Reprinted from
THE AMERICAN JOURNAL OF PSYCHOLOGY
Vol. LXV, pp. 497-516
October, 1952

THE ROLE OF SELECTED STIMULUS-VARIABLES IN THE PERCEPTION OF THE UNVOICED STOP CONSONANTS

By ALVIN M. LIBERMAN, University of Connecticut, PIERRE DELATTRE, University of Pennsylvania, and FRANKLIN S. COOPER, Haskins Laboratories, New York

In earlier reports we have described a method for investigating the perception of speech and other complex sounds.[1] This method uses a spectrographic display as a basis for controlling the acoustic stimuli, and depends on an instrument, called a 'pattern playback,' which converts spectrograms into sound and thus makes it possible to evaluate by ear the effects of a wide variety of experimental modifications in the acoustic (spectrographic) pattern. As part of the development of the playback method, and in a preliminary attempt to strip the speech stream down to its phonemic essentials, we undertook to simplify the spectrographic pattern and yet preserve the intelligibility of the message.[2]

The first step in the simplification procedure was to copy from spectrograms of connected speech those patterns which appeared most prominently on visual inspection. These patterns, painted by hand on a clear cellulose acetate base, were converted into sound and then, on a trial-and-error basis, they were altered in various details in order to restore intelligibility and gain still greater simplicity. The result was a set

* Accepted for publication November 5, 1951. This research was made possible by funds granted by the Carnegie Corporation of New York and the University of Pennsylvania Faculty Research Fund.

[1] F. S. Cooper, Reading machines, in *Research on Guidance Devices and Reading Machines for the Blind, Committee on Sensory Devices, Nat. Acad. Sci.*, 1947, 38-41, also Appendix F; Spectrum analysis, *J. Acoust. Soc. Amer.*, 22, 1950, 761-762; F. S. Cooper, A. M. Liberman, and J. M. Borst, The interconversion of audible and visible patterns as a basis for research in the perception of speech, *Proc. Nat. Acad. Sci.*, 37, 1951, 318-325.

[2] We shall deal here only with those acoustic stimuli which determine phoneme, syllable, or word identification, leaving aside the various other kinds of information (*e.g.* the identity and mood of the speaker) which are presumably carried in the speech wave.

of highly simplified and somewhat schematized spectrograms, corresponding to 20 test-sentences, which yielded a median intelligibility of about 85% when converted into sound by the playback and presented to naïve listeners. Fig. 1 shows a sample of the kind and degree of simplification that was achieved, and, for comparison, a spectrogram of the sounds as they appear in their original, spoken form.[3] In prepar-

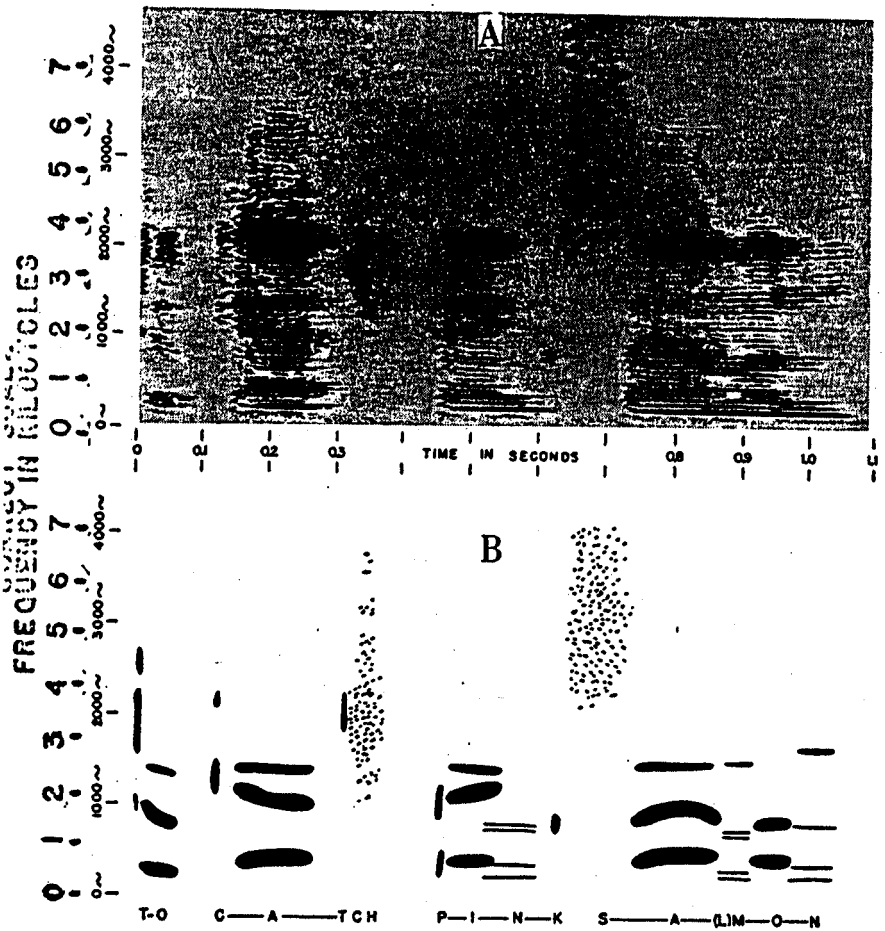

FIG. 1. SHOWING SPECTROGRAM (A) AND SIMPLIFIED VERSION (B) OF A SPOKEN PHRASE

[3] This spectrogram and those of Fig. 2 were produced by a spectrograph similar in principle, but considerably different in detail, from the one developed originally at the Bell Telephone Laboratories. (For a description of the original sound spectrograph see a series of articles by members of the Bell Telephone Laboratories in

ing the schematized spectrograms we attempted merely to find an intelligible version of simplified speech, rather than to explore the effects of systematic variations; hence, the spectrogram of Fig. 1 and the others like it represent only a first approximation to the stimulus-essentials of phoneme communication, and, at best, the results obtained suggest the directions which more intensive and specific investigations might reasonably take. One result which has seemed to us to deserve more careful study relates to the stimuli for the perception of *p*, *t*, and *k*; the experiment reported here is concerned with that problem.

Perceptually, acoustically, and also in terms of the articulatory movements which produce them, the unvoiced stop consonants comprise a distinct class of speech sounds. That they form a class in the perception of speech is indicated by the results of Wiener and Miller, who asked *Ss* to identify nonsense syllables spoken against a background of noise, and found that the unvoiced stops were confused most frequently with other unvoiced stops; in only a small percentage of cases were the unvoiced stops identified as sounds belonging to any other class of phoneme.[4] Acoustically, the stops are similar to each other and different from the other categories of speech sounds in that they are relatively short bursts of comparatively low energy, including most often a rather wide range of frequencies. The unvoiced stops (*p*, *t*, and *k*) are presumably distinguished from their voiced cognates (*b*, *d*, and *g*) on the basis that the voiced stops, like all voiced sounds, have acoustic energy at the fundamental frequency of the voice, whereas the unvoiced stops do not.[5] From the standpoint of articulatory movements, the three unvoiced stops are produced in common by the relatively quick opening of the oral tract with the consequent release of the interrupted breath stream. (For *p* the interruption is produced by joining the upper and lower lips, for *t* by raising the tongue tip against the upper teeth or the alveols, and for *k* by raising the back of the tongue against the velum or soft palate.) The vocal cords do not vibrate during the phonation of *p*, *t*, and *k*, which accounts for the absence of acoustic energy at the fundamental frequency of the voice.

Some of the acoustic features which presumably distinguish one stop from another can be seen in Fig. 2. The energy in any one of the stops is usually spread over a rather wide range of frequencies, but the range is typically less than the total frequency-scale of the spectrogram, and a characteristic difference among the stops appears to lie in the frequency at which the energy centers. Without attempting to specify the actual frequencies, Potter, Kopp, and Green have noted that the energy of *p* is concentrated at a relatively low frequency position, that of *t* centers at a high value, while that of *k* is at a high frequency for the front vowels (*e.g. i* as in *eat*), at a middle frequency for the mid-vowels (*e.g.* ʌ as in *up*), and at a low frequency for the back vowels (*e.g. u* as in *moot*).[6] These investigators have also

---

J. Acoust. Soc. Amer., 18, 1946.) The spectrograph from which our spectrograms were made was designed specifically for use with the pattern playback. It yields film transparencies which record relative sound intensities over a range of approximately 40 db.

[4] F. M. Wiener and G. A. Miller, Some characteristics of human speech, in *Summary Technical Report of Division 17, NDRC, Vol. 3, Transmission and Reception of Sounds under Combat Conditions*, 66-67.

[5] There may be other acoustic differences between voiced and unvoiced stops, but they are less apparent than the presence or absence of voicing.

[6] R. K. Potter, G. A. Kopp, and H. C. Green, *Visible Speech*, 1947, 79-103.

seen that an initial stop normally affects the frequency-positions of the formants[1] at the start of the vowel which follows, the magnitude and direction of the effect being somewhat different for each of the three stop sounds. There may be other characteristic differences among the stop sounds, as, for example, in intensity and in the time-interval between explosion and voicing, but these differences are less obvious from an examination of spectrograms than the differences in frequency-position or the effects on the following vowel. In general, the acoustic features which differentiate *p*, *t*, and *k*, can be defined only in the broadest terms on the basis of the spectrographic evidence now available. That evidence is quite sufficient, however, to enable us to select the acoustic variables which should be manipulated in our experiments, and there is enough evidence to indicate that experimental manipulation will, in fact,

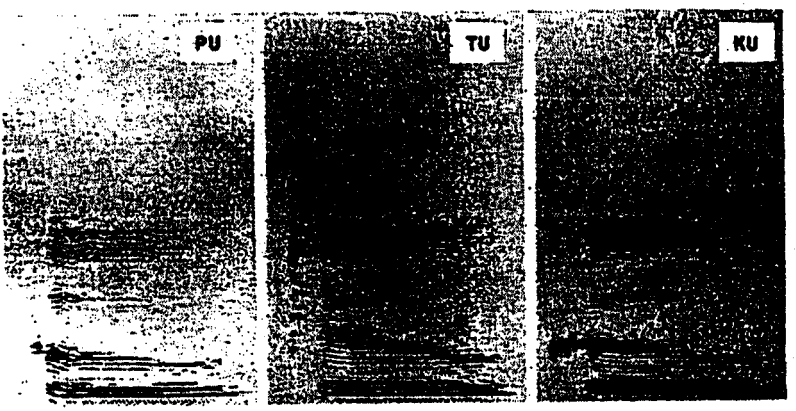

FIG. 2. SPECTROGRAPHIC REPRESENTATIONS OF UNVOICED STOP CONSONANTS BEFORE VOWELS

be necessary if we are to find the essential stimulus-correlates for the perception of the unvoiced stops, inasmuch as it appears that for each of these phonemes there are several distinctive acoustic features which do not vary independently in a collection of speech samples.

The experience gained in producing 'synthetic' speech from simplified spectrograms permits several very tentative conclusions in regard to the acoustic stimuli for *p*, *t*, and *k*, and forms the basis for the present extension of that experimental approach to the problem. It was found in working with the simplified spectrograms that stops as a class could be approximated satisfactorily by representing the sounds as vertical bars (see Fig. 1), and it was reasonably clear that the position of the bar on the frequency scale

[1] The term 'formant,' as applied to vowels and to the so-called resonant consonants (*l*, *m*, *n*, *r*, etc), refers to a frequency region in which there is a relatively high concentration of acoustic energy.

was a most important factor, though not necessarily the only one, in determining whether the stop would be heard as *p*, *t*, or *k*. With the possible exception of *t*, however, there appeared to be no fixed and single frequency-position which characterized the stop. Adjustments in the frequency-position of the bar had to be made according to the nature of the vowel which followed, the need for this adjustment having been shown most clearly by one case in which a particular bar (in one frequency-position) was heard as *k* before one vowel and as *p* before another. If confirmed, this result would mean that the identification of the stop depends, in the case of *p* and *k* at least, not on the acoustic characteristics of the hand-drawn stop (bar) itself, but rather on the stop in relation to the following vowel. In the present experiment we shall attempt to extend and refine these preliminary observations. For that purpose we shall adopt a bar-like figure as a schematized representation of the unvoiced stop consonant, systematically vary its position on the frequency-scale before each of seven cardinal vowels, and determine the effect of these variations on the auditory identification of the stop as *p*, *t*, or *k*.

Apparatus and Procedure

Apparatus. The sounds used in this study were produced by the pattern-playback, an instrument shown in schematic form in Fig. 3. A variable density tone-wheel modulates a thin sheet of light and produces a 120 ~ fundamental together with its

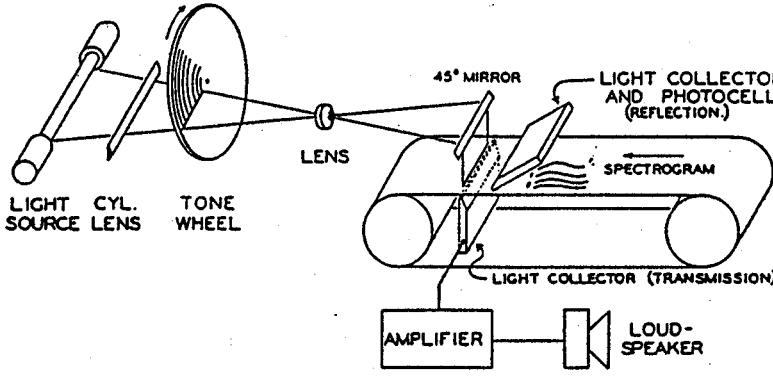

Fig. 3. Operating Principle of the Pattern-Playback

first 50 harmonics, the tones being arranged across the sheet on a scale which matches the frequency scale of the spectrogram. As a painted spectrogram moves through the light, those modulated beams which correspond in position (and hence in frequency) to the painted portions of the spectrogram are reflected into an optical system and

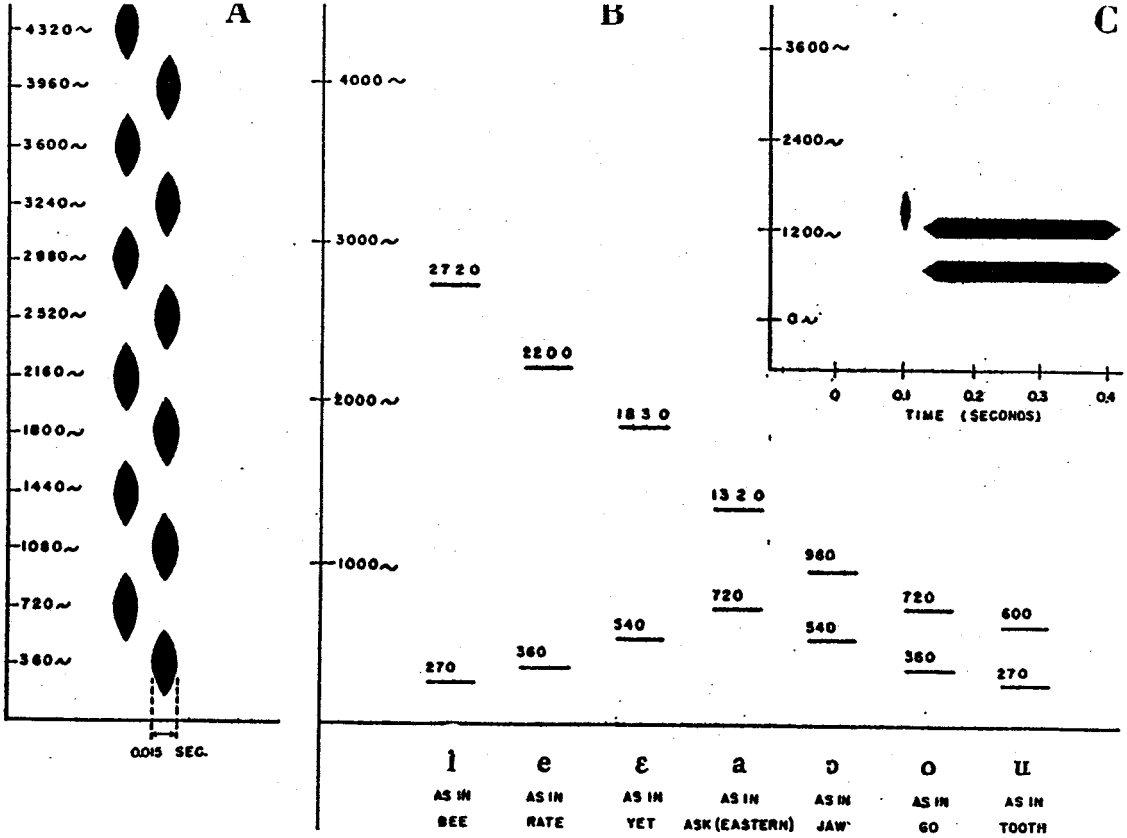

FIG. 4. SHOWING FREQUENCY-POSITIONS (A) OF THE SCHEMATIC STOPS; (B) OF THE VOWEL-FORMANTS; AND (C) ONE OF THE SCHEMATIC SYLLABLES

led to a phototube, whose current is then amplified and supplied to a loud-speaker.[8] In this way the playback produces sounds which have approximately the frequency-, time-, and intensity-characteristics of the spectrographic picture. (Intensity is controlled by varying the reflectance of the paint or the width of the lines.)

The playback provides an attenuation of approximately 6 db. per octave, which corresponds roughly to the distribution of energy in normal speech.

Total harmonic distortion in the individual tones produced by the playback does not exceed 3%; most of this is second harmonic.

Stimuli. The purpose of this experiment required that a schematic stop be presented for auditory identification (as *p*, *t*, or *k*) at various frequency positions before each of several vowels. On the basis of spectrographic evidence and our earlier experience with the simplified spectrograms, we decided that an adequate sample of frequency positions could be had by placing the schematic stop at intervals of 360 ~ within the range 360 to 4320 ~. Seven schematic vowels were used, representing a comprehensive and rather systematic selection from the so-called vowel triangle.

In Fig. 4 A we see the 12 frequency-positions of the schematic stops and in Fig. 4 B the positions of the vowel formants. The 12 frequency-positions of the schematic stop before each of seven vowels made a total of 84 consonant-vowel syllables to be presented to the *Ss*. One of these syllables is shown in Fig. 4 C.

The schematic stop produces a reasonably adequate and indifferent stop-like sound, and was selected after trial-and-error exploratory work. It has a height equal to five contiguous harmonics of the 120~ fundamental, a maximal width of 15 m.sec., and is drawn as an ellipse to reduce transients.

For all the schematic vowels, except *u*, the formants have the shape and dimensions of those shown in Fig. 4 C. To produce the *u* sound it was necessary to reduce the width of the second formant by approximately two-thirds.

The schematic vowels are taken from the results of an earlier study, the purpose of which had been to synthesize the cardinal vowels by the use of two formants only.[9] When converted into sound by the playback, these schematic vowels rather closely approximate normal vowel color, and, with a group of students in phonetics as *Ss*, they had been found to be very highly identifiable.

It should, perhaps, be noted that the sounds produced by the schematic vowels differ slightly from the nearest American English equivalents given at the bottom of Fig. 4. As can be seen in the figure, the schematic vowels maintain a steady state, which is to say they are pure and without diphthongization. In American English speech, on the other hand, at least two of our seven vowels, namely *e* and *o*, are quite noticeably diphthongized. Also, the acoustic differentiations between and among the schematic vowels are probably somewhat greater than those which are produced by an average American speaker.

To increase the intensity of the schematic stop relative to the vowel, we added a small amount of black pigment to the paint used in drawing the vowel formants,

[8] The playback speaks either from photographic copies of actual spectrograms, using the modulated light which is *transmitted* through the relatively transparent portions of the film, or from spectrograms which are painted on a clear base of cellulose acetate, in which case the sound is controlled by the light *reflected* from the painted areas. In this experiment only the painted spectrograms have been used.

[9] Pierre Delattre, Liberman, and Cooper, Voyelles synthétique à deux formantes, et voyelles cardinales, Le Maître Phonétique, December 1951 (in press).

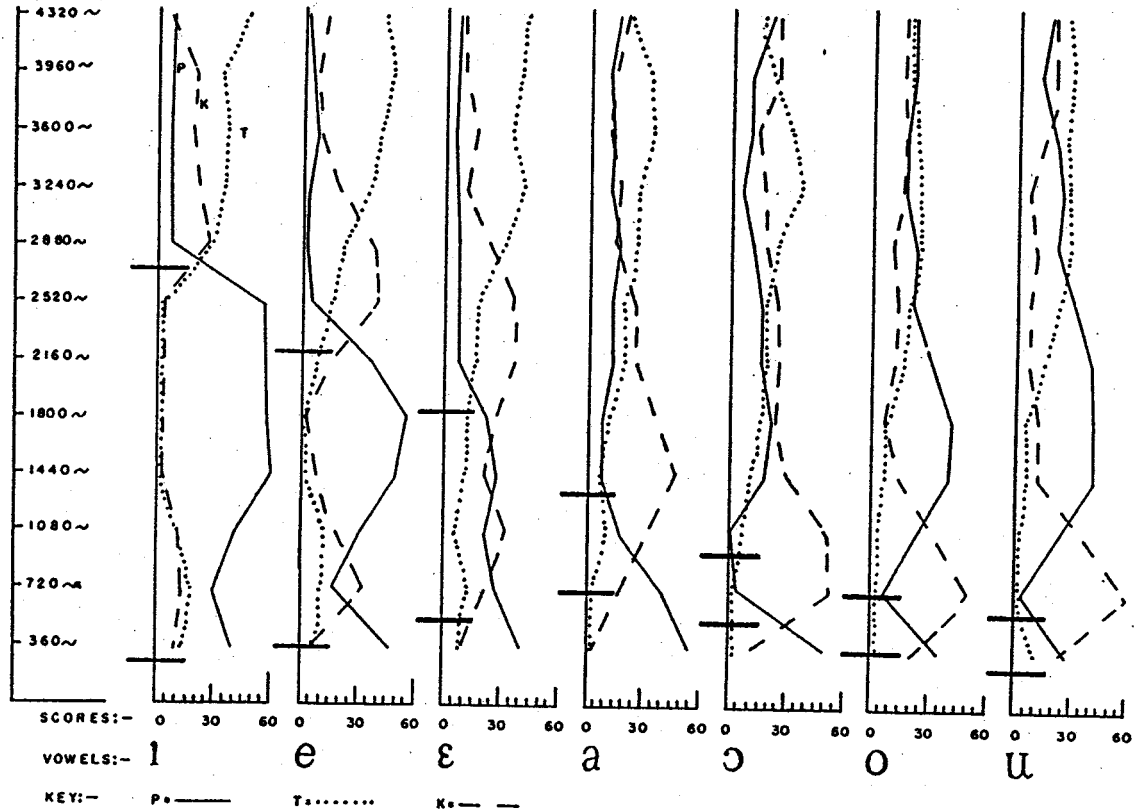

FIG. 5. DISTRIBUTIONS OF *p*, *t*, *k* JUDGMENTS

(The number of times each schematic stop was judged as *p*, *t*, or *k* (labelled 'scores') is plotted for each vowel against the frequency-position of the schematic stop. The formant positions of the schematic vowels are also shown.)

thereby reducing the reflectance of that paint to a level equivalent to 3 db. below the flat white with which the schematic stops were drawn.

Each of the syllables was converted into sound by the playback and was recorded on magnetic tape.

Presentation of stimuli. The syllables were arranged for presentation in a 'random' order, subject to the following restrictions: (1) within each successive block of seven stimuli each vowel appears once, and (2) within each successive block of 12 stimuli each stop position appears once. There was, of course, the general restriction that each combination of stop position and vowel appear only once in the entire series of 84 stimuli.

The recorded syllables were spliced into a master magnetic tape according to the random order described above. To facilitate judgment by *S*, a second recording of each syllable was made and spliced in at such a distance from its twin that each syllable would be presented and then repeated after 0.2 sec. Successive pairs of syllables were separated by 4½ sec. of blank tape; thus *S* had 4½ sec. in which to make his identification of one syllable before being presented with the next.

Disk recordings were made from the magnetic tape master, and all the judgments secured in this experiment were made in regard to sounds reproduced phonographically from the disks.

For every *S* the entire list of 84 syllables was presented twice. The order of presentation of the syllables was reversed for the second series of judgments.

S was instructed to listen to each of the syllables and to identify the initial stop consonant as *p*, *t*, or *k*. He was strongly urged to make such an identification for every syllable heard, even though in some cases the judgment might represent no more than a guess, and he was asked, also, to qualify every judgment by indicating on a four-point scale the degree of confidence he felt in his judgment. An analysis of the results showed a very close relation between the mean of the confidence ratings assigned to any particular identification and the number of *Ss* who agreed in making the identification. That relation was, indeed, so very high that the rating data did not reveal anything which cannot be seen in the frequency distributions of the *p*, *t*, and *k* judgments; therefore, the ratings will not be dealt with further in this report.

To simplify *S*'s task we asked him to identify and record only the initial stop consonant. He was specifically instructed not to try to identify the vowel, or, at least, not to bother to record his identification.

Subjects. Judgments regarding stimuli were secured from 30 *Ss*, who were obtained in two groups: 18 from the University of Connecticut, all of whom were students in an undergraduate course in psychology; and 12 from the University of Pennsylvania, 9 of whom were students in an undergraduate course in phonetics and 3 of whom were advanced students in phonetics.

Results

Fig. 5 shows how the judgments of *p*, *t*, and *k* vary according to the frequency-position of the schematic stop in relation to the schematic vowel with which it was paired. This is essentially a three-dimensional plot, showing the 12 frequency-positions of the schematic stops along the y-axis, the vowels along the x-axis (arranged in order of their front-to-back posi-

tions of articulation), and the number of judgments of *p*, *t*, and *k* for each consonant-vowel combination on auxiliary scales parallel to the x-axis (*i.e.* z-axis scales folded into the x,y-plane). There is a total of 60 judgments for each syllable, each of 30 *Ss* having been given the stimulus-series twice.

Looking first at the distribution of *k* judgments, we find in the case of the vowel *u* that the schematic stop at 720 ~ was called *k* 56 times out of 60, and that the number of *k* judgments falls sharply on either side of that mode, levelling off, on the high frequency-side, at 10 judgments out of 60 when the schematic stop reaches 1440 ~. By comparison with *k* before *u*, the other distributions of *k* judgments are somewhat less sharply peaked and the modes do not represent quite so high a percentage of the total number of judgments. Except in the case of *i*, however, all of the *k* distributions show definite humps or peaks which rise above the levels of the *p* and *t* judgments and reflect considerable agreement among the subjects that schematic stops in certain frequency positions sounded more like *k* than like *p* or *t*.

It is possible that there would have been less agreement in the identification of the schematic stops if the *Ss* had been presented with a wider variety of stimuli and required to select their judgments from among a correspondingly greater number of phonemes. In that connection, however, it is worth noting again the finding of Wiener and Miller that listeners seldom confused an unvoiced stop with a sound belonging to some other class of phonemes;[10] an unvoiced stop, if heard at all, was most often identified as one of the unvoiced stop sounds, even though, within the limits of that category, the specific identification as *p*, *t*, or *k* might have been incorrect. It is also relevant to recall here that schematized stops similar to those of the present experiment were used previously in the simplified spectrograms of sentences (described earlier in this paper), and were found there to be quite highly identifiable. On the basis of these considerations we should suppose that the identifications made in the present experiment will not be radically changed when the schematic stops are presented for judgment among all the other phonemes of American speech, either in the context of meaningful sentences or as parts of nonsense syllables.

Perhaps the most obvious and notable fact about the distributions of *k* judgments is the movement of the mode or peak upward on the frequency scale in the vowel series from *u* to *i*. This movement rather clearly follows the second formant of the vowel as it moves up in frequency, and it would appear that the schematic stop which sounds most like *k* is the one which lies at a frequency slightly above (or, perhaps, in a few cases on a level with) the frequency of the second formant of the vowel which follows.

In the distributions of *k* judgments before *e* and ɛ there is quite clearly

[10] *Op. cit.*, 66-67.

a second mode at a point slightly above the first formant. A suggestion of that mode appears also in *i*, but it does not rise above the *p* or *t* judgments and would not be noted here if it did not fit the pattern set by the vowels which precede it in the series.

The presence of two modes at *e* and ɛ suggests that the *k* impression might be strengthened if we were to place schematic stops at the frequency positions corresponding to each of the two modes. We have tried this, in a preliminary way, and have found no dramatic improvement in the *k*. Whether there is any improvement at all can be determined only by more extensive and systematic observations.

The fact that the distribution of *k* judgments tends to be bimodal for the front vowels (*i*, *e*, and ɛ) and unimodal for the back vowels (*u*, *o*, and ɔ) may be related to some evidence we have regarding the number of formants necessary to produce these two classes of vowel sounds. With the back vowels, in which the first and second formants are normally close together in frequency, a reasonable approximation to the vowel sound can be produced by a single formant placed at a frequency intermediate between the frequency positions of the first and second formants; in the front vowels, on the other hand, where the two formants are widely separated in frequency, a single intermediate formant will not produce the vowel sound. A tentative rule, then, might be that the schematic stop is heard as *k* when the acoustic energy is just above (or at) the frequency positions of those formants (whether one or two) which are essential for the identification of the vowel.

In addition to the shift in frequency position of the *k* mode through the vowel series, there is a tendency for the height of the mode to decrease from *u* to *i*. The amount of decrease is not very great in the series *u* through *a*, nor is it precisely regular; in ɛ and *e* the apparent decrease might be accounted for by the splitting into two modes; but in *i* it is clear that the schematic stops did not sound very much like *k*.

The difference in the number of *k* judgments between *u* and *i* may mean that additional cues are necessary for the identification of *k* when it precedes an extreme front vowel such as *i*. We cannot define these cues with confidence, but, on the basis of what is known about the articulatory movements required to produce *ki* as against *ku*, we can suspect the presence of at least one peculiar feature in the normal utterance of *ki*. To articulate *k*, a speaker places the back of his tongue at a point relatively far back on the roof of the mouth and then quickly lowers the tongue to produce the *k* explosion. A back vowel, such as *u*, is also formed with the back of the tongue arched toward the back of the palate. In contrast to these movements, the articulation of an extreme front vowel, such as *i*, consists normally in raising the front of the tongue toward the front of the mouth. Phoneticians have established that the difference in tongue position between *k* and *i* affects the articulation of *k* before *i*, causing the point of tongue-palate contact for *k* to move forward in the mouth; with *u*, on the other hand, there is no such effect. One might suppose, conversely, that an initial *k* would influence an *i* which follows it, at least to the extent of producing some transition between the *k* explosion and the steady state of the vowel, and that this effect would be considerably less, or even absent, when the vowel is *u*.

In this connection, Potter, Kopp, and Green have pointed out that an initial *k* causes a 'rounding together' of the second and third formants (sometimes the third and fourth formants), and they have suggested that this rounding is more apparent in the front vowels than in the back vowels.[11] If *k* does, in fact, have some effect on the vowel formants of *i*, but not on *u*, this effect might be an important cue to the identity of *k* when it precedes *i*, and the complete absence of this cue in our schematized syllables would then account for the fact that the schematic stop before *i* was not often heard as *k*.

Another possibility, of course, is that the *ki* syllable is simply less identifiable than *ku*, even when all the significant cues are present. Such a difference in identifiability, if it does exist, should not be assumed to result from the subjects' having had less experience with the one combination than with the other, inasmuch as Dewey's phonetic count of English shows *k* before the front vowels *i*, *e*, and ɛ to be at least as frequent in occurrence as *k* before the back vowels *u*, *o*, and ɔ.[12]

It may be relevant to our obtained difference between *ki* and *ku* that in the evolution of several languages, including English in the case of words coming from the Latin by way of Old French, the *k* sound has been unstable before the front vowels—*e.g.* Latin *centum* [*k*ɛntum] becomes English *cent* [*s*ɛnt], while it survives as *k* when it precedes the back vowels—*e.g.* Latin *cortem* [*k*ɔrtɛm] becomes English *court* [*k*ɔrt]. (The fact, pointed out in the preceding paragraph, that the *k* sound seems to occur in English with equal frequency before the front and back vowels is to be attributed to the presence in the language of a relatively large number of borrowed words and words of Germanic origin.) The instability of *k* when it introduces a front vowel can be attributed to purely articulatory, that is motor, factors which cause *k* to shift toward consonants which are somewhat closer to the front vowels, or it can be assumed to be caused by an inherent difficulty in identifying *k* when it is followed by a front vowel.

The distributions of *p* judgments are to a large extent the inverse of the distributions for *k* (in the range 300 to 3000 ~). In *i*, *e*, ɛ, and, perhaps, *a*, the second (higher) formant of the vowel sets an upper limit to the frequency position of the schematic stop which will sound like *p*. Then, beginning with *a* and becoming increasingly apparent in the back vowel series (*a* through *u*), a mode appears in the *p* distribution somewhere above the second formant of the vowel, reaching its highest point at *u*. In this *latter* series the second formant of the vowel becomes a lower limit to the *p* sound.

For all the vowels the judgment *p* predominates as the response to the lowest stop position (360 ~). The number of *p* judgments for this lowest position appears to increase regularly in the *i* through *a* series, with one slight inversion at *e*, and then to decrease in very orderly fashion in the series *a* to *u*. Although the lowest stop position attracts more judgments of

[11] *Op. cit.*, 97-103.
[12] G. Dewey, *Relative Frequency of English Speech Sounds*. 1923, 66.

p than of *k* or *t*, it is the most preferred position for *p* in only three of the seven vowels (ɛ, *a*, and ɔ); therefore the critical aspect of the stimulus for *p* is not simply its low position on the frequency scale.

The distributions of *t* judgments are quite broad and flat, without any obvious peaks or modes, and it is difficult to define a *t* point or region beyond the statement that *t* occupies the frequency range lying above the

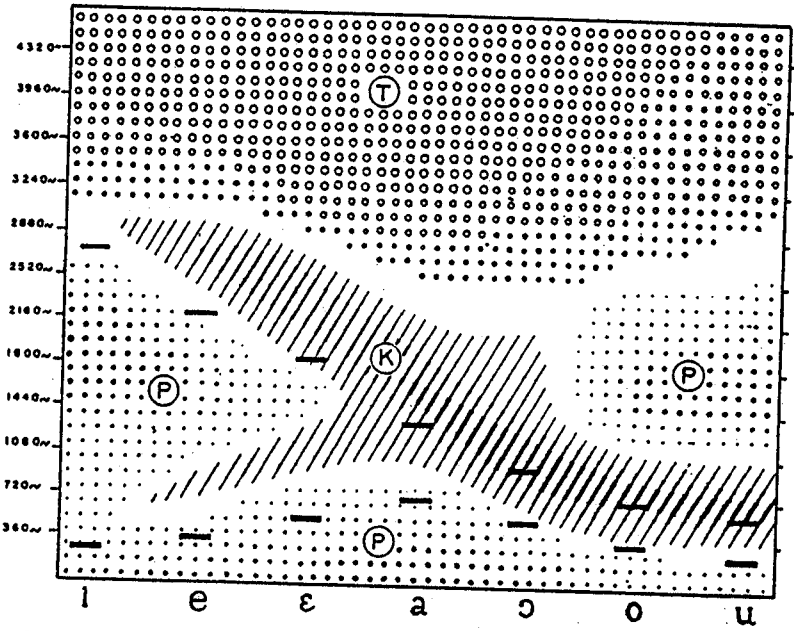

FIG. 6. MAP OF THE AREAS IN WHICH THE JUDGMENT *p*, *t*, OR *k* PREDOMINATES

stop positions which are judged to be *p* or *k*. There is little else to be seen in the *t* distributions except, perhaps, that the *S*s had a greater disposition to judge the stop as *t* when it preceded the front vowels (*i*, *e*, and ɛ) than when it preceded the back vowels (*u*, *o*, and ɔ).

In Fig. 6 we have attempted to represent the results of this experiment in rather general form. The zones of the figure correspond to those areas (*cf.* Fig. 5) in which one of the judgments (*p*, *t*, *k*) occurred more often than the other two; within each zone, the thickness of the line or the size of the dot or circle indicates roughly the extent to which that judgment was preferred.

The reproducibility of the data can be seen by comparing the curves of

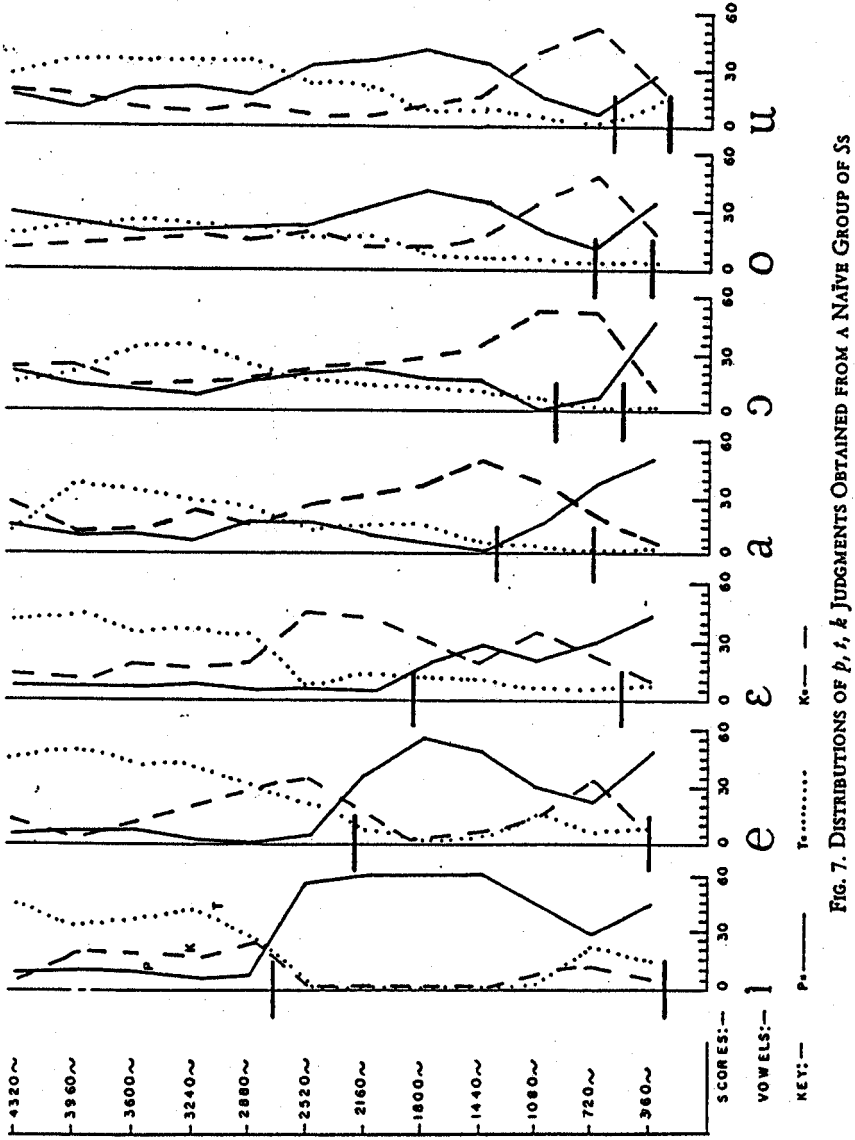

FIG. 7. DISTRIBUTIONS OF *p*, *t*, *k* JUDGMENTS OBTAINED FROM A NAÏVE GROUP OF *Ss*

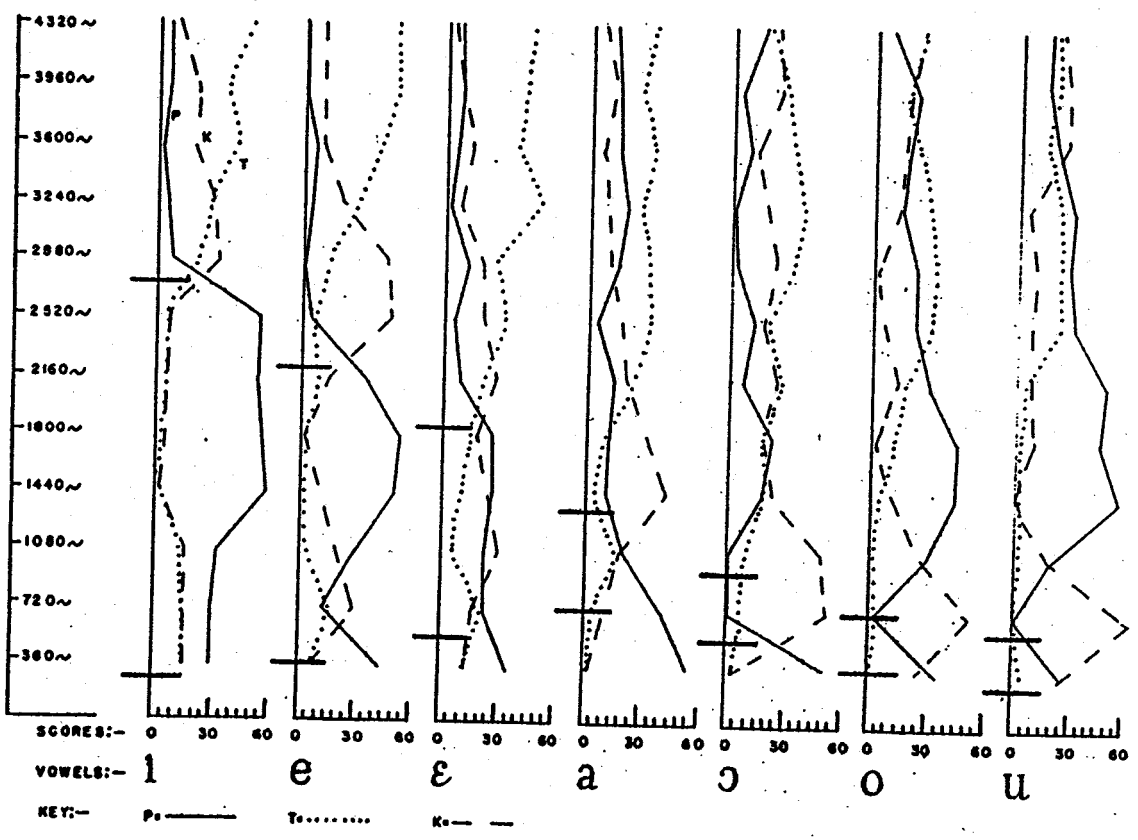

FIG. 8. DISTRIBUTIONS OF *p*, *t*, *k* JUDGMENTS OBTAINED FROM A SOPHISTICATED GROUP OF *Ss*

Figs. 7 and 8. In these figures the data are plotted separately for our two groups of *Ss*, the one being relatively naïve in regard to phonetics and the other relatively sophisticated. The two sets of curves are obviously very similar.

As a further test of reliability, we have compared the curves corresponding to the judgments made on the first and second times through the series of syllables. These curves, which are not shown here, resemble each other quite as much as do the curves for the two groups of subjects in Figs. 7 and 8.

The results of this experiment show clearly the influence of the following vowel on the perception of the schematic stops *p* and *k*.[13] The modes of the *p* and *k* distributions shift in frequency position according to the frequencies of the formants in the vowel which follows, and these shifts are so great, relatively, that the *k* mode crosses over the *p* mode in the vowel series *i* through *u*. Perhaps the clearest single example of this effect is seen (Fig. 5) in the judgments of the schematic stop at 1440 ~. Before *i* this schematic stop was called *p* 59 times out of 60; in the series *i* to *a* there was a progressive decrease in the number of *p* judgments, and a correspondingly progressive increase in *k*, until at *a* this same schematic stop was called *k* 47 times out of 60; then from *a* to *u* the number of *k* judgments decreased, as *p* increased, and at *u* this stop was called *p* 42 times out of 60. In brief, the identity of the *same* schematic stop was judged in a large majority of cases to be *p* when paired with *i* and *u*, but to be *k* when paired with *a*. If these results with schematized syllables truly represent the situation as it is in the perception of actual speech, then it must be concluded that the vowel plays a critical part in the auditory perception of *p* and *k*, and in that event the irreducible acoustic correlate for *p* and *k* is the sound pattern corresponding to the consonant-vowel syllable.

It may be of interest, also, that our results relate to the assumption that the perception of speech depends ultimately on the proprioceptive return

---

[13] We have made some preliminary tests to determine how the 12 schematic stops are judged when they are presented without the schematic vowels. For the three lowest stopped positions, at 360, 720, and 1080 ~, the judgments of *k* predominate, and show a rather high mode at the last-named position. The three next higher positions, at 1440, 1800, and 2160 ~, are called *p*, *t*, and *k* almost equally, with *t* being very slightly favored over *p* and *k*. Above 2160 ~ the frequency of judgments of *t* increases somewhat, largely at the expense of *k*, while *p* remains as it was.

We expect, in a later publication, to present these results in more nearly complete form, and to discuss their relation to the results obtained with the schematic syllables. In the present paper we can only suggest that the *Ss*' identifications of the isolated schematic stops do not appear to be inconsistent with our conclusion that the vowel plays an important part in the perception of a *p* or *k* which precedes it.

from the articulatory movements which are made in speaking. One test of this assumption depends on, and is suggested by, the fact that the relation between the articulatory and acoustic events is a rather complex one, *i.e.* that relatively large changes in articulation can, under certain conditions, produce relatively small acoustic differences, and vice versa. Given this kind of relation between articulation and the acoustic result, and given the assumption that the ultimate cues for the perception of speech arise from the movements of articulation, we should expect that the relation between perception and articulation will be considerably simpler (more nearly direct) than the relation between perception and acoustic stimulus, or, to put it another way, that for any group of speech sounds the perceived similarities (and differences) will correspond more closely to the articulatory than to the acoustic similarities among the sounds. It will, of course, be very difficult to test this expectation with all the sounds of a language, since we cannot readily arrange the sounds along comparable scales of similarity in the acoustic, articulatory, and perceptual domains. We can, however, look for extreme cases in which the question of similarity is reduced to a categorical matter of identity or difference; in the results of our experiment we find several cases which reach, or very closely approximate, those extremes. One such case (the schematic stop at 1440 ~) was described in the preceding paragraph, where it was seen that acoustically identical schematic stops before *i*, *a*, and *u* can add to these vowels initial sounds which, as perceived, are as distinctively different as the *p* of *pi* and *pu* and the *k* of *ka*.[14] On the basis of what is known about the acoustic and articulatory aspects of this particular situation, it is altogether reasonable to assume that these acoustically identical schematic stops could be most closely approximated in the act of speaking only by very different kinds of articulatory movements, and we should suppose that the proprioceptive stimuli which will result from these movements would also be very different. With the mouth set to articulate *i* or *u*, a speaker would presumably use a movement of the lips to produce a sound having roughly the same acoustic characteristics as our schematic stop at 1440 ~; for the vowel *a*, the same acoustic result would be most closely approximated by lowering the back of the tongue from a point of contact with the roof of the mouth. (Our schematic stops are considerably simpler acoustically than the sounds

[14] The quality of *p* or *k* cannot be assumed to be given by the schematic vowel alone, since a listener does not hear an initial *p* or *k* when any one of the schematic vowels is presented without the schematic stop. It should be noted that this is true only of steady-state vowels. We have some very preliminary indications that when the vowel formants are appropriately 'bent' an initial *p* or *k* can, perhaps, be heard even though the schematic stop has been entirely omitted.

produced in actual speech, and it is likely, also, that the acoustic identity which we have been able to impose on the schematic stops is only approximated in speaking.) We assume, then, that the sound patterns corresponding to the syllables *pi* (or *pu*) and *ka* set off the appropriate movements in the listener (that is, the articulatory movements which the listener would make in attempting to reproduce these acoutic patterns) or perhaps only the short-circuited neural equivalents of these movements, with the result that the initial schematic phonemes *p* and *k*, which can be entirely identical as acoustic stimuli, become clearly differentiated in proprioceptive terms and are therefore finally perceived as distinctively different sounds. Thus we have, at the one extreme, a case in which there is complete identity of acoustic stimuli (in the consonant part of the syllable), considerable difference in movement and proprioception, and, in accordance with the proprioceptive rather than the acoustic stimuli, considerable difference in perception. At the other extreme, one would look for a case in which there is (again in the consonant part of the syllable) a large acoustic difference, no articulatory difference, and no perceived difference. In its extreme form that case is not to be found in our data, but we have a reasonable approximation to it in the schematic stops which sounded most like *k*. As can be seen in Fig. 5 or Fig. 6, the frequency position of the 'best' *k* is quite different for *u*, *o*, *ɔ*, *a*, *ɛ*, and *e*, and these differences are large enough to cover a very significant part of our total frequency range. The nearest spoken equivalents to these schematic *k*'s will be produced by articulatory movements which are not precisely identical (the point of tongue contact is toward the back of the mouth for the back vowels and moves forward when the *k* precedes a front vowel), but the differences in articulation are certainly very small by comparison with the articulatory differences between *k* and *p*. We may say, then, in regard to the schematic stops which were heard as *k* that the perception reflects the basic similarity in the articulatory patterns rather than the relatively gross differences among the acoustic stimuli; though this is not the extreme case that we should have liked ideally to find, it does come sufficiently close to provide some additional support for the assumption that the perception of speech depends, in any final analysis, on the proprioceptive stimuli which arise from the movements of articulation.

Summary

Earlier work on the perception of speech yielded a method by which an investigator makes experimental modifications of spectrographic displays and then, with an appropriate playback instrument, converts the spectrograms into sound for aural evaluation. Exploratory research with this

method produced reasonably intelligible sentences from highly simplified spectrograms, and provided a basis for more intensive research into the stimulus-essentials for the perception of individual phoneme, syllable, or word units.

The present study is an intensive investigation of this kind into the effect of one acoustic variable on the perception of unvoiced stop consonants which introduce a consonant-vowel syllable. For that purpose a schematized, spectrographic representation of the unvoiced stop consonants was placed at each of 12 frequency-positions before each of seven schematized vowels. The schematic stop was 600 ~ high and 15 m.sec. wide, as it appeared on the spectrogram, and the 12 positions of the stop sampled the range 360 to 4320 ~ at interval of 360 ~; the schematic vowels were composed of two formants, and were so chosen as to represent a systematic sampling of the vowel triangle. These spectrographic patterns, each one corresponding to a consonant-vowel syllable, were converted into sound by a playback and presented in random order to 30 *Ss* with instructions to identify the initial stop-like sound as *p*, *t*, or *k*.

The *Ss*' responses indicate that the identification of *p* and *k* did not depend solely on the frequency position of the schematic stop, but rather on this position in relation to the schematic vowel which followed. Thus, for each of the vowels, the distribution of *k* judgments showed a pronounced mode (that is, a strong preference for the *k* identification) at the schematic stop which lay slightly above the second formant of the vowel; this mode moved upward along the frequency-scale with the rise in frequency of the second vowel-formant in the back-to-front vowel series, *u*, *o*, *ɔ*, *a*, *ɛ*, *e*, and *i*. For *ɛ*, *e*, and perhaps *i* in which the two formants of the vowel are far apart in frequency, there was a second mode in the *k* distribution at a point just above the first vowel-formant. The height of the *k* mode tended to vary with the vowel, but if we exclude *i*, where none of the schematic stops was identified very often as *k*, the amount of variation was not great. In the frequency range 300 to 3000 ~, within which the *k* and *p* modes lie, the distributions of *p* judgments were in large part the inverse of the *k* distributions.

Above 3000 ~ the schematic stops were judged most often to be *t*. There was very little effect of the schematic vowel on the identification of *t*.

The judgments made in regard to one schematic stop (at 1440 ~) exhibit most clearly the influence of the schematic vowel on the perception of the schematic stop. At the one end of the vowel series, before *u*, this stop was called *p* 70% of the time; from *u* through *o* and *ɔ* the number of *p* judgments steadily decreased as *k* increased, until at *a* 78% of the judg-

ments were to the effect that the stop was *k*; this trend was then reversed from *a* through *ɛ* and *e*, and finally, at *i* the stop was once again identified as *p*, this time in 98% of the responses. This and the other results of the experiment demonstrate that in the perception of the schematic *p* and *k* before schematic vowels, and perhaps for their equivalents in normal speech also, the irreducible acoustic stimulus is the sound pattern corresponding to the consonant-vowel syllable.