Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 24, No. 6, 597–606, November, 1952

Some Experiments on the Perception of Synthetic Speech Sounds*

FRANKLIN S. COOPER, PIERRE C. DELATTRE,† ALVIN M. LIBERMAN,‡ JOHN M. BORST, AND LOUIS J. GERSTMAN
Haskins Laboratories, New York, New York
(Received August 4, 1952)

Synthetic methods applied to isolated syllables have permitted a systematic exploration of the acoustic cues to the perception of some of the consonant sounds. Methods, results, and working hypotheses are discussed.

THE program of research on which we are engaged was described in general terms at the preceding Speech Communication Conference.[1] As we pointed out there, and in more detail in another paper,[2] our work on the perception of speech was based on the assumption that we would have a flexible and convenient experimental method if we could use a spectrographic display to control or manipulate speech sounds. Workers at the Bell Telephone Laboratories had developed the sound spectrograph, which made it instrumentally feasible to obtain spectrograms of relatively long samples of connected speech, and it had become evident that the spectrographic transform has important advantages over the oscillogram as a way of displaying speech sounds to the eye. We were interested in using the spectrogram, not merely as a representation of speech sounds, but also as a basis for modifying and, in the extreme case, creating them. For that purpose we built a machine called a pattern playback, which converts spectrographic pictures into sound, using either photographic copies of actual spectrograms or, alternatively, "synthetic" patterns which are painted by hand on a cellulose acetate base. Having determined first that the playback would speak quite intelligibly from photographic copies of actual spectrograms, we proceeded to prepare hand-painted patterns of test sentences[3] which were, by comparison with the original spectrograms, very highly simplified (see Fig. 1). In drawing the hand-painted spectrograms we tried, as the first step, to reproduce as well as we could those aspects of the original pattern which were most apparent to the eye, and then, by working back and forth between hand-painted spectrogram and sound, we modified the patterns, usually by trial and error, until the simplified spectrograms were rather highly intelligible.

The work with simplified spectrograms did not provide unequivocal answers to questions about the minimal and invariant patterns for the various sounds of speech, but it did enable us to develop our techniques, and, further, it suggested certain specific problems which appeared to warrant more systematic investigation. In our research on these problems we have departed from the procedure of progressively simplifying the spectrograms of actual speech and have undertaken instead to study the effects on perception of variations in isolated acoustic elements or patterns. Thus, we can hope to determine the separate contributions to the perception of speech of several acoustic variables and, ultimately, to learn how they can be combined to best effect.

STOP CONSONANTS: BURSTS OF NOISE

A careful inspection of actual spectrograms suggests, and our experience with simplified spectrograms seems to confirm, that one of the variables that may enable a listener to differentiate *p*, *t*, and *k* is the position along the frequency scale of the brief burst of noise which constitutes the acoustic counterpart of the articulatory explosion. In an attempt to isolate this variable and determine its role in perception, we prepared a series of schematized burst-plus-vowel patterns in which bursts at each of twelve frequency positions were paired with each of seven cardinal vowels. As can be seen in Fig. 2, the bursts were constant as to size and shape, and the vowels, which maintained a steady state throughout, were composed of two formants only.[4] All of the combinations of burst and vowel—a total of 84 syllable patterns—were converted into sound and presented in random order to 30 college students with instructions to identify the initial component of the syllable as *p*, *t*, or *k*.

Figure 3 shows, for each of the vowels, how the subjects' identifications varied according to the frequency position of the burst. In general, it appears that this one variable—the frequency position of the burst—provides the listener with a basis for distinguishing among *p*, *t*, and *k*. We see that high frequency bursts were heard as *t* for all vowels. Bursts at lower frequencies were heard as *k* when they were on a level with, or slightly above, the second formant of the vowel; otherwise they were heard as *p*. It is clear that for *p* and *k*

* This research was made possible in part by funds granted by the Carnegie Corporation of New York and in part through the support of the Department of Defense in connection with Contract DA49-170-sc-274.
† Also at the University of Pennsylvania, Philadelphia, Pennsylvania.
‡ Also at the University of Connecticut, Storrs, Connecticut.
[1] F. S. Cooper, J. Acoust. Soc. Am. 22, 761–762 (1950).
[2] Cooper, Liberman, and Borst, Proc. Natl. Acad. Sci. 37, 318–325 (1951).
[3] We employed sentence lists prepared by Egan and co-workers. See J. Egan, O.S.R.D. Report No. 3802, Psycho-Acoustic Laboratory, Harvard University, November 1, 1944.
[4] For a complete account of the experimental work leading to the choice of the formant frequencies of these vowels, see Delattre, Liberman, and Cooper, Le Maître Phonétique No. 96, 30–36 (July–December, 1951).

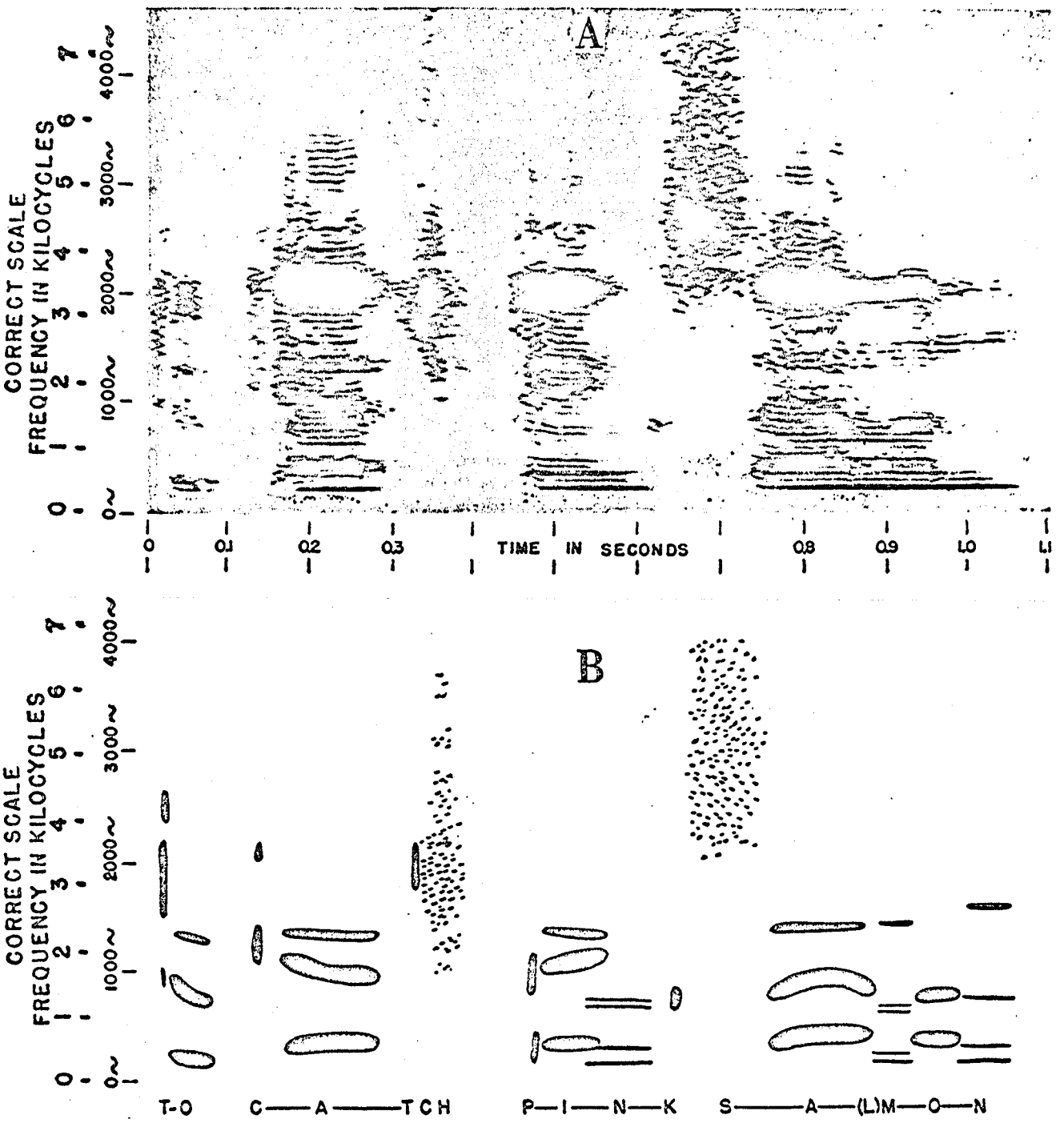

FIG. 1. (A) Sound spectrogram of human speech. (B) Simplified version of the same phrase, painted by hand. Both of these spectrographic patterns are intelligible after conversion into sound by means of the pattern playback. (Reproduced by courtesy of the American Journal of Psychology.)

the identification of the consonant depended, not solely on the frequency position of the burst of noise, but rather on this position in relation to the vowel. In other words, the perception of these stimuli, and also, perhaps, the perception of their spoken counterparts, requires the consonant-vowel combination as a minimal acoustic unit.[5]

[5] For a detailed account of the experiment and a further discussion of the results, see Liberman, Delattre, and Cooper, Am. J. Psychol. (to be published).

STOP CONSONANTS: TRANSITIONS

We turned next in our study of the stop consonants to another aspect of the acoustic pattern which is often evident in spectrograms, namely, the consonant-vowel transitions. These transitions are seen as rapid shifts in the frequency positions of the vowel formants where vowel and consonant join and are typically most marked for the second formant, although they are usually present in some degree for the other formants as well.

The interpretation of these transitions is a major problem. In articulatory terms it is clear that the positions of the speech organs for consonant and vowel are, in general, different and that the rapid movement from one position to the other will usually produce an equally rapid shift in the acoustic output. The parallel interpretation in perceptual terms is that these rapid changes in the sound stream are no more than the necessary transitions (hence, the name) between the sounds that serve to identify successive phonemes; by implication, the transitions are merely nulls which dilute, or even confuse, the acoustic message.

An alternative interpretation is that these rapid changes are heard as important distinguishing characteristics of the sound stream and may indeed serve as a principal acoustic cue for the perception of the consonant-vowel combination—the syllable or "half-syllable," as the case may be.[6] Since a vowel is usually loud and long (hence, identifiable by itself) whereas a consonant is often weak or of very short duration, the practical effect is that the transitional portion of the vowel is transferred to the acoustic counterpart of the consonant. But whether one considers the syllable as separable in this restricted sense or as an indissoluble unit, the second interpretation of transitions gives far more weight to their role in speech perception than the term "transition" would imply. The first step in exploring this question experimentally was to select one vowel and to draw synthetic spectrograms in which a

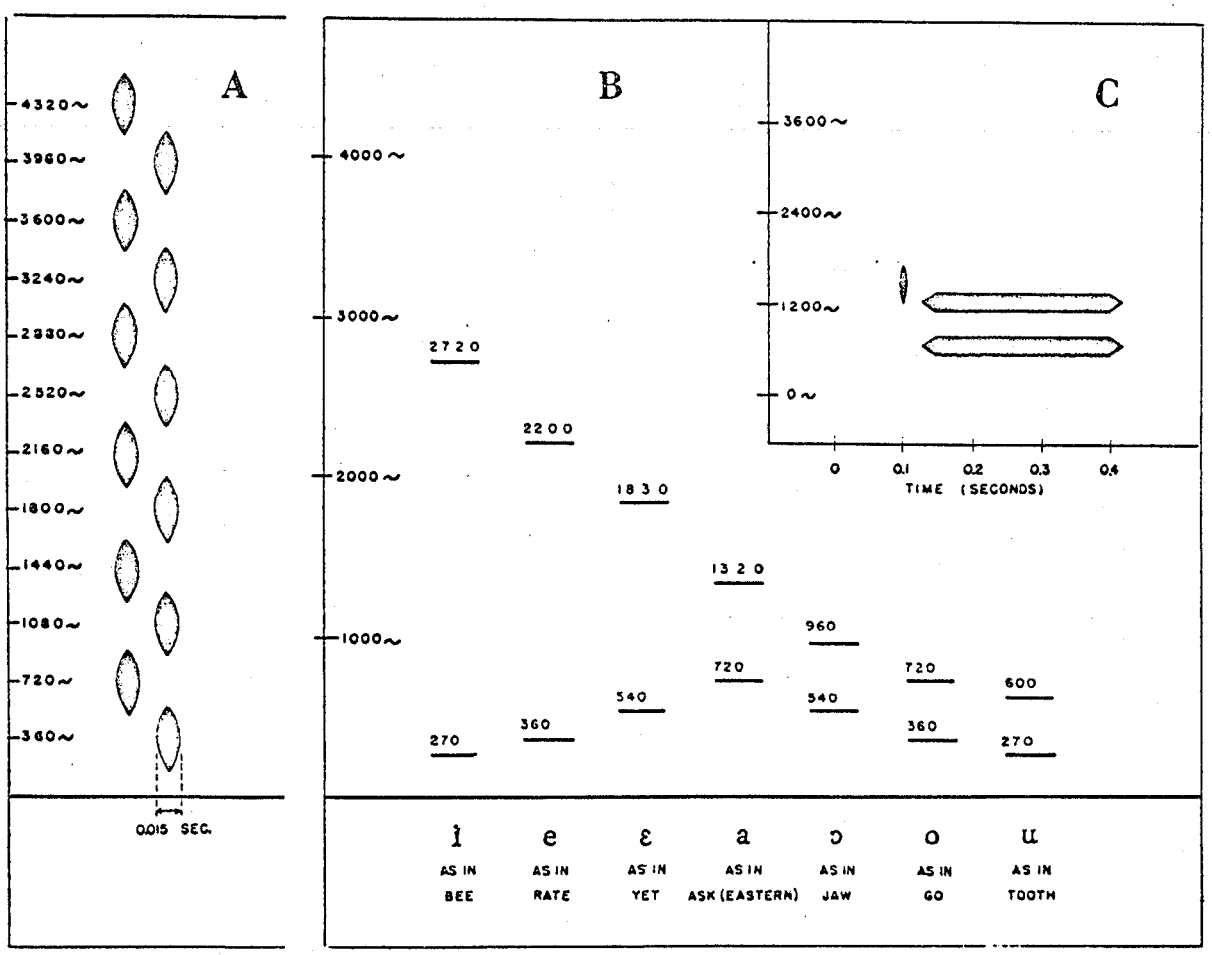

FIG. 2. Stimulus patterns used in determining the effect of burst position on the perception of the unvoiced stop consonants. (A) Frequency positions of the twelve bursts of noise. (B) Frequency positions of the formants of the two-formant vowels with which the bursts were paired. (C) One of the 84 "syllables" formed by pairing a burst of noise and a two-formant vowel. (Reproduced by courtesy of the American Journal of Psychology.)

[6] M. Joos, Language, Suppl. 24, 122, 1948.

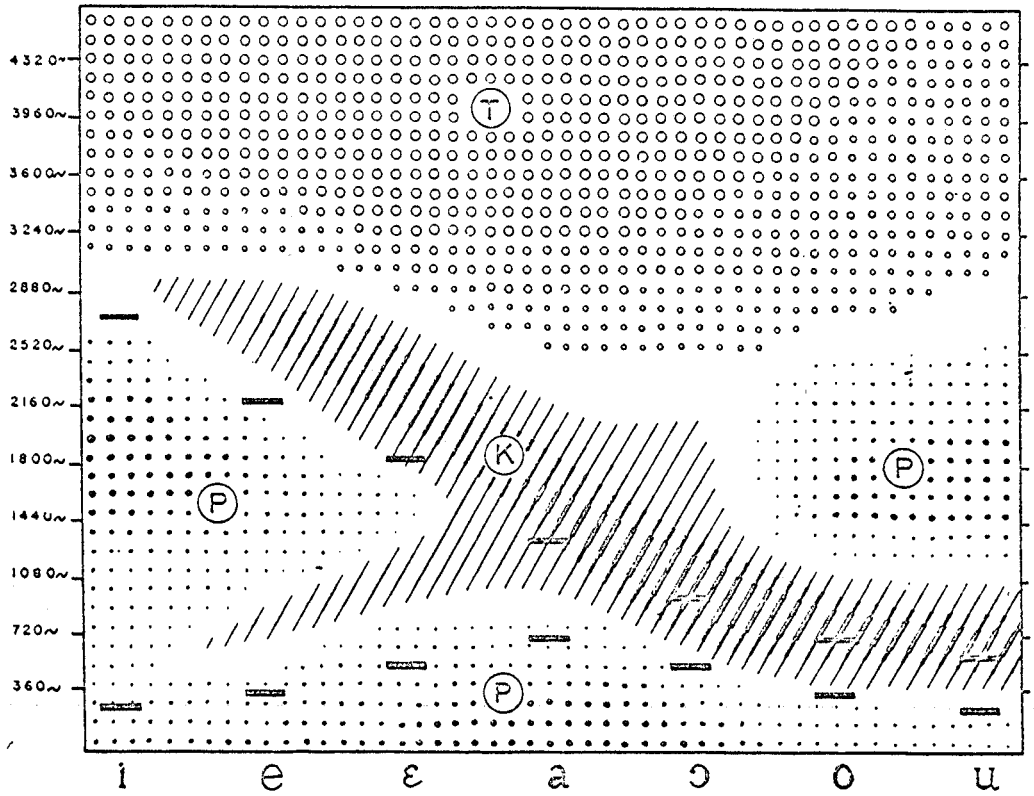

FIG. 3. Preferred identifications by 30 listeners of the stimuli of Fig. 2. The twelve center frequencies of the bursts of noise are shown along the *y* axis; the seven vowels are arranged in the order front-to-back along the *x* axis, with formant positions given. The zones show the burst-vowel combinations for which one of the three responses was dominant and indicate roughly the extent of dominance. (Reproduced by courtesy of the American Journal of Psychology.)

variety of "transitions" were added to the two-formant version of the vowel. Such a series is shown in Fig. 4. In the upper line, the first formant has always the same rising transition, but there is a systematic variation in the transitions of the second formant: rising sharply at the left of the figure, straight in the center, and falling steeply at the right. In the lower line, the same sequence of second-formant transitions is repeated, but the first formant has a very small rising transition. One observation that came from this sort of exploratory work was that the transitions of the first formant appear to contribute to voicing of the stop consonants, while transitions of the second formant provide a basis for distinguishing among *b*, *d*, and *g*, or their cognates *p*, *t*, and *k*. [The sounds corresponding to these painted spectrograms were presented by means of magnetic tape recordings.[7] These sounds were generated by passing the patterns of Fig. 4 through the pattern playback.]

Our first attempts to generalize from the second-formant transitions for *ba*, *da*, and *ga* to the corresponding transitions for a different vowel showed quite clearly that matters would be more complicated—that we were again dealing with interactions or interrelations between the acoustic counterparts of consonant and vowel when they occur together as a syllable.

This exploratory work was followed by systematic tests of a range of second-formant transitions applied to each of the seven vowels that had previously been used in the *PTK*-burst experiment. The resulting test syllables are very much like those shown in Fig. 4 except that the extent of second-formant transitions was increased by one degree at the left and two at the right, giving a total of eleven different degrees of transition. Thus, with seven vowels, there were seventy-seven consonant-vowel stimuli to be judged. The results are

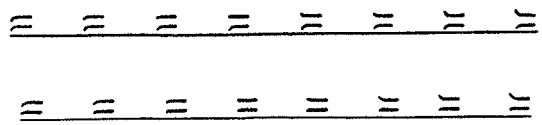

FIG. 4. Variations in the onset of vowel formants used in exploring the role of transitions. When the patterns shown in the upper line are converted into sound by the playback, the syllables *ba*, *da*, and *ga* are heard in succession as the second-formant transitions vary from rising to falling. The upper and lower lines differ only in the extent of the first-formant transitions; this seems to contribute to the voiced (upper) or unvoiced (lower) characteristics of the consonants.

[7] The authors will supply, at cost, copies of the sound demonstration on magnetic tape or disk.

shown in the upper left-hand corner of Fig. 5. There are, for each vowel, three bars showing the distribution of judgments among *b*, *d*, and *g* as a function of the direction and extent of the second-formant transition. All first formants had rising transitions similar to those in the upper half of Fig. 4. The length of the bar gives a rough indication of the range of different transitions included within the group judgment for each sound and, hence, some indication of the degree of over-lap or confusion among the sounds. Specifically, the connecting lines pass through the median judgments, and the bars end at the quartile points. Thus, the array shows that most of the subjects heard a rising second-formant transition as *b* and that falling transitions might be heard either as *g* or *d*, depending on the vowel.

In the lower left-hand quadrant of Fig. 5 are the results of a comparable test in which all of the first formants were straight, or "unvoiced." Also, the two right-hand arrays of Fig. 5 give comparable data for the two sets of test stimuli mentioned above when, however, the subjects were instructed to choose among *p*, *t*, and *k*. In a general way, the four arrays are similar. The results agree in the predominance of *b* (or *p*) judgments for rising second-formant transitions and in the existence of a crossover between *d* (or *t*) and *g* (or *k*) judgments for falling transitions. It does appear that the cognate relationships between *ptk* and *bdg* are effectively cued by the second-formant transitions. A problem remains, however, of finding adequate cues for the distinction between voiced and unvoiced stops. Transitions of the first formant make some difference, and, of course, the presence or absence of a "voice bar" at the fundamental frequency plays a role.

The same data are presented somewhat more directly in Fig. 6. Vowel color is now displayed along the *y* axis, and the extent of second-formant transition is the *x* dimension, as shown pictorially at the lower left; the heights of the "mountains" show the percentage distributions of the responses indicated by the column headings (*b*, *d*, *g*, or *p*, *t*, *k*) for the various stimulus

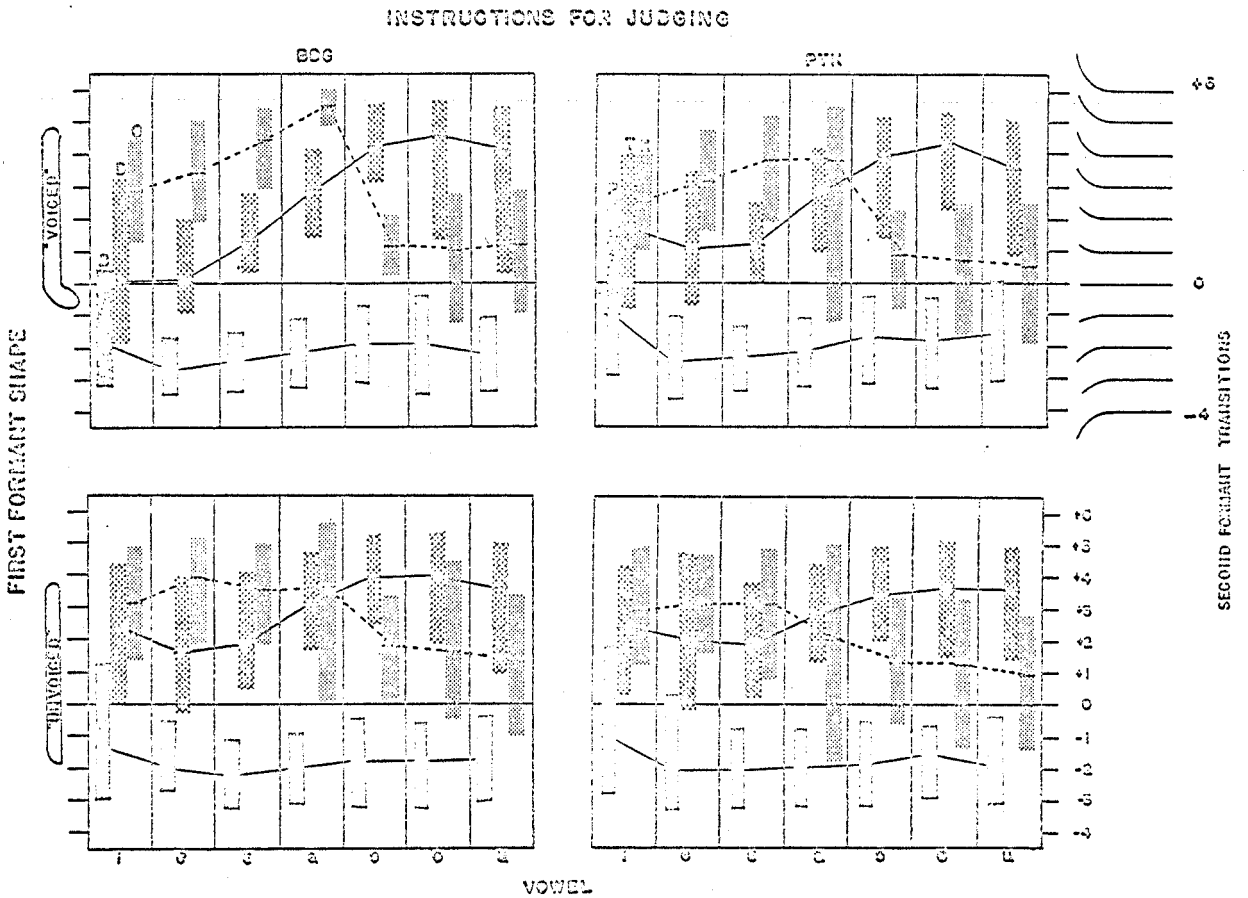

FIG. 5. Results of four experiments on the contribution of second-formant transitions to the identification of the stop consonants. The upper left quadrant is derived from the responses *b*, *d*, or *g* given by 33 subjects to a set of syllables similar to those in the upper row of Fig. 4, that is, syllables characterized by a constant rising transition of the first formant, eleven degrees of transition of the second-formant (*y* axis), and seven vowels (*x* axis). The connecting lines go through the transitions eliciting median responses for each vowel; the lengths of the bars show the quartile ranges. The upper right quadrant indicates the responses of a numerically equal, but different group of subjects to the same stimuli, but with instructions to limit their responses to *p*, *t*, or *k*. The lower two quadrants are based on experiments which differ from the above only in the employment of stimulus patterns similar to those of the lower row of Fig. 4, that is, of patterns with no transitions in the first formants.

combinations of transition plus vowel. One is somewhat reassured, in dealing with experiments of this kind, to find that some of the stimuli do yield unanimous agreement and that the variations, both with degree of transition and with vowel color, seem to be smooth and continuous. At the same time, it is evident that these transitions do not suffice in all cases; there are a few consonant-vowel combinations for which none of the transitions gives an unambiguous cue. However, a comparison of the data on transitions with the previous results for bursts shows that most of the ambiguous cases would possibly be resolved if *both* of these cues were being used.

NASAL RESONANTS: *M* AND *L*

A class of sounds which are cognate to *b*, *d*, *g*, and *p*, *t*, *k* consists of the nasal resonants *m*, *n*, and ŋ. Another series of exploratory experiments indicated that each of these consonants involves a vowel transition and also a steady-state resonant sound whose intensity and frequency characteristics are different from those of the vowel. We were interested, in the first instance, in segregating the effects of the transitions, and this seemed to require that we find a neutral position for the resonant portion which would convey the impression of resonant nasal consonants as a class without providing important cues to the identity of the particular consonant. This is probably an oversimplification, but it does permit us to collect data for a comparison of the resonants with the voiced and voiceless stops. We have run a first set of tests in which the previous seven vowels and eleven degrees of transition were paired in all possible combinations in syllables which also contained a neutral nasal resonance portion. The consonant was placed in terminal position since initial ŋ does not occur in English. This work is still in process; hence, no figure will be presented. In a general way, we find about the same distributions that appeared in the *BDG*-transition test. Thus, the second-formant transitions which were regularly heard as *b* (or *p*) in the preceding test now give the cognate *m*, and there is a comparable crossover in which *n* parallels *d* (or *t*) and ŋ parallels *g* (or *k*). There are some indications in the data that we are not dealing with a monovalent stimulus in this case; probably we shall have to explore variations in the supposedly neutral resonance.

The exploratory work that precedes systematic tests of the kind that we have been discussing tends to become divergent almost without limit, but also it turns up interesting leads, such as the example shown in Fig. 7.

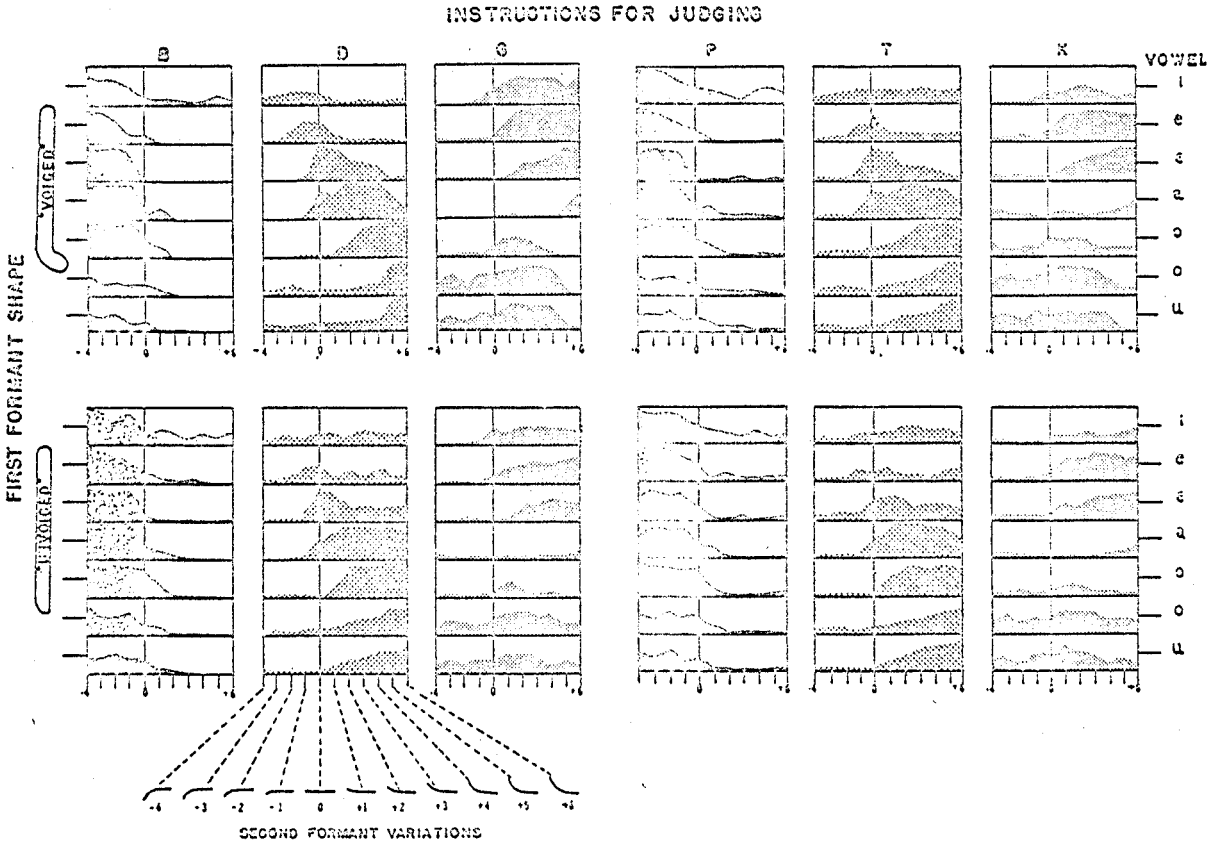

FIG. 6. Distributions of judgments of the stop consonants obtained with variations in second-formant transitions. The percentage of *b*, *d*, and *g* (or *p*, *t*, and *k*) responses is plotted for each vowel as a function of the direction and degree of second-formant transition (*x* axis). The medians and quartiles of these distributions were shown in the corresponding quadrants of Fig. 5.

We find that a transition from higher to lower frequency which is followed by a steady-state resonant sound is often heard as *m* but may at times sound like *l* instead. Our best guess at the moment and on the basis of cut-and-try experiments with only two vowels is that the distinctions between *l* and *m* are multiple, involving (a) the rate of transition of the second formant—a gradual transition favors *l*, a rapid transition favors *m*; (b) the frequency position of the low formant of the resonant portion—*l* is favored by a higher frequency; and (c) the behavior of the second formant in passing from vowel to resonance—if the second formant of the resonant portion forms a plausible continuation of the second formant of the vowel, one tends to hear *l*, whereas a sudden discontinuity contributes to an *m* impression. The first three lines of Fig. 7 illustrate these three pattern differences; the fourth line shows a composite pattern which incorporates all three differences. These are tentative results, but they indicate the kind of thing that one finds in the exploratory phase. [The sounds which correspond to these syllables were demonstrated, line by line, and in both forward and reverse directions.]

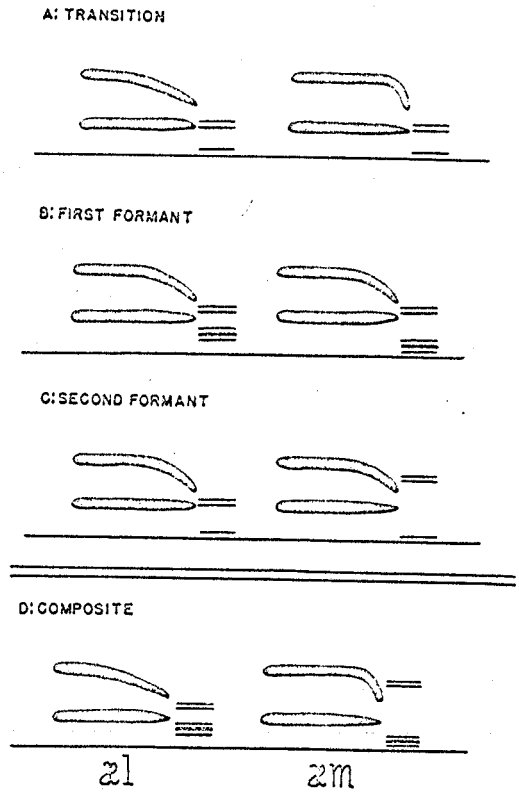

FIG. 7. Spectrographic patterns showing three acoustic cues which contribute to the perceptual differentiation of *l* and *m* following the vowel *æ*.

VOWELS

In this review we shall pass over a sizable block of work on two-formant and one-formant synthetic vowels, except to say that some of the results are most readily explained on the basis that the ear can, and sometimes does, perform an averaging operation on two formants which lie close together; thus, the first and second formants of the back vowels may at times be replaceable by a single formant, or the second and third formants of *i* by a single high formant. We have not so far found it necessary to use three formants to obtain reasonably good vowel color for the cardinal vowels, but an exploratory investigation has indicated that *transitions* of the third formant may contribute to consonant identification. Of course, the behavior of the third formant in spectrograms of the Midwestern *r* and of nasal vowels is well known.[8]

SOME FUTURE DIRECTIONS

The general directions in which the present work should be extended are fairly obvious. We have studied various acoustic cues in isolation. We can reasonably expect that the synthetic sounds will be identified with greater accuracy if two or three cues are provided *simultaneously*. We can even hope that *not more* than two or three acoustic cues will be required to give high intelligibility, even though the resulting sounds may still not be entirely lifelike. In addition, it is quite possible that such speech will be more resistant to noise than normal speech. As to the effectiveness of multiple cues, we know already that a transition added to a burst of noise improves the stop consonants, but we have yet to investigate what adjustments in burst position and in extent of transition may be required for the best combination of these two cues and just how much improvement will result. Also, while it is clear that bursts and transitions complement each other in the sense that when one cue is weak, the other is usually strong, nevertheless, there may remain some syllables for which both cues together may not suffice, and one must then search for other cues. One such possibility is a transition in the third formant of the vowel, and we do have some exploratory evidence of contributions from this quarter. However, the problem is not merely to find additional acoustic cues which make a contribution, but rather to sift out the two or three most efficient cues; that is, we should like eventually to rank-order the cues in terms of their *relative* contributions to intelligibility. Also, we need to run tests in which a greater variety of stimuli are presented and wider ranges of judgments are allowed, until finally, all of the phonemes of American English have been studied in their usual combinations.

The step from phoneme combinations to connected speech will involve a variety of additional problems, but we ought, eventually, to be able to synthesize connected speech on the sole basis of rules, or principles, of the same general kind that we are beginning to derive for the stops and the resonants. This is not our primary

[8] M. Joos, Language, Suppl. 24, 93 (1948); also, Potter, Kopp, and Green, *Visible Speech* (D. Van Nostrand Company, Inc., New York, 1947), p. 220 ff.

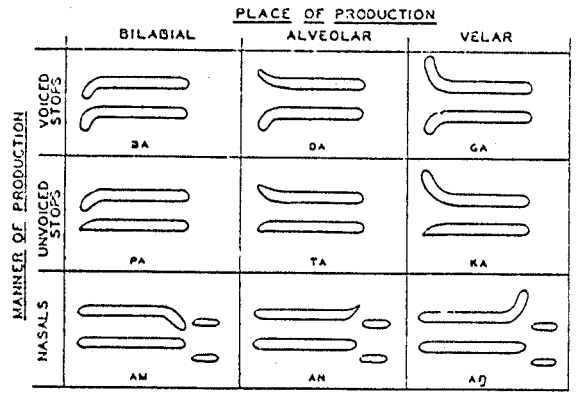

FIG. 8. An arrangement according to articulatory categories of spectrographic patterns which are heard as syllables consisting of the voiced stops, unvoiced stops, and nasal resonants paired in each case with the vowel *a*.

objective, but it does provide an over-all check on the validity of the acoustic descriptions.

SOME SPECULATIONS

In addition to the pragmatic objective of synthesizing speech and giving simplified acoustic descriptions of the speech sounds, we may hope that eventually an acoustic counterpart of the linguistic structure of the language might emerge—and indeed, that the regularities of the one structure might complement those of the other. Figure 8, for example, shows one attempt to correlate some of our acoustic data with the articulatory and linguistic patterns of English. There the schematic patterns for the voiced and voiceless stops and the nasal resonants (paired with the vowel *a*) are arranged in a 3-by-3 array based on articulatory features. It does seem that the acoustic data fit naturally into the array, with the distinctions among columns being given by the transitions of the second formant, and, among rows, by three "markers," namely, rising transitions of the first formant for the voiced stops, no transitions of the first formant (also bursts of noise not shown in the figure) for the unvoiced stops, and a steady resonant portion for the nasal resonants. [The playback sounds corresponding to the patterns of Fig. 8 were played, row by row and column by column.] We should probably not try to generalize from these limited data; we have not yet made the corresponding comparisons for a range of vowels, and some changes in interpretation may be necessary when we do.

As a second point, it may be of interest to examine the data from the point of view that perception involves a set of binary choices. You will recall that bursts of noise preceding a vowel were always heard as *t* when the center frequency of the burst was high, but that low bursts were heard as *p* or *k*, depending on the vowel that followed:

$$\text{Bursts:}\quad \begin{matrix}\text{High}\ (+)= & t\\ \text{Low}\ (-)= & p \text{ or } k\end{matrix}.$$

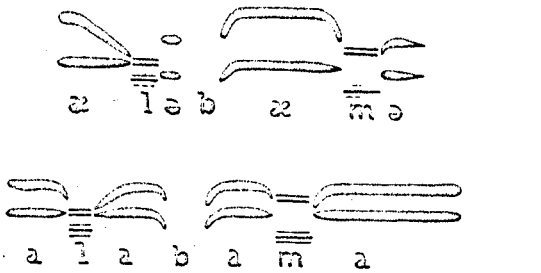

FIG. 9. Simplified spectrograms of the word "Alabama," drawn, in so far as this was possible, according to the rules derived from our research on the component phones and then modified somewhat to mimic a Southern U. S. Pronunciation (upper figure) and a French pronunciation (lower figure).

You will also recall that transitions of the second formant, if rising, were always heard as *p*, and if falling as *t* or *k*, depending on the vowel that followed:

$$\text{Transitions:}\quad \begin{matrix}\text{Falling}\ (+)= & t \text{ or } k\\ \text{Rising}\ (-)= & p\end{matrix}.$$

We have then a basis for deciding among *p*, *t*, and *k*, when both cues are given: *p*= − − (low burst, rising transition), *t*= + + (high burst, falling transition), *k*= − + (low burst, falling transition). It would appear, then, that perceptual distinctions among *p*, *t*, and *k* might conceivably be made on the basis of only two separate binary decisions.[9] If this is correct, we should be able to synthesize satisfactory stop consonants without regard to the exact placement of bursts or to the precise degree of transition, but merely on the basis of "high" or "low" bursts and transitions. We are by no means confident that this can be done.

For a third point, let us return to the general subject of transitions. It seems fairly clear that transitions are important in speech perception, and one could wish for a name that would carry *this* implication rather than its opposite. You have seen how the identification of a particular transition (or burst) seems to depend also on the vowel, so that, apparently, one is perceiving an acoustic unit having the approximate dimensions of a syllable or half-syllable. Now this is not really very surprising if spectrograms are taken at face-value, but we—and perhaps some other workers as well—had undertaken to find the "invariants" of speech, a term which implies, at least in its simplest interpretation, a one-to-one correspondence between something half-hidden in the spectrogram and the successive phonemes of the message. It is precisely this kind of relationship that we do *not* find, at least for these stripped-down stops and nasal resonants. It may be useful to phrase this departure from a one-to-one correspondence between phoneme and sound in the technical jargon of cryptography, thereby borrowing a well-established

[9] We should, perhaps, point out that the kind of binary scheme being considered here differs in several respects from the system put forward by Jakobson, Fant, and Halle, Technical Report No. 13, Acoustics Laboratory, M.I.T. May, 1952.

distinction, and say that we seem to be dealing, at the acoustic level, with an *encoded* message rather than an *enciphered* one—or, more probably, with a mixture of code and cipher. But the important point, however phrased, is a caution that one may not always be able to find the phoneme in the speech wave, because it may not exist there in free form; in other words, one should not expect always to be able to find acoustic invariants for the *individual* phonemes.

The problem of speech perception is then to describe the decoding process either in terms of the decoding mechanism or—as we are trying to do—by compiling the code book, one in which there is one column for acoustic entries and another column for message units, whether these be phonemes, syllables, words, or whatever.

One more bit of speculation, if we may. The results of the *PTK*-burst experiment—and also the results with transitions—provide some extreme cases which suggest that the perceived similarities and differences between speech sounds may correspond more closely to the similarities and differences in the *articulatory* domain than to those in the *acoustic* domain; that is to say, the relation between perception and articulation may be simpler than the relation between perception and the acoustic stimulus. In Fig. 3, the set of bursts which were called *k* differ markedly in acoustic terms, despite the fact that they are heard as the same speech sound and are spoken in about—although not quite—the same way. On the other hand, the bursts at 1440 cps are identical sounds in acoustic terms, but they are heard as different speech sounds when paired with different vowels, e.g., *pi*, *ka*, and *pu*. Here, the perceived *differences* in the consonant are in contrast to the acoustic "*similarities*," but they might very well parallel articulatory differences if it is reasonable to assume that a person, in attempting to duplicate the sound of these bursts, would find it easiest to use his lips when his mouth is set to say *i* or *u* (close vowels) but would find it easiest to use the arch of the tongue with his mouth in position to say *a* (open vowel).

These are examples of what we mean in saying that perception may at times be more closely and simply related to the articulatory movements than to the acoustic stimuli. This is not a new concept—the central idea has been stated in various ways by various workers[10] —but we do believe that these considerations must be taken into account in any theory of speech perception; obviously, they are most directly related to the functioning of the decoding mechanism.

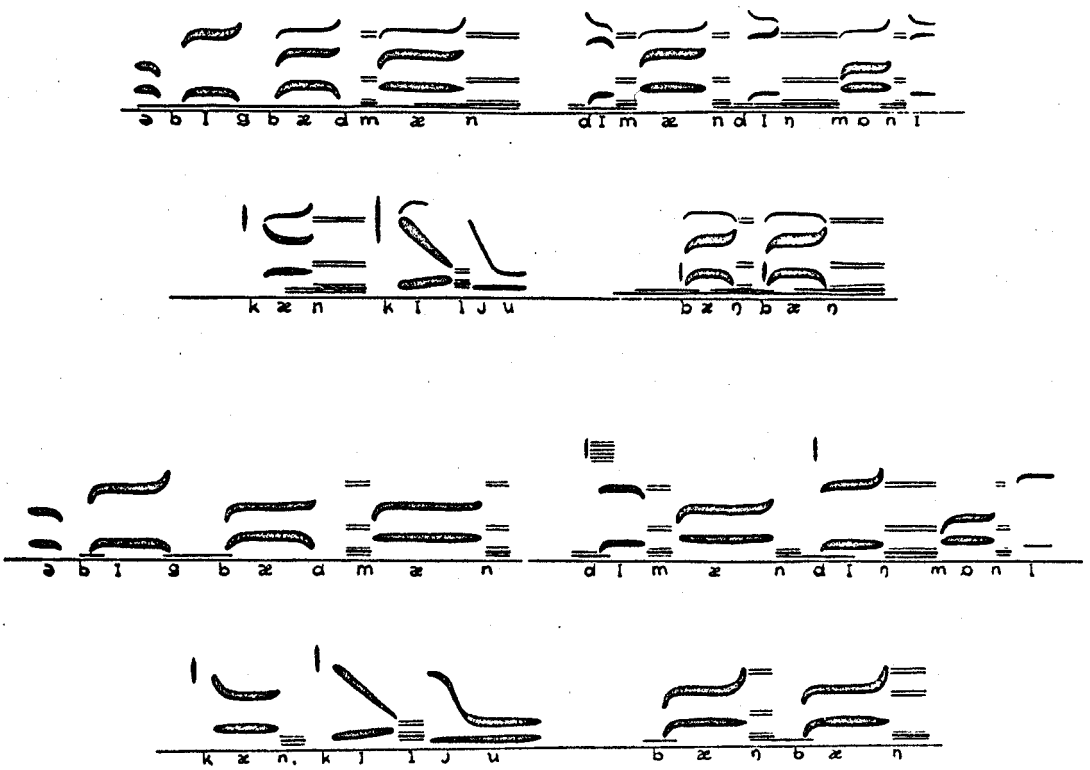

FIG. 10. Two versions of a sentence employing principally stop and resonant consonants. The lower version is a first draft which was painted directly from the typewritten text in accordance with the rules derived from our experiments. Revisions by ear (including the use of some third-formant transitions) resulted in the upper version. Both were highly intelligible when converted into sound by the playback.

[10] Notably, R. H. Stetson, *Motor Phonetics* (Oberlin College, 1951); also, M. Joos, Language, Suppl. 24, 98 (1948).

SYNTHESIS OF CONNECTED SPEECH

In discussing future directions for the general program of work that has been described here, we mentioned the synthesis of connected speech as a long-range objective. It is possible, of course, to attempt synthesis using only the limited information we now have about only a few of the sounds of speech. We shall play for you some examples of words and sentences which were synthesized on the basis of rules derived from our experiments. It is fairly evident that the rules alone are inadequate at this stage and that these examples do benefit from extrapolations of the cardinal vowels to the vowels of American English and from some hunches about diphthongization, syllable length, and stress. However, in all cases, the words were created *de novo* —without reference to actual spectrograms—and employed bursts and transitions for the production of the stop and resonant consonants. [This portion of the demonstration consisted of the following recordings from the playback: (1) "Alabama," from the patterns of Fig. 9. The upper version yielded a Southern dialect; the lower version gave the word as it might have been pronounced by a Frenchman. (2) Spondees taken from the lists prepared at the Psycho Acoustic Laboratory: "backbone, bonbon, outlaw, pancake, cookbook, cupcake, nutmeg." (3) Sentences: "Oh my aching back." "At M.I.T. meet Lick and Locke." "A playback can talk back." (4) The sentence of Fig. 10, in two versions; the lower is the first draft, painted directly from the typewritten page; the upper version benefited from revisions by ear. Some transitions of the third formant were introduced, in addition to the use "by rule" of second-formant transitions and bursts.]

Apparently most of you understood some, if not all, of the examples; even so, it is clear that much remains to be done to achieve a working mastery of the rules governing the acoustic stimuli by which we perceive speech.