

AN EXPERIMENTAL STUDY OF  
THE ACOUSTIC DETERMINANTS OF VOWEL COLOR;  
OBSERVATIONS ON ONE- AND TWO-FORMANT VOWELS  
SYNTHESIZED FROM SPECTROGRAPHIC PATTERNS

PIERRE DELATRE\*, ALVIN M. LIBERMAN,† FRANKLIN S. COOPER AND LOUIS  
J. GERSTMAN

## 1. INTRODUCTION<sup>1</sup>

This paper will report the results of an attempt to synthesize the 16 cardinal vowels of the International Phonetic Association<sup>2</sup> by converting hand-painted spectrographic patterns into sound. The conversion from spectrogram to sound is accomplished with a special-purpose instrument (called a pattern playback) which, by making it possible to listen to spectrograms, provides a basis for evaluating the effects of a wide variety of experimental modifications in the acoustic pattern, and thus affords a convenient method for determining the role of various acoustic features in the perception of speech.<sup>3</sup> This method is most appropriate, perhaps, for investigating the dynamic, or constantly varying, aspects of speech sounds; it is not primarily designed, nor is it necessarily ideally suited, for work with steady-state vowels, though it can be used, as it has been

\* Also University of Pennsylvania.

† Also University of Connecticut.

<sup>1</sup> Some of the data of this study have been published in an earlier paper (P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Voyelles synthétiques à deux formantes et voyelles cardinales," *Le Maître Phonétique*, 96, 30-7 (1951). In that article we were concerned primarily to display the results of our syntheses graphically and to compare that graph with the IPA articulatory charts.

<sup>2</sup> The vowels which served as our models were produced by one of the authors, a phonetician whose native language is French. Naturally his conception of the color of the cardinal vowels might differ slightly from that of Daniel Jones whose pronunciation of these vowels on a well-known recording has long served as the standard. However, we should guess that the differences, if any, are very small, since for 11 of the 16 vowels the IPA offers French vowels as guides to pronunciation. (See *The Principles of the International Phonetic Association*, University College, London, 1949).

<sup>3</sup> Descriptions of the technique and of some of the results obtained with it are to be found in: F. S. Cooper, "Spectrum Analysis," *Journal of the Acoustical Society of America*, 22.761-2 (1950); F. S. Cooper, A. M. Liberman, J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proceedings of the National Academy of Sciences*, 37.318-25 (1951); A. M. Liberman, P. Delattre, F. S. Cooper, "The Role of Selected Stimulus Variables in the Perception of the Unvoiced Stop Consonants," *American Journal of Psychology* (In press); F. S. Cooper, P. Delattre, A. M. Liberman, J. M. Borst, L. J. Gerstman, "Some Experiments on the Perception of Synthetic Speech Sounds," *Journal of the Acoustical Society of America* (In press).

in this study, for that purpose. Our interest in producing these steady-state vowels with the pattern playback stemmed from our need for highly simplified, but still identifiable, synthetic vowels which could be used in combination with some of the consonants we were interested in studying.

Previous investigations,<sup>4</sup> using various techniques of analysis and synthesis, have suggested that three, two, and, in some cases, one formant might be sufficient to produce all the vowel colors, and have defined, at least within broad limits, the frequency positions which those formants should occupy. In a recent monograph, for example, Joos<sup>5</sup> reports that he was able to produce reasonable approximations to normal vowel color with no more than two formants, though he hesitates to draw a final conclusion because his work with these synthetic vowels was essentially exploratory. Our own preliminary investigations convinced us that two formants were, indeed, sufficient to produce identifiable vowels. It was clear, however, that much would be gained by trying to find those formant positions which would give the closest approximations to various vowel colors. The attempt to find those positions becomes, then, the primary purpose of this investigation. Taking the 16 cardinal vowels of the IPA as our standard, we have separately and systematically varied the frequency positions of each formant in synthetic two-formant patterns and then selected from among these many sounds those which formed the best vowels. We have tested the identifiability of the vowels which were finally selected, and we have investigated in a preliminary fashion the effects of variations in the relative intensities of the two formants. On the basis of the results of varying the relative intensities of the two formants, we were led, finally, to experiment with single-formant vowels.

## 2. APPARATUS AND TECHNIQUE FOR PRODUCING THE SYNTHETIC VOWELS

The playback with which the vowels of this study were synthesized produces 50 modulated bands of light corresponding in frequency to a fundamental of 120 cps with its harmonics up to and including the 50th at 6000 cps. These modulated beams of light are made to scan a spectrogram from left to right along the time dimension, the light being spread across the spectrogram to match the frequency scale of the spectrogram. Painted portions of the spectrogram then act to select the light which is modulated at frequencies corresponding to the position of the painted areas and to reflect this light into a phototube whose current is fed, after amplification, into a loudspeaker. The intensity of the sound can be controlled by altering the width of the painted lines, by using paints of different reflectance, or by placing filters in the path of the light beams.

The playback "equalizes" the intensities of the sounds it produces by reducing intensity 9 db per octave in the frequency range above 1500 cps. This rate of

<sup>4</sup> Summaries of and references to many of these studies are to be found in T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan, 1941. More recent studies are described or referred to in: H. K. Dunn, "The Calculations of Vowel Resonances, and an Electrical Vocal Tract," *Journal of the Acoustical Society of America*, 22.740-53 (1950); Gordon E. Peterson and Harold L. Barney, "Control Methods Used in a Study of the Vowels," *Journal of the Acoustical Society of America*, 24.175-84 (1952).

<sup>5</sup> M. Joos, *Acoustic Phonetics*, Language Monographs, No. 23, Baltimore, 1948, 83.

attenuation corresponds roughly to the distribution of energy in normal human speech.

The vowel sounds of this study were synthesized by drawing a spectrographic representation of the desired vowel and then converting the spectrogram into sound. An inspection of spectrograms of spoken vowels reveals, generally, from three to five frequency regions (called *formants*) in which there is a relatively high concentration of acoustic energy. For a male voice pitched at about 120 cycles, the spectrogram typically shows from two to four harmonics within each formant. One also sees that the harmonic nearest the center of the formant is usually more intense than the harmonics on either side. For the synthetic vowels, we have used not more than two formants; in regard to width of formant and to the distribution of energy within the formant, we have tried to make the synthetic vowels correspond rather closely to actual speech. (The width of the formants and the distribution of energy within them, as well as their frequency positions, are shown in Fig. 1 of Section 3.)

Although the playback tones are spaced 120 cycles apart, it is possible to achieve the effect of finer gradations of formant frequency, i.e., 10 to 30 cycles, by altering the relative widths of the three contiguous harmonics which comprise the formant, thereby shifting its center of intensity. Wherever we have resorted to this procedure, we have estimated the equivalent frequency of the "unbalanced" formant on the basis of the relative widths of the lines; otherwise we have given the frequency value corresponding to the middle harmonic.

### 3. PROCEDURE FOR SELECTING THE FORMANT FREQUENCIES AND THE RESULTS OBTAINED

We first did a considerable amount of exploratory research to satisfy ourselves that reasonable approximations to the cardinal vowels could, in fact, be synthesized from two-formant hand-painted spectrograms, and to determine how best to set up a series of vowel-like sounds from which the most nearly adequate vowels might be selected. Using the results of this exploratory work as a basis, we then drew two series of two-formant patterns (a total of 235 vowel patterns) which represented a fairly systematic and comprehensive variation in frequency position for each of the formants. In the one set of vowel patterns the frequency of the *lower formant* was fixed in turn at each of four values, while for each of these settings the higher formant was varied in steps of 120 cycles over a rather considerable range; in the other set, the frequency of the *higher formant* was fixed while the frequency of the lower formant was varied in steps of approximately 30 cycles. (As is well known, relatively smaller steps at the low frequencies produce as much difference for the ear as relatively larger steps at the higher frequencies.)

The first set of vowels can be divided into four groups according to the frequency position of the fixed lower formant (formant 1). For the four positions of this formant, we used frequency values corresponding to the four frequency regions into which formant 1 normally falls in French sustained vowels (1: /i y u/, 2: /e ø o/, 3: /ɛ œ ɔ/, 4: /a α/), choosing French vowels for this pur-

pose because of their known similarity to cardinal vowels. Thus, the first part of the first set combined a fixed formant frequency of 250 cycles (which is the position of formant 1 in a French pronunciation of /i/) with a formant 2 which varied from 280 cycles to 3000 cycles in 20 steps of 120 cycles each. We listened to these vowels many times and finally selected those sounds which, in our judgment, most closely approximated the four cardinal vowels /i y u u/. Because it appeared that the steps of 120 cycles did not permit us to get the very best approximation to the desired vowel, we proceeded, by the method described in the preceding section, to try to improve each vowel by making finer adjustments in the frequency positions of formant 2. Table I summarizes the procedures and results obtained with the first set of vowels.

TABLE I

Results of the attempt to find the closest two-formant (synthetic) approximations to the 16 cardinal vowels of IPA obtained with the first formant fixed at each of 4 positions (corresponding to /i e ε a/ of French) while the second formant varied over the range indicated below.\*

Frequency at which 1st formant was fixed	Frequency range within which 2nd formant was varied	Vowel approximated	Frequency of second formant which gave best approximation
250	480-3000	i	2900
		y	1900
		u	1050
		u	700
360	500-2520	e	2400
		ø	1650
		ɣ	1100
		o	800
510	840-2040	ε	2000
		œ	1450
		Λ	1150
		ɔ	950
720	960-1800	æ	1650
		a	1300
		ɒ	1200
		ɑ	1100

\* It should be remembered that in order to obtain any frequency which was not an integral multiple of 120 cycles we had to resort to the unbalancing procedure described in Section 2. The frequency values given here are estimates based on the relative intensities of the harmonics composing the formant.

In the second set of experimental vowels we fixed formant 2 at the values given in Table I and went on then to vary the frequency of formant 1 (in steps of 30 cycles) over frequency ranges which extended a considerable distance on either side of the fixed frequency we had used in the first set. After much listening to this second set of vowels we concluded that the formant 1 frequencies which gave the best vowels were identical with those we had used in the first set of experimental vowels, except for the formant 1 of /æ a ɒ ɑ/. In the first set we had fixed the first formant of these vowels at 720 cps; after listening to the second set, we decided that better approximations to the vowels /æ a ɒ ɑ/ were obtained when formant 1 was moved up to 750 cps.

In Fig. 1 are shown the synthetic spectrograms which, when converted into sound, most closely match the 16 cardinal vowels. Figure 2 shows the position of each of these synthetic vowels in a two-dimensional space in which the Y-axis

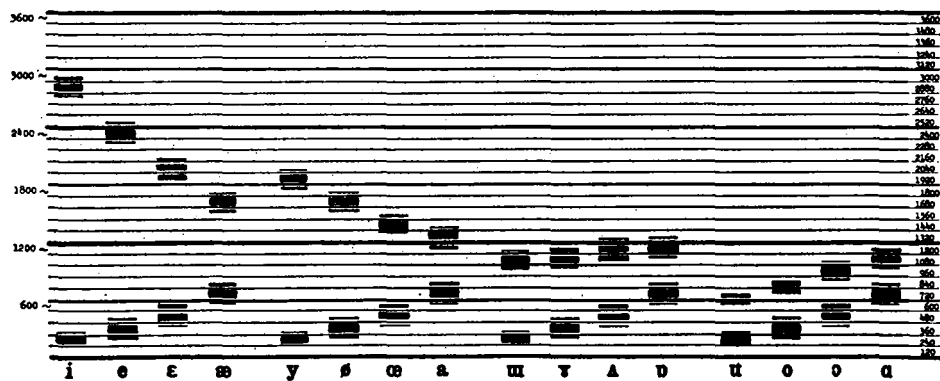


FIG. 1. The synthetic approximations to the cardinal vowels of the IPA, showing the precise composition and frequency position of each formant.

gives the frequency of the first formant and the X-axis the frequency of the second formant.<sup>6</sup> It should be noted that these results were obtained with sounds having a fundamental frequency of 120 cps; no attempt has been made in this study to determine what adjustments in formant-frequency positions might be needed for other values of the fundamental.

To test the identifiability of our synthetic vowels we arranged the final group of 16 sounds in three test formats and presented them to students in phonetics for judgment. Test A included the entire group of 16 cardinal vowels /i e ε æ y ø œ a u ɤ ʌ ɒ u o ɔ ɑ/; Test B consisted of the 10 vowels (out of the group of 16) which are to be found in French /i e ε y ø œ u o ɔ a/; and Test C contained the seven "outside" vowels /i e ε a ɔ u/, a group of sounds which should be reason-

<sup>6</sup> A disk recording of the synthetic vowels described here is available. This recording includes the synthetic vowels that were selected finally as the closest approximations to the 16 cardinal vowels and separately the series of 235 experimental vowels from which the selection was made. The disk can be obtained, at cost, by writing to The Haskins Laboratories, 305 East 43rd Street, New York, N. Y.

ably familiar to speakers of English. For each test the vowels were first presented in quick succession, and identified, in order to familiarize the subject with the total range of vowel-like sounds which were to be included in the test. The synthetic patterns were then presented in random order (at the rate of one every four seconds) for identification by the subjects. The tests were presented to eleven subjects in the order A, B, C, and then, a second time, in reverse order. Five additional subjects took Tests B and C. Most of the subjects had English as a native language, but were studying phonetics and French.

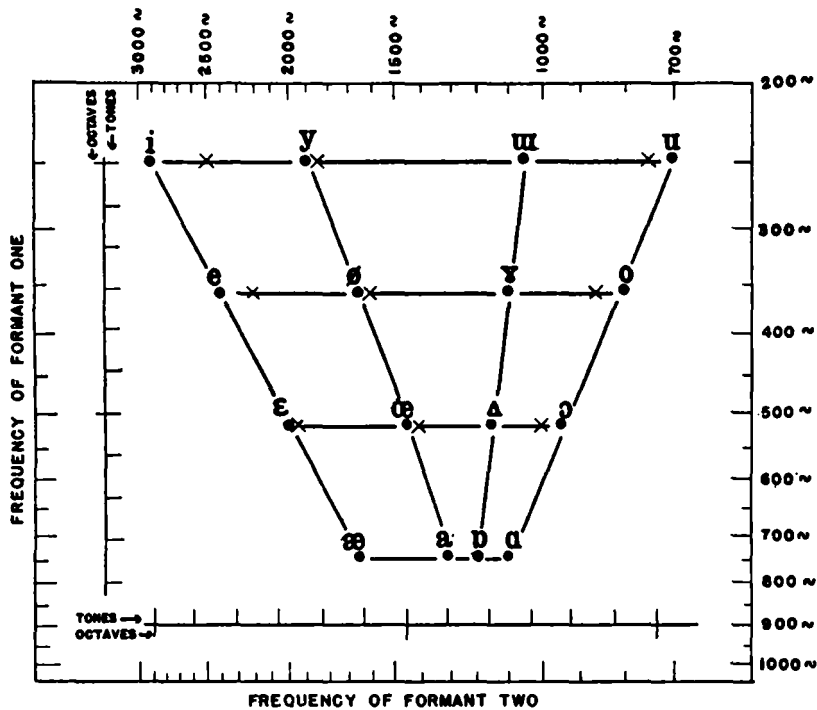


FIG. 2. The acoustic locations of the synthetic vowels, plotted according to the frequency positions of the first and second formants. To make this acoustic plot resemble more closely the familiar vowel quadrilateral (whose coordinates have an articulatory reference), we have reversed the usual order of the frequency scales on both the vertical and horizontal axes and have made the frequency scales logarithmic. (For a detailed comparison of articulatory and acoustic vowel charts see: Martin Joos, *op. cit.*, pp. 49-59.)

In Figs. 3, 4, 5, we see the formant positions of the synthetic vowels, plotted as in Fig. 1, and in the circles corresponding to the location of the vowels are values which summarize the results of the tests. The upper number in each circle is simply the percentage of times the synthetic vowel was judged correctly. The lower number is a composite score, made up of the percentage of correct identifications plus the percentages of identifications which were incorrect by not more than one "step" on the vowel chart (e.g., an /æ/ identified as /ε œ a/ is counted a one-step error).

It is apparent that in Test C the accuracy of judgment was quite high, and,

except for six percent of the judgments of /u/, such errors as did occur were never more than confusions with the nearest neighboring vowel. The percentage of correct identifications is somewhat lower in Test B and lower still in Test A, a result which may quite safely be attributed to the differences among the tests in the number of stimuli to be identified, and to the fact that Tests A and B—particularly Test A—contain vowel sounds /ø œ u ɣ/ with which our subjects are not very familiar.<sup>7</sup>

It is also to be seen in Figs. 3, 4, 5, that the vowels differed greatly among themselves in identifiability. In Test A, for example, /i/ was judged correctly 100% of the time while /ɣ/ and /ʌ/ were called correctly in only 18% of the

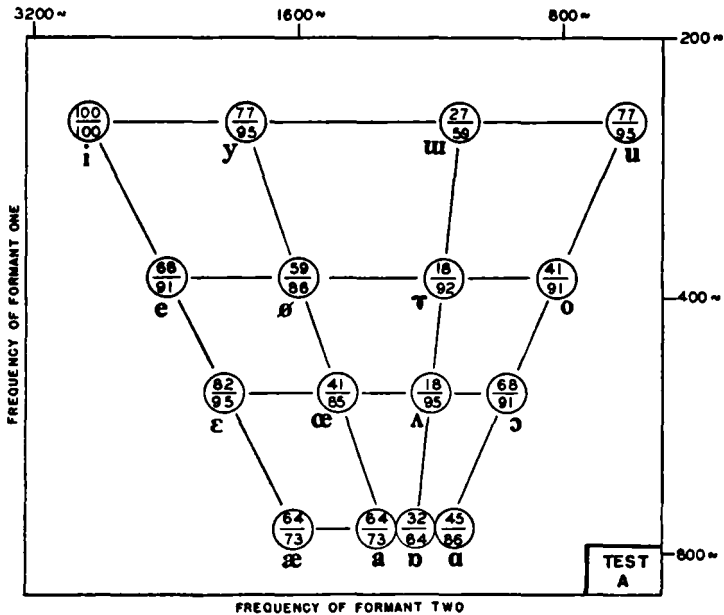


FIG. 3. The accuracy of listeners' identifications of the 16 synthetic vowels. Within each circle, the value given above the line is the percentage of correct identifications and the value below the line is the percentage of correct identifications plus the percentage of identifications which were incorrect to the extent of only one step in the vowel chart.

judgments. These differences are explicable, perhaps, in terms of the differential familiarity of the vowels and, also, by the positions of the vowels in the vowel chart. In general, those vowels which occur in English were identified with fewer errors than those which do not (e.g., /u/ and /ɣ/, which occur in none of the Latin or Germanic languages and which were unknown to our subjects until the time of the test, were identified most poorly of all). We note, also, some

<sup>7</sup> For comparison with the identifications of our synthetic sounds, we presented to the same subjects, in such a way as to duplicate the test conditions of Tests B and C, a group of vowels spoken by one of the authors. (We did not include the vowels of Test A because we doubted our ability to speak several of the vowels correctly.) In all of the identifications of these spoken vowels there was only one error.

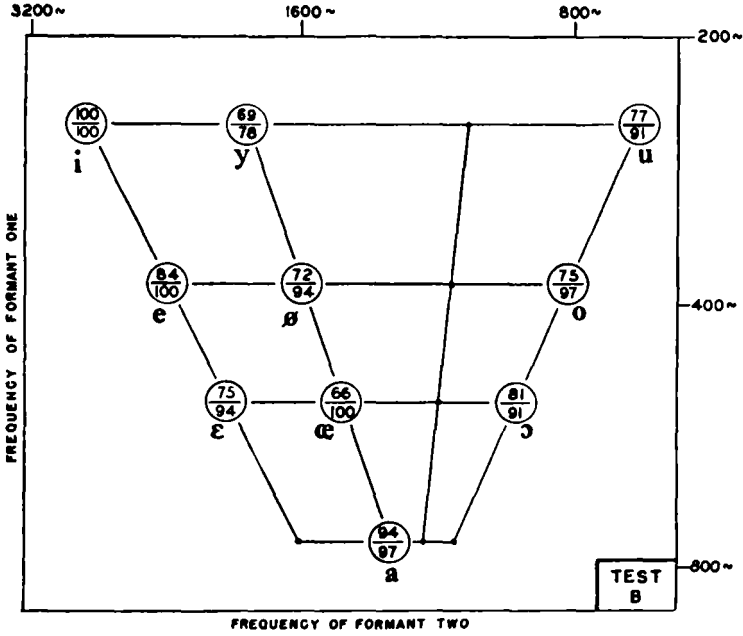


FIG. 4. The accuracy of listeners' identifications of the 10 synthetic vowels that have analogues in French. Percentages of correct identification are indicated as in Fig. 3.

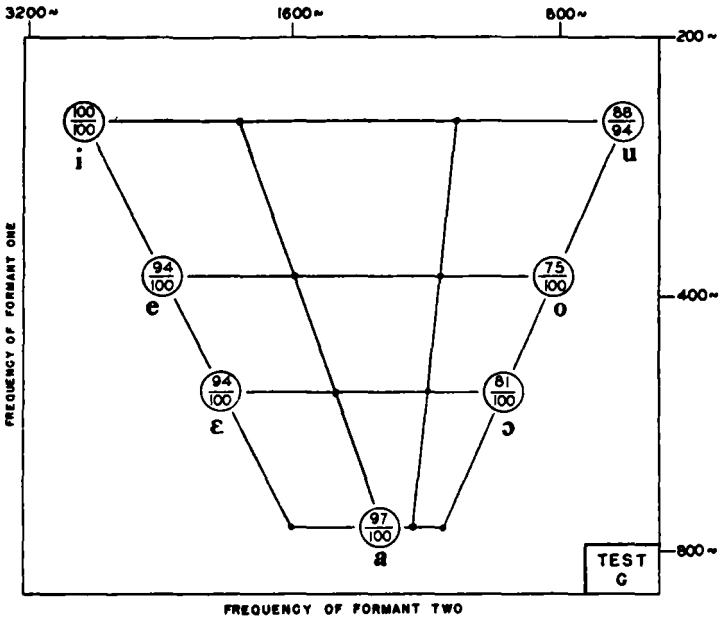


FIG. 5. The accuracy of listeners' identifications of seven of the "outside" synthetic vowels. Percentages of correct identification are indicated as in Fig. 3.



evidence that a familiar vowel at the outside edges of the chart will be identified more easily than an equally familiar vowel which has its locus on the inside quadrilateral (e.g., / $\Delta$ / was identified less accurately than / $\circ$ /). If we assume that distances on our vowel charts (Figs. 2, 3, 4, 5) correspond, however roughly, to perceived similarities, we may suppose that the "outside" vowels, having fewer neighboring vowels with which to be confused, are at an advantage by comparison with the "inside" vowels.

#### 4. EFFECTS OF INTENSITY VARIATIONS

Starting with the "equalized"<sup>8</sup> formant intensities of our synthetic vowels, we proceeded to reduce by small steps the intensity of each formant separately (until the formant was effectively extinguished), and to observe the effects on the vowel sound as heard. The reduction in intensity was accomplished by interposing filters of appropriate density in the path of the modulated light beams, thus reducing the intensity of the light before it reached the painted formant.

4.1. *The effects of reducing the intensity of the lower formant.* For those vowels in which the two formants are relatively far apart /i e  $\epsilon$  y  $\emptyset$   $\text{\textcircled{a}}$ /, progressive reductions in the intensity of the first formant resulted finally, for each of these vowels, in a destruction of vowel color. Beyond a certain amount of attenuation, somewhat different for each of the vowels listed above, the vowel color was replaced by a non-vowel sound which corresponded to the pitch of the higher formant. In no case did a reduction in the intensity of the lower formant cause a clear *shift* in vowel color.

In the case of those vowels whose two formants are relatively close together /u o  $\text{\textcircled{a}}$  / a very different result was found. Here reductions in the intensity of the lower formant caused, not a loss of color, but rather a change in color toward an adjacent vowel. /u/ went to close / $\text{\textcircled{a}}$ /, /o/ to open / $\text{\textcircled{a}}$ /, / $\text{\textcircled{a}}$ / to /a/, and /a/ to / $\text{\textcircled{a}}$ /. This finding can, perhaps, be accounted for if we assume that the ear effectively averages two vowel formants which are close together, receiving from these two formants an impression which is highly similar to that which would be heard from one formant placed at a position somewhere intermediate between them. We should suppose, then, that reducing the intensity of the lower formant, for example, would have the effect of increasing the higher formant's relative contribution to the "mean" and would thus effectively raise the mean formant. On this basis reducing the intensity of the lower formant of /u/ could raise the mean impression to a level equivalent to the mean impression received from the normal formant intensities of a close / $\text{\textcircled{a}}$ /. All of this suggests, of course, that

<sup>8</sup> For the purposes of this part of the investigation, all the vowel formants, including those of /o/ and /u/, were made equally wide; when converted into sound, however, they were not of the same intensity, inasmuch as the playback is so constructed as to produce a reduction in intensity along the frequency scale amounting to 9 db per octave in the frequency range above 1500 cps. As we pointed out in Section 2, this attenuation corresponds roughly to the distribution of energy in normal human speech; it should be noted, however, that this rate of reduction in intensity is correct only on the average and hence is not necessarily perfectly appropriate for any particular speech sound.

those back vowels in which the formants are close enough to permit averaging are, in effect, one-formant vowels.<sup>9</sup>

With the vowels /u ɤ ʌ/, reductions in the intensity of the first formant caused /u/ to become vague /ɑ/, /ɤ/ to become vague /a/, and /ʌ/ to become vague /æ/. It should be noted that these color changes are superficially quite different from those which occurred when the first formant of the back vowels was reduced in intensity. In the case of the back vowels the shift in color was to a vowel quite near to the original (e.g., /ɔ/ went to /ɑ/), whereas the vowels /u ɤ ʌ/ move over a rather considerable distance, e.g., /u/ shifts to /ɑ/. (The term "distance" may here be taken in its literal sense to refer to the two-dimensional space in which the synthetic vowels are arrayed in Fig. 2.) As has been pointed out, the shifts which occur with the back vowels suggest that the ear averages the two not-widely-separated formants of the back vowels, which is to say that the two formants should be replaceable by one formant which is properly intermediate. In the case of /u ɤ ʌ/ we may suppose, to take one group of examples, that the second formant frequency of /ɤ/ (1100 cps) is equal to the mean impression of the two formants (750 cps and 1300 cps) of /a/. Reducing the intensity of the lower formant of /ɤ/ leaves us then with a single formant whose frequency happens to correspond to the "mean" impression created by the lower and higher formants of a vowel /a/, although the latter is quite far removed from /ɤ/ on the vowel chart.

It was also noted that /a ʌ ɔ ɑ/ became slightly nasalized when the intensity of the first formant was reduced by as much as 7 db.

4.2. *The effects of reducing the intensity of the higher formant.* For /i e ε y ø œ/, and also /u ɤ ʌ/, small attenuations of the higher formant caused the vowel to acquire a quality that can best be described as "dull." With further reductions in the intensity of the higher formant each of these vowels changed to a close form of that particular back vowel whose first formant it shares. Thus /i y u/ went to close and rather vague /u/; /e ø ɤ/ became a close and vague /o/; /ε œ ʌ/ became a close and vague /ɔ/. Reducing the intensity of the second formants of the back vowels /u o ɔ/ has, of course, the same final result since /u/ has the same first formant as /i y u/, /o/ has the same first formant as /e ø ɤ/, and /ɔ/ has the same first formant as /ε œ ʌ/: /u/ became a close /u/, /o/ became a close /o/, and /ɔ/ became a close /ɔ/. When the second formant of /æ a ɒ ɑ/ is reduced there is a shift in vowel color to open /ɔ/.

In all these cases reduction in the intensity of the higher formant leaves us finally with a single formant of relatively low frequency, and we should suppose, on the basis of our assumptions about "averaging," that the vowel color of that one formant would correspond to an "average" impression produced by the two

<sup>9</sup> Fletcher pointed out in 1929 that early investigators had failed to detect the relatively weak second formant of the back vowels and had concluded, therefore, that these vowels were "singly resonant." (See H. Fletcher, *Speech and Hearing*, New York, D. Van Nostrand, 1929). Our results suggest that these back vowels may, indeed, be single-formant vowels, but only in the very special sense that, because of "averaging," formants 1 and 2 of the spoken vowel can be replaced, in any synthetic production, by a single formant which is located at an intermediate position.

formants of some back vowel. For /æ a ɒ ɑ/ we see clearly that this happens: the lower formant of these vowels (750 cps) is intermediate between the first (510 cps) and second (950 cps) formants of /ɔ/, and we find, as reported in the preceding paragraph, that the first formant of these vowels, by itself, does indeed sound like a vowel in the /ɔ/ category. We might have expected, similarly, that the first formants of /ε œ ʌ ɔ/ (510 cps) would sound like /o/ (first formant, 360 cps, second formant 800 cps), and that the first formant of /e ø ɣ o/ would sound like /u/ (first formant, 250 cps, second formant, 700 cps). In fact, /ε œ ʌ ɔ/ move only to close /ɔ/, when the higher formant is omitted, and /e ø ɣ o/ go only as far as close /o/. We should suppose, then, that in the case of the vowels /o/ and /u/ the correct average is at a value very close to the first formant.

##### 5. ONE-FORMANT EQUIVALENTS OF THE SYNTHETIC TWO-FORMANT VOWELS

The results described in the preceding section imply that the two formants of some of the back vowels are replaceable by a single formant. We have attempted by several techniques to secure evidence which might bear rather more directly on this question. In one exploratory investigation we tried simply to find those single formants which will most closely approximate the 10 synthetic cardinal vowels /i e ε æ a ɒ ɑ ɔ o u/ which lie on the perimeter of the vowel chart. For each two-formant vowel we listened to a series of single formants which sampled the frequency range between the first and second formants of the two-formant vowel. (These single formants were  $2\frac{1}{2}$  harmonics wide, and successive single formants were 120 cps apart.) The sounds were arranged by pairs: the first member of the pair was the two-formant vowel, and the second member of the pair was one of the above series of one-formant vowels. The judgments were made by a guest phonetician and the authors of this paper. In general, the results were reasonably consistent with the conclusions obtained by varying the relative intensities of the two formants (Section 4). For the vowels /u o ɔ ɒ ɑ a/, where the two formants are relatively close together, the judges were able to find reasonably good one-formant equivalents. The best one-formant approximations were very near in frequency to the *first* formants (of the two-formant synthetic vowels) for the two extreme back vowels /u/ and /o/, and were, in general, at positions progressively nearer the middle of the frequency interval separating the two formants of the synthetic vowels in the series /ɔ ɒ ɑ a/. It was difficult to find one-formant equivalents for /æ ε e/, in which the two-formants are rather far apart, but for /i/, which has the greatest separation of first and second formants, a single formant near the normal position of the *second* formant seemed to produce the /i/ color rather well.<sup>10</sup>

It is perhaps worth noting that the results were quite different when we attempted to match single intermediate formants to arbitrarily chosen two-formant patterns which did not sound like vowels. In these cases we found it

<sup>10</sup> It is relevant here that /i/ is an extreme vowel (none has a higher second formant), and the kind of judgment-anchoring which so often occurs at the extremes of a stimulus series might account for the fact that a high-frequency formant was so often identified as /i/.

generally impossible to locate a single formant which sounded like the two-formant pattern. It was our impression that the two-formants of the non-vowel patterns, however close together they might be, did not fuse into a single sound, but tended rather to be heard as two-component chords. (We have noticed the same kind of thing in listening to those synthetic cardinal vowels which are outside the range of our linguistic experience.) Inasmuch as the two formants of the non-vowels do not blend into a single sound, one ought not, perhaps, to expect to find an equivalent "average."

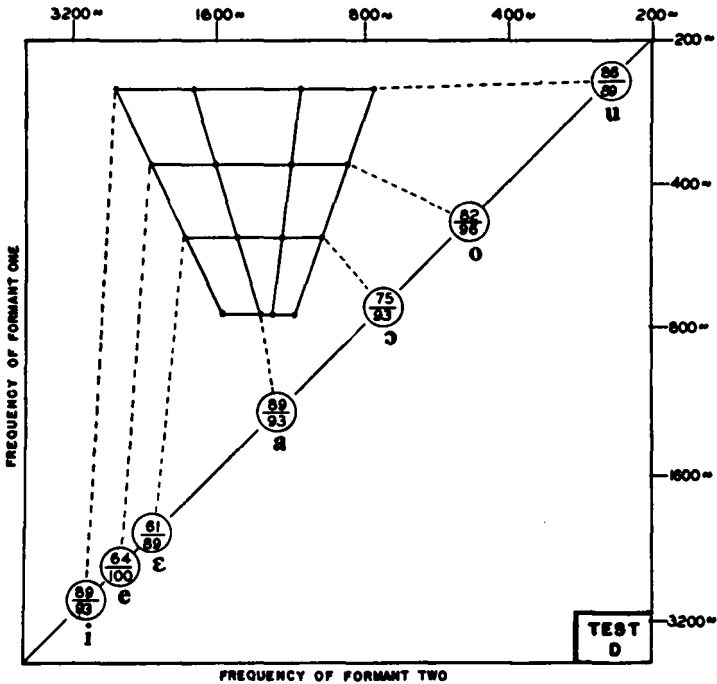


FIG. 6. The accuracy of listeners' identifications of the seven one-formant vowels. Each one-formant vowel has a locus on the straight line where formant one equals formant two. The broken lines direct the eye to the original two-formant positions on the vowel chart. Percentages of correct identification are indicated as in Fig. 3.

As a further test of the apparent averaging of the two-formant vowels, we constructed a series of 24 one-formant representations which sampled the range 240 cps to 3000 cps at intervals of 120 cps. (Again, each formant was  $2\frac{1}{2}$  harmonics wide.) These sounds were arranged in a random order and presented to five listeners (one phonetician and the authors) with instructions to try to identify each sound as a vowel and to indicate on a three-point scale how "good" a vowel it was. There was considerable agreement among the judges in identifying the vowels /u/ through /ɔ/—/u/ at 240 cps, /o/ at 360 and 480 cps, and /ɔ/ at 600 and 720 cps—and these one-formant sounds were felt by all the judges to be rather highly identifiable. The judges did not agree so well in identifying

the other stimuli, and were generally less confident about their judgments, except in the case of formants above 2760 cps which were unanimously called /i/. For /e ε a/, the closest approximations appeared to be 2520 cps, 2160 cps, and 1200 cps, respectively.<sup>11</sup>

Using the results of these two exploratory studies as a basis, we selected the best one-formant approximations to seven of the "outside" cardinal vowels /i e ε a ɔ o u/ and arranged them for Test D in a format identical with that used to determine the identifiability of the two-formant vowels in Test C (see

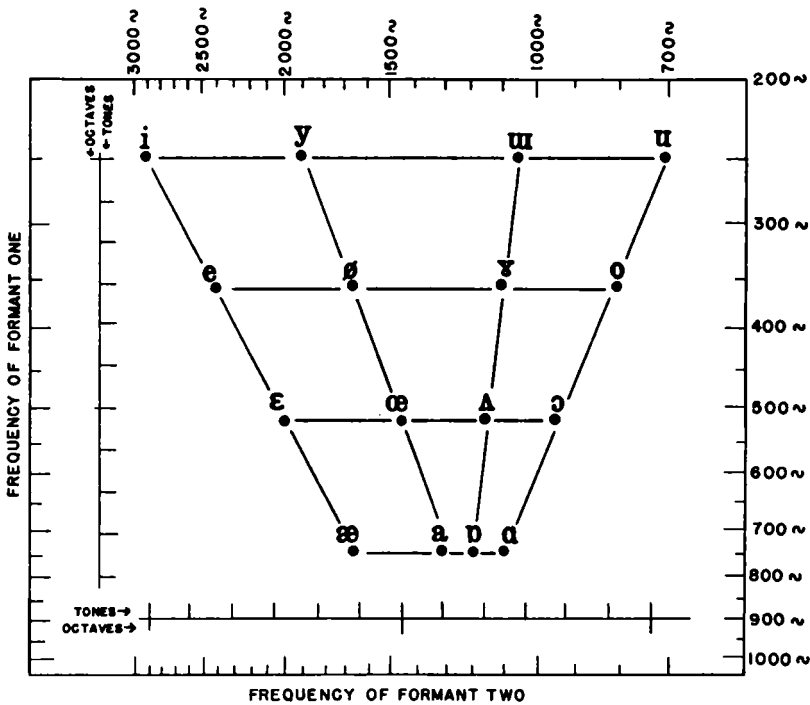


FIG. 7. Acoustic locations of spoken French vowels for comparison with the synthetic vowels of Fig. 2. The French vowels, indicated by crosses, are plotted according to the frequencies of their first and second formants.

Section 3). The center frequencies of the single-formant vowels in Test D were: /u/—240 cps; /o/—480 cps; /ɔ/—720 cps; /a/—1200 cps; /ε/—2160 cps; /e/—2520 cps; /i/—3000 cps. These sounds were presented to 14 of the 16 listeners who had judged the two-formant vowels of Tests B and C. The results of this test are presented in Fig. 6. By comparing the percentages of correct identification in Test D with the analogous values in Test C (Fig. 5) we see

<sup>11</sup> Köhler matched single pure tones to the pitch of sung vowels and arrived at equivalent frequencies of 3480 for /i/, 2265 for /e/, 1140 for /a/, 470 for /o/, and 225 for /u/. These values are very close to those which we arrived at by the very different technique described above. Köhler's results are reported in E. G. Boring, *Sensation and Perception in the History of Experimental Psychology*, Appleton-Century-Crofts, New York, 1942, 373-4.

that the identifications were, in general, less accurate for the one-formant than for the two-formant vowels. Of greater interest, however, is the fact that, with the exception of /i/, the one-formant front vowels lost relatively more in identifiability than the one-formant back vowels. (The percentages of correct identification go from 100 for /i/, 94 for /e/, and 94 for /ɛ/, in the two-formant version, to 89 for /i/, 64 for /e/, and 61 for /ɛ/ in the one-formant version; the values for the back vowels, on the other hand, go from 88 for /u/, 75 for /o/, and 81 for /ɔ/ in the two-formant version, to 86 for /u/, 82 for /o/, and 75 for /ɔ/ in the one-formant case.) This result would appear in general to confirm the conclusion already presented that the back vowels (whose formants are close together) can be rather closely approximated by a single formant, whereas the front vowels, with the possible exception of /i/, cannot.

#### 6. COMPARISON OF SPOKEN AND SYNTHETIC VOWELS

Figure 7 provides a basis for comparing the formant frequencies of our two-formant synthetic vowels with the frequencies of spoken French vowels.<sup>12</sup> Of the 16 synthetic vowels, 10 have their analogues in spoken French.

We see that the frequencies of the lower formants are about the same for synthetic and spoken vowels in the case of the close, mid-close, and mid-open vowels. For the open vowels the formant 1 frequencies of the synthetic vowels are slightly higher than those of the spoken vowels.

The frequencies of the higher formant show much larger differences between synthetic and spoken vowels. The largest difference, as can be seen in the figure, is for /i/. A most extreme French /i/ has its second formant at 2500 cps, while our synthetic /i/ has the second formant at 2900 cps, a difference of 1.3 tones. For /e/ there is a difference which is somewhat smaller ( $\frac{3}{4}$  of a tone) but in the same direction; also, for /ɛ y ø œ/, the second formant of the synthetic vowel is higher than that of the spoken vowel, although the difference is less than in /i/ or /e/, amounting only to about  $\frac{1}{4}$  of a tone. These differences between synthetic and spoken vowels can, perhaps, be accounted for on the basis of two assumptions: (1) that the third formant (of spoken vowels) makes a significant contribution to vowel color in the case of these (front) vowels, and (2) that the ear effectively averages two formants—the second and third, in this case—which are relatively close together. (The latter assumption has been made and supported in the preceding sections of this paper.) What we have done, then, in our synthetic vowels is to arrive at a "compromise" second formant which has a position somewhere between the second and third formants of spoken vowels, this compromise formant being closer to the third formant (as in /i/) or closer to the second (as in /ɛ/) depending on the relative importance of the second and third formants in determining vowel color.

If one attempts to define the essential acoustic features of vowels by examining spectrograms of actual speech, he is tempted to conclude that the third formant

<sup>12</sup> For a specification of the frequency positions of the first two (lower) formants of French spoken vowels, see Pierre Delattre, "Un triangle acoustique des voyelles orales du français," *French Review*, XXI, 6, May, 1948.

is of negligible importance for the perception of the vowel, since its frequency position changes very little from vowel to vowel. This is, of course, an altogether reasonable inference from the spectrographic data, but if our interpretation of our own results is correct, we must assume that the third formant, however stationary, can be quite important for its contribution to the "mean" impression of formants two and three when these are close together.

For the back vowels /u o ɔ/ we see that the synthetic versions have second formants somewhat higher than the spoken vowels. This difference amounts to  $\frac{2}{3}$  to  $\frac{1}{2}$  tone; we have found, however, that the higher formant of these back vowels can be lowered in frequency to the extent of two tones or more without destroying the cardinal vowel color.

#### SUMMARY

A series of 235 two-formant patterns, representing systematic variations in formant-frequency position, were prepared and converted into sound by an instrument called a "pattern playback." From this experimental series, the authors selected those sounds which, in their collective judgment, most closely approximated the 16 cardinal vowels of the International Phonetic Association. These 16 synthetic sounds were presented to a group of students in phonetics for identification as vowels; the results indicated that the synthetic vowels were rather highly identifiable.

Variations in the relative intensities of the formants in the two-formant synthetic vowels caused some of the sounds simply to lose vowel color and to become vague, while others seemed to change color and to become different vowels. The changes in color which did occur are all consistent with the assumption that the ear effectively "averages" two formants which are relatively close together (as is the case for the back vowels), and receives from them an over-all quality roughly equivalent to that which would be produced by a single intermediate formant.

To test this assumption more directly, an attempt was made to locate those single formants which, as heard, most closely resemble the synthetic two-formant vowels. In general, it was relatively easier to find satisfactory one-formant equivalents for the back vowels (whose first and second formants are rather close in frequency) than for the front vowels (where the frequency separation of the formants is relatively great). An exception to this generalization is the vowel /i/, which can be very closely approximated by a single high-frequency formant. The best one-formant approximations to seven of the cardinal vowels /i e ε a ɔ o u/ were selected, and their identifiability was tested with the same listeners and by the same technique used to measure the identifiability of the two-formant synthetic vowels. As might be expected on the assumption of "averaging" and from the observations about the relative ease or difficulty of finding one-formant equivalents, the identifiability of the back vowels was not appreciably less in the one-formant than in the two-formant form, while, with the exception of /i/, the difficulty of identifying the one-formant front vowels was rather considerably increased.

A comparison of the frequency positions of the two-formant synthetic vowels with the first- and second-formant positions of the corresponding French vowels *as spoken* shows certain differences, especially in the second formants of the front vowels. These differences appear to be explicable on the assumption that the third formant of the spoken front vowels makes an important contribution to their color and, further, that the second formant of the synthetic vowels, which is generally higher than the second formant of the spoken vowels, represents again an "average" of formants which lie close together in the spoken vowel.

In conclusion we should like to acknowledge that 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.

*Haskins Laboratories*