unpublished Ph.D. dissertation, University of Connecticut.*

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1958). Effect of third-formant transitions on the perception of the voiced stop consonants. *J. acoust. Soc. Amer.*, 30, 122.

Harris, K. S., Bastian, J. and Liberman, A. M. (1961). Mimicry and the perception of a phonemic contrast induced by silent interval: electromyographic and acoustic measures. *J. acoust. Soc. Amer.*, 33, 842 (Abstract).

Ladefoged, P. (1959). The perception of speech. In *Mechanisation of Thought Processes*, Vol. I (London).

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, 65, 497.

Liberman, A. M., Delattre, P. C., Cooper, F. S. and Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of stop and nasal consonants. *Psychol. Monogr.*, 68, No. 8 (Whole No. 379).

Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 54, 358.

Liberman, A. M. (1957). Some results of research on speech perception. *J. acoust. Soc. Amer.*, 29, 117.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153.

LIBERMAN, A. M., HARRIS, K. S., KINNEY, J. A. and LANE, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *J. exp. Psychol.*, 61, 379.

LISKER, L. (1957). Closure duration and the inter-vocalic voiced-voiceless distinction in English. *Language*, 33, 42.

LISKER, L. (in preparation). On separating some acoustic cues to the voicing of intervocalic stops in English.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81.

POLLACK, I. (1952). The information of elementary auditory displays. *J. acoust. Soc. Amer.*, 24, 745.

POLLACK, I. (1953). The information of elementary auditory displays. II. *J. acoust. Soc. Amer.*, 25, 765.