Control Methods Used in a Study of the Vowels

GORDON E. PETERSON AND HAROLD L. BARNEY

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received December 3, 1951)

Relationships between a listener's identification of a spoken vowel and its properties as revealed from acoustic measurement of its sound wave have been a subject of study by many investigators. Both the utterance and the identification of a vowel depend upon the language and dialectal backgrounds and the vocal and auditory characteristics of the individuals concerned. The purpose of this paper is to discuss some of the control methods that have been used in the evaluation of these effects in a vowel study program at Bell Telephone Laboratories. The plan of the study, calibration of recording and measuring equipment, and methods for checking the performance of both speakers and listeners are described. The methods are illustrated from results of tests involving some 76 speakers and 70 listeners.

INTRODUCTION

CONSIDERABLE variation is to be found in the processes of speech production because of their complexity and because they depend upon the past experience of the individual. As in much of human behavior there is a self-correcting, or servomechanism type of feedback involved as the speaker hears his own voice and adjusts his articulatory mechanisms.[1]

In the elementary case of a word containing a consonant-vowel-consonant phoneme[2,3] structure, a speaker's pronunciation of the vowel within the word will be influenced by his particular dialectal background; and his pronunciation of the vowel may differ both in phonetic quality and in measurable characteristics from that produced in the word by speakers with other backgrounds. A listener, likewise, is influenced in his identification of a sound by his past experience.

Variations are observed when a given individual makes repeated utterances of the same phoneme. A very significant property of these variations is that they are not random in a statistical sense, but show trends and sudden breaks or shifts in level, and other types of nonrandom fluctuations.[4] Variations likewise appear in the successive identifications by a listener of the same utterance. It is probable that the identification of repeated sounds is also nonrandom but there is little direct evidence in this work to support such a conclusion.

A study of sustained vowels was undertaken to investigate in a general way the relation between the vowel phoneme intended by a speaker and that identified by a listener, and to relate these in turn to acoustical measurements of the formant or energy concentration positions in the speech waves.

In the plan of the study certain methods and techniques were employed which aided greatly in the collection of significant data. These methods included randomization of test material and repetitions to ob-

tain sequences of observations for the purpose of checking the measurement procedures and the speaker and listener consistency. The acoustic measurements were made with the sound spectrograph; to minimize measurement errors, a method was used for rapid calibration of the recording and analyzing apparatus by means of a complex test tone. Statistical techniques were applied to the results of measurements, both of the calibrating signals and of the vowel sounds.

These methods of measurement and analysis have been found to be precise enough to resolve the effects of different dialectal backgrounds and of the nonrandom trends in speakers' utterances. Some aspects of the vowel study will be presented in the following paragraphs to illustrate the usefulness of the methods employed.

EXPERIMENTAL PROCEDURES

The plan of the study is illustrated in Fig. 1. A list of words (List 1) was presented to the speaker and his utterances of the words were recorded with a magnetic tape recorder. The list contained ten monosyllabic words each beginning with [h] and ending with [d] and differing only in the vowel. The words used were *heed, hid, head, had, hod, hawed, hood, who'd, hud,* and *heard.* The order of the words was randomized in each list, and each speaker was asked to pronounce two different lists. The purpose of randomizing the words in the list was to avoid practice effects which would be associated with an unvarying order.

If a given List 1, recorded by a speaker, were played back to a listener and the listener were asked to write down what he heard on a second list (List 2), a comparison of List 1 and List 2 would reveal occasional
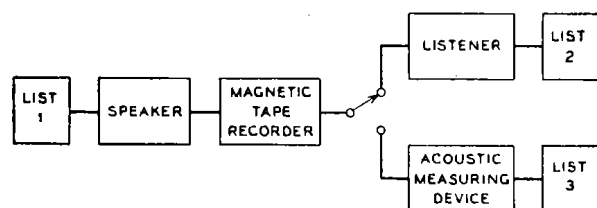
[1] Bernard S. Lee, J. Acoust. Soc. Am. **22**, 824 (1950).
[2] B. Bloch, Language **24**, 3 (1948).
[3] B. Bloch, Language **26**, 88 (1950).
[4] R. K. Potter and J. C. Steinberg, J. Acoust. Soc. Am. **26**, 807 (1950).

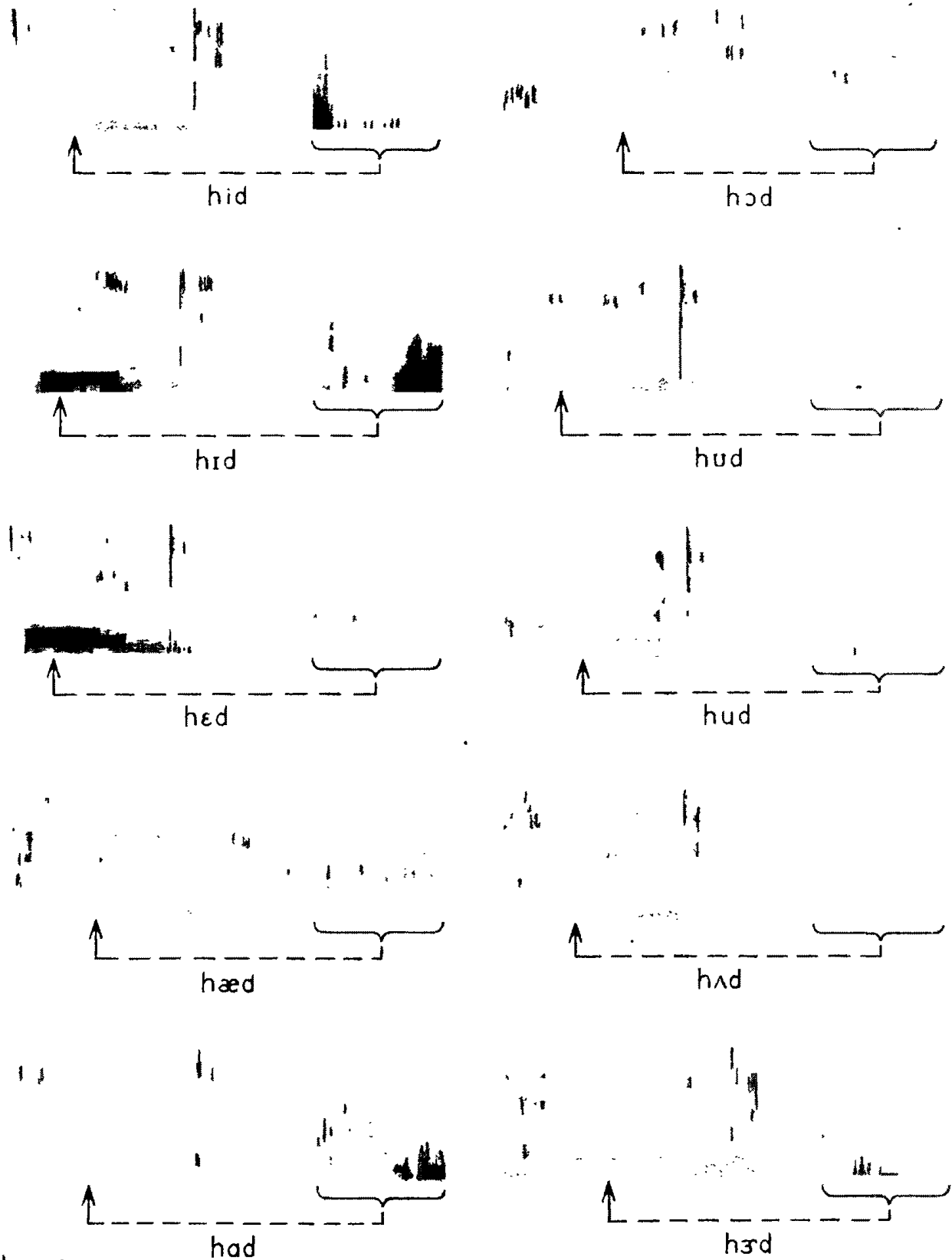FIG. 1. Recording and measuring arrangements for vowel study.

Fig. 2. Broad band spectrograms and amplitude sections of the word list by a female speaker.

differences, or disagreements, between speaker and listener. Instead of being played back to a listener, List 1 might be played into an acoustic measuring device and the outputs classified according to the measured properties of the sounds into a List 3. The

three lists will differ in some words depending upon the characteristics of the speaker, the listener, and the measuring device.

A total of 76 speakers, including 33 men, 28 women and 15 children, each recorded two lists of 10 words,

making a total of 1520 recorded words. Two of the speakers were born outside the United States and a few others spoke a foreign language before learning English. Most of the women and children grew up in the Middle Atlantic speech area.[5] The male speakers represented a much broader regional sampling of the United States; the majority of them spoke General American.[5]

The words were randomized and were presented to a group of 70 listeners in a series of eight sessions. The listening group contained only men and women, and represented much the same dialectal distribution as did the group of speakers, with the exception that a few observers were included who had spoken a foreign language throughout their youth. Thirty-two of the 76 speakers were also among the 70 observers.

The 1520 words were also analyzed by means of the sound spectrograph.[6,7]

Representative spectrograms and sections of these words by a male speaker are shown in Fig. 3 of the paper by R. K. Potter and J. C. Steinberg;[4] a similar list by a female speaker is shown here as Fig. 2.[8] In the spectrograms, we see the initial [h] followed by the vowel, and then by the final [d]. There is generally a part of the vowel following the influence of the [h] and preceding the influence of the [d] during which a practically steady state is reached. In this interval, a section is made, as shown to the right of the spectrograms. The sections, portraying frequency on a horizontal scale, and amplitude of the voiced harmonics on the vertical side, have been measured with calibrated Plexiglass templates to provide data about the fundamental and formant frequencies and relative formant amplitudes of each of the 1520 recorded sounds.

LISTENING TESTS

The 1520 recorded words were presented to the group of 70 adult observers over a high quality loud speaker system in Arnold Auditorium at the Murray Hill Laboratories. The general purpose of these tests was to obtain an aural classification of each vowel to supplement the speaker's classification. In presenting the words to the observers, the procedure was to reproduce at each of seven sessions, 200 words recorded by 10 speakers. At the eighth session, there remained five men's and one child's recordings to be presented; to these were added three women's and one child's recordings which had been given in previous sessions, making again a total of 200 words. The sound level at the observers' positions was approximately 70 db re 0.0002 dyne/cm², and varied over a range of about 3 db at the different positions.

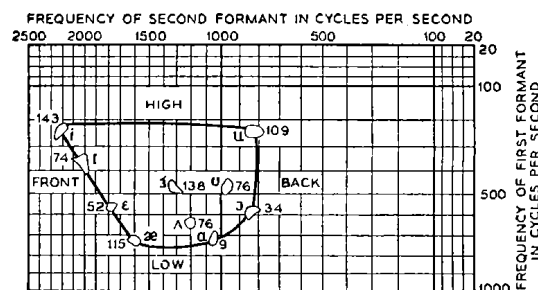In selecting the speakers for each of the first seven

FIG. 3. Vowel loop with numbers of sounds unanimously classified by listeners; each sound was presented 152 times.

sessions, 4 men, 4 women, and 2 children were chosen at random from the respective groups of 33, 28, and 15. The order of occurrence of the 200 words spoken by the 10 speakers for each session was randomized for presentation to the observers.

Each observer was given a pad containing 200 lines having the 10 words on each line. He was asked to draw a line through the one word in each line that he heard. The observers' seating positions in the auditorium were chosen by a randomizing procedure, and each observer took the same position for each of the eight sessions, which were given on eight different days.

The randomizing of the speakers in the listening sessions was designed to facilitate checks of learning effects from one session to another. The randomizing of words in each group of 200 was designed to minimize successful guessing and the learning of a particular speaker's dialect. The seating positions of the listeners were randomized so that it would be possible to determine whether position in the auditorium had an effect on the identification of the sounds.

DISCUSSION OF LISTENING TEST RESULTS

The total of 1520 sounds heard by the observers consisted of the 10 vowels, each presented 152 times. The ease with which the observers classified the various vowels varied greatly. Of the 152 [i] sounds, for instance, 143 were unanimously classified by all observers as [i]. Of the 152 sounds which the speakers intended for [ɑ], on the other hand, only 9 were unanimously classified as [ɑ] by the whole jury.

These data are summarized in Fig. 3. This figure shows the positions of the 10 vowels in a vowel loop in which the frequency of the first formant is plotted against the frequency of the second formant[9] on mel scales;[10] in this plot the origin is at the upper right. The numbers beside each of the phonetic symbols are the numbers of sounds, out of 152, which were unanimously classified as that particular vowel by the jury. It is of interest in passing that in no case did the jury agree unanimously that a sound was something other than what the speaker intended. Figure 3 shows that

[5] C. K. Thomas, *Phonetics of American English*, The Ronald Press Company (New York, 1947).

[6] Koenig, Dunn, and Lacy, J. Acoust. Soc. Am. 17, 19 (1946).

[7] L. G. Kersta, J. Acoust. Soc. Am. 20, 796 (1948).

[8] Key words for the vowel symbols are as follows: [i] heed, [ɪ] hid, [ɛ] head, [æ] had, [ɑ] father, [ɔ] ball, [ʊ] hood, [u] who'd, [ʌ] hud, [ɝ] heard.

[9] R. K. Potter and G. E. Peterson, J. Acoust. Soc. Am. 20, 528 (1948).

[10] S. S. Stevens and J. Volkman, Am. J. Psychol. 329 (July, 1940).

[i], [ɝ], [æ], and [u] are generally quite well understood.

To obtain the locations of the small areas shown in Fig. 3, the vowels were repeated by a single speaker on twelve different days. A line enclosing all twelve points was drawn for each vowel; the differences in the shapes of these areas probably have little significance.

When the vowels are plotted in the manner shown in Fig. 3, they appear in essentially the same positions as those shown in the tongue hump position diagrams which phoneticians have employed for many years.[11] The terms "high, front, low back" refer to the tongue positions in the mouth. The [i], for instance, is made with the tongue hump high and forward, the [u] with the hump high and back, and the [ɑ] and [æ] with the tongue hump low.

It is of interest that when observers disagreed with speakers on the classification of a vowel, the two classifications were nearly always in adjacent positions of the vowel loop of Fig. 3. This is illustrated by the data shown on Table I. This table shows how the observers classified the vowels, as compared with the vowels intended by the speakers. For instance, on all the 152 sounds intended as [i] by the speakers, there were 10,267 total votes by all observers that they were [i], 4 votes for [ɪ], 6 votes for [ɛ], and 3 votes for [ɔ]. Of the 152 [ɑ] sounds, there was a large fraction of the sounds on which some of the observers voted for [ɔ]. [ɪ] was taken for [ɛ] a sizable percentage of the time, and [ɛ] was called either [ɪ] or [æ] (adjacent sounds on the vowel loop shown in the preceding Fig. 3) quite a large number of times. [ɑ] and [ɔ], and [ʌ] and [ɑ] were also confused to a certain extent. Here again, as in Fig. 2, the [i], [ɝ], [æ], and [u] show high intelligibility scores.

It is of considerable interest that the substitutions shown conform to present dialectal trends in American speech rather well,[12] and in part, to the prevailing vowel shifts observable over long periods of time in most languages.[13] The common tendency is continually to shift toward higher vowels in speech, which correspond to smaller mouth openings.

The listener, on the other hand, would tend to make the opposite substitution. This effect is most simply described in terms of the front vowels. If a speaker produces [ɪ] for [ɛ], for example [mɪn] for [mɛn] as currently heard in some American dialects; then such an individual when serving as a listener will be inclined to write *men* when he hears [mɪn]. Thus it is that in the substitutions shown in Table I, [ɪ] most frequently became [ɛ], and [ɛ] most frequently became [æ]. The explanation of the high intelligibility of [æ] is probably based on this same pattern. It will be noted along the

vowel loop that a wide gap appears between [æ] and [ɑ]. The [a] of the Romance languages appears in this region. Since that vowel was present in neither the lists nor the dialects of most of the speakers and observers the [æ] was usually correctly identified.

The [i] and the [u] are the terminal or end positions in the mouth and on the vowel loop toward which the vowels are normally directed in the prevailing process of pronunciation change. In the formation of [i] the tongue is humped higher and farther forward than for any other vowel; in [u] the tongue hump takes the highest posterior position in the mouth and the lips are more rounded than for any other vowel. The vowels [u] and [i] are thus much more difficult to displace, and a greater stability in the organic formation of these sounds would probably be expected, which in turn should mean that these sounds are recognized more consistently by a listener.

The high intelligibility of [ɝ] probably results from the retroflexion which is present to a marked degree only in the formation of this vowel; that is, in addition to the regular humping of the tongue, the edges of the tongue are turned up against the gum ridge or the hard palate. In the acoustical pattern the third formant is markedly lower than for any other vowel. Thus in both physiological and acoustical phonetics the [ɝ] occupies a singular position among the American vowels.

The very low scores on [ɑ] and [ɔ] in Fig. 3 undoubtedly result primarily from the fact that some members of the speaking group and many members of the listening group speak one of the forms of American dialects in which [ɑ] and [ɔ] are not differentiated.

When the individuals' votes on the sounds are analyzed, marked differences are seen in the way they classified the sounds. Not only did the total numbers of agreements with the speakers vary, but the proportions of agreements for the various vowels was significantly different. Figure 4 will be used to illustrate this point. If we plot total numbers of disagreements for all tests, rather than agreements, the result is shown by the upper chart. This shows that [ɪ], [ɛ], [ɑ], [ɔ], and [ʌ] had the most disagreements. An "average" observer would be expected to have a distribution of disagreements similar in proportions to this graph. The middle graph illustrates the distribution of disagreements given by observer number 06. His chief difficulty was in distinguishing between [ɑ] and [ɔ]. This type of distribution is characteristic of several observers. Observer 013, whose distribution of disagreements is plotted on the bottom graph, shows a tendency to confuse [ɪ] and [ɛ] more than the average.

The distributions of disagreements of all 70 observers differ from each other, depending on their language experience, but the differences are generally less extreme than the two examples shown on Fig. 4. Thirty-two of the 70 observers were also speakers. In cases where an observer such as 06 was also a speaker, the remainder of the jury generally had more disagreements

[11] D. Jones, *An Outline of English Phonetics* (W. Heffer and Sons, Ltd., Cambridge, England, 1947).

[12] G. W. Gray and C. M. Wise, *The Bases of Speech* (Harper Brothers, New York, 1946), pp. 217–302.

[13] L. Bloomfield, *Language* (Henry Holt and Company, New York, 1933), pp. 369–391.

with his [ɑ] and [ɔ] sounds than with the other sounds he spoke. Thus it appears that if a speaker does not differentiate clearly between a pair of sounds in speaking them, he is unlikely to classify them properly when he hears others speak them. His language experience, as would be expected, influences both his speaking and his hearing of sounds.

Since the listening group was not given a series of training sessions for these tests, learning would be expected in the results of the tests.[14] Several pieces of evidence indicate a certain amount of practice effect, but the data are not such as to provide anything more than a very approximate measure of its magnitude.

For one check on practice effect, a ninth test was given the jury, in which all the words having more than 10 disagreements in any of the preceding eight tests were repeated. There was a total of about 175 such words; to these were added 25 words which had no disagreements, picked at random from the first eight tests. On the ninth test, 67 words had more disagreements, 109 had less disagreements, and 24 had the same number of disagreements as in the preceding tests. The probability of getting this result had there been no practice or other effect, but only a random variation of observers' votes, would be about 0.01. When these data are broken down into three groups for the men, women and children speakers, the largest differences in numbers of disagreements for the original and repeated tests was on the childrens' words, indicating a larger practice or learning effect on their sounds. The indicated learning effect on men's and women's speech was nearly the same. When the data are classified according to the vowel sound, the learning effect indicated by the repetitions was least on [i], [ɝ], and [u], and greatest on [ɑ] and [ɔ].

Another indication that there was a practice effect lies in the sequence of total numbers of disagreements by tests. From the second to the seventh test, the total number of disagreements by all observers diminished consistently from test to test, and the first test had considerably more disagreements than the eighth, thus strongly indicating a downward trend. With the speakers randomized in their order of appearance in the eight tests, each test would be expected to have approximately the same number of disagreements. The probability of getting the sequence of numbers of total disagreements which was obtained would be somewhat less than 0.05 if there were no learning trend or other non-random effect.

It was also found that the listening position had an effect upon the scores obtained. The observers were arranged in 9 rows in the auditorium, and the listeners in the back 4 rows had a significantly greater number of disagreements with the speakers than did the listeners in the first 5 rows. The effect of a listener's position
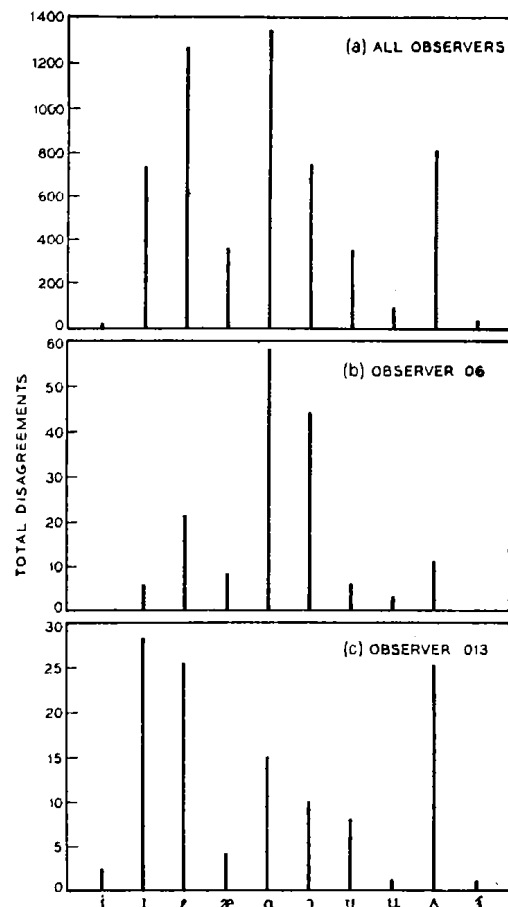

FIG. 4. Observer disagreements in listening tests.

within an auditorium upon intelligibility has been observed previously and is reported in the literature.[15]

ACOUSTIC MEASUREMENTS

Calibrations of Equipment

A rapid calibrating technique was developed for checking the over-all performance of the recording and analyzing systems. This depended on the use of a test tone which had an envelope spectrum that was essentially flat with frequency over the voice band. The circuit used to generate this test tone is shown schematically in Fig. 5. It consists essentially of an overloading amplifier and pulse sharpening circuit. The wave shapes which may be observed at several different points in the test tone generator are indicated in Fig. 5.

The test tone generator may be driven by an input sine wave signal of any frequency between 50 and 2000 cycles. Figure 6(a) shows a section of the test tone with a 100 cycle repetition frequency, which had been recorded on magnetic tape in place of the word lists by the speaker, and then played back into the sound spectrograph. The departure from uniform frequency response of the over-all systems is indicated by the shape of the envelope enclosing the peaks of the 100

[14] H. Fletcher and R. H. Galt, J. Acoust. Soc. Am. 22, 93 (1950).

[15] V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture* (John Wiley and Sons, New York, 1950), pp. 180–181.
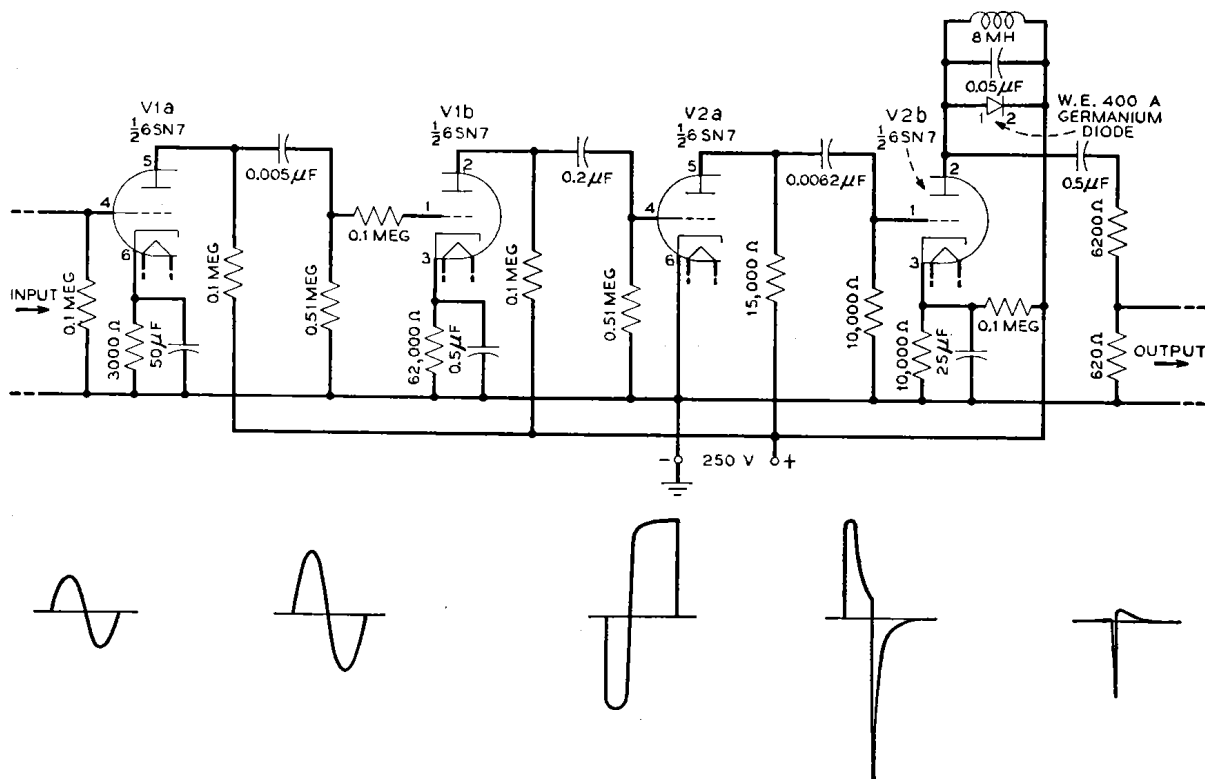
FIG. 5. Schematic of calibrating tone generator.

cycle harmonics. With the 100 cycles from the Laboratories standard frequency oscillator as the drive signal, the frequency calibration of the systems may be checked very readily by comparison of the harmonic spacing on the section with the template scale. The amplitude scale in 6(a) is obtained by inserting a pure tone at the spectrograph in 5 db increments. The frequency scale for spectrograms may also be calibrated as shown in Fig. 6(b). The horizontal lines here are representations of the harmonics of the test tone when the test tone generator is driven by a 500 cycle standard frequency. These lines further afford a means of checking the amount of speed irregularity or wow in the over-all mechanical system. A calibration of the time scale may be obtained by using the test tone generator with 100 cycle drive and making a broad band spectrogram as shown in Fig. 6(c). The spacings between vertical striations in this case correspond to one-hundredth of a second intervals.

In the process of recording some of the word lists, it was arranged to substitute the calibrating test tone circuit for the microphone circuit, and record a few seconds of test tone between the lists of words. When the word lists were analyzed with the spectrograph, the accompanying test tone sections provided a means of checking the over-all frequency response of the recorder and analyzer, and the frequency scale of the sectioner.

The effect of speed variations in either the recorder or the sound spectrograph is to change the frequency scale. A series of measurements with the 100 cycle test tone showed that the tape recorder ran approximately one percent slower when playing back than it did on recording.

The speed variations on the sound spectrograph were measured with the test tone applied directly, and the maximum short time variations were found to be ±0.3 percent. Such direct calibrations of the frequency scale of the spectrograph, during a period of four weeks when most of the spectrographic analysis was done, showed maximum deviations of ±30 cycles at the 31st harmonic of the 100 cycle test tone. During that period a control chart[16] of the measurements of the 3100 cycle component of the test tone showed a downward trend of about 10 cycles, which was attributed to changes in the electronic circuit components of the spectrograph. As a result of these calibration tests, it was concluded that the frequency scale of the sound spectrograph could be relied upon as being accurate within ±1 percent.

Formant Measurements

Measurements of both the frequency and the amplitude of the formants were made for the 20 words recorded by each of the 76 speakers. The frequency position of each formant was obtained by estimating a weighted average of the frequencies of the principal components in the formant. (See reference 4 for a discussion of this procedure.) When the principal components in the formant were symmetrically distributed about a dominant component, such as the second formant of [ʌ] *hud* in Fig. 2, there is little ambiguity

16 "A.S.T.M. manual on presentation of data," Am. Soc. Testing Materials (Philadelphia, 1945), Appendix B.

in choosing the formant frequency. When the distribution is asymmetrical, however, as in the first formant of [ɜ] *heard* in Fig. 2, the difference between estimated formant frequency and that assigned by the ear may be appreciable.

One of the greatest difficulties in estimating formant frequencies was encountered in those cases where the fundamental frequency was high so that the formant was poorly defined. These factors may account for some, but certainly not all, of the differences discussed later

(a) NARROW BAND SECTION
OF 100 CYCLE CALIBRATING TONE

(b) NARROW BAND SPECTROGRAM
OF 500 CYCLE CALIBRATING TONE

(c) WIDE BAND SPECTROGRAM
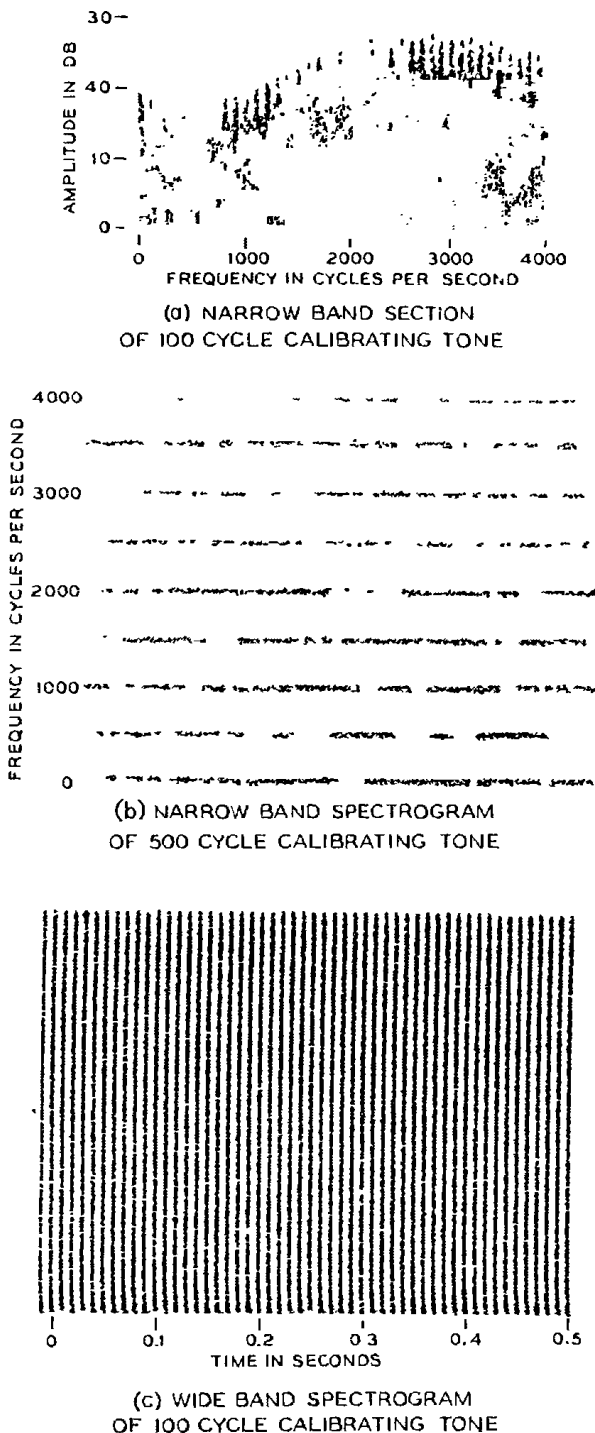OF 100 CYCLE CALIBRATING TONE

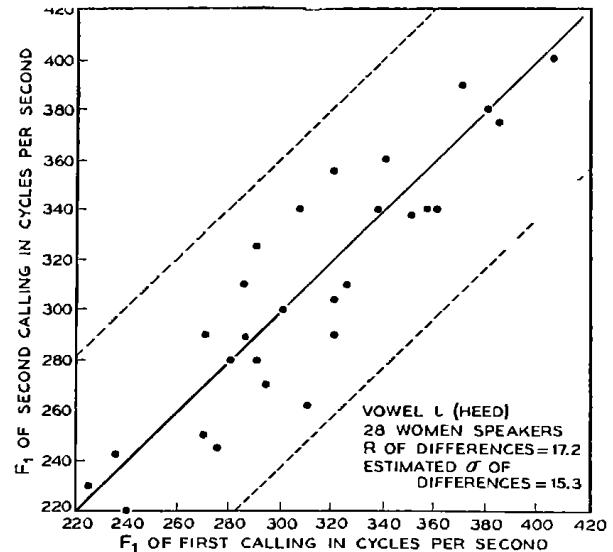FIG. 6. Spectrograms and section of calibrating tone.

FIG. 7. Accuracy-precision chart of first formant frequencies of [i] as spoken by 28 women.

between vowel classification by ear and by measured values of formant frequencies.

Amplitudes were obtained by assigning a value in decibels to the formant peak. In the case of the amplitude measurements it was then necessary to apply a correction for the over-all frequency response of the system.

The procedure of making duplicate recordings and analyses of the ten words for each of the speakers provided the basis for essential checks on the reliability of the data.

One method by which the duplicate measured values were used is illustrated by Fig. 7. This is a plot of the values for the first formant frequency F_1 of [i] as in *heed*, as spoken by the 28 female subjects. Each point represents, for a single speaker, the value of F_1 measured for the *heed* in the first list, *versus* the value of F_1 for the *heed* in the second list. If the F_1 for the second list or calling was greater than that for the first calling, the point lies above a 45-degree line; if it is less, the point lies below the 45-degree line. The average difference $\bar{R}$ between the paired values of F_1 for first and second callings, was 17.2 cycles. The estimated standard deviation σ derived from the differences between pairs of F_1 values was 15.3 cycles. The dotted lines in Fig. 7 are spaced ±3 σ cycles from the 45-degree line through the origin. In case a point falls outside the dotted lines, it is generally because of an erroneous measurement.

Each of the three formant frequencies for each of the 10 vowels was plotted in this way. There were 760 such points for each formant, or a total of 2280 points plotted on 90 accuracy-precision charts like Fig. 7. Of these 2280 points, 118 fell outside the ±3 σ limits. On checking back over the measurements, it was found that 88 of the points were incorrect because of gross measurement errors, typographical errors in transcribing the data, or because the section had been made during the influence period of the consonants instead of in the
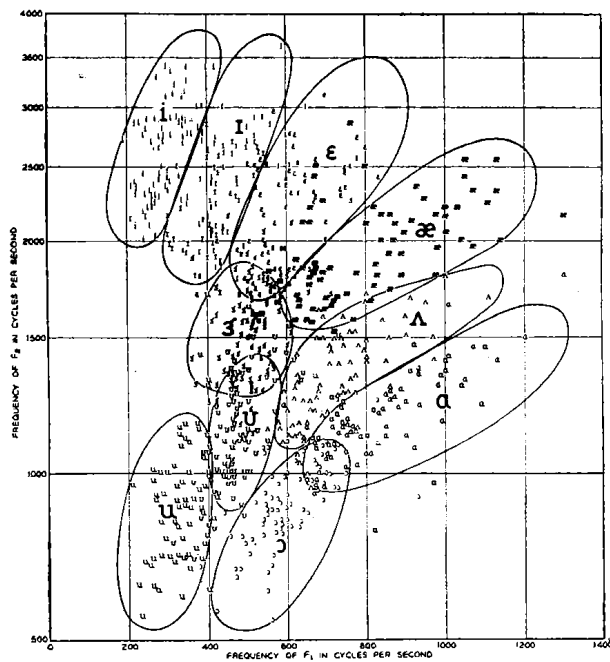
FIG. 8. Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

fore it is assumed that the data are not statistically random, but that there are statistically significant differences between speakers. Since the measurements for pairs of callings were so nearly alike, as contrasted with the measurements on the same sound for different speakers, this indicated that the precision of measurements with the sound spectrograph was sufficient to resolve satisfactorily the differences between the various individuals' pronunciations of the same sounds.

RESULTS OF ACOUSTIC MEASUREMENTS

In Fig. 3, as discussed previously, are plotted areas in the plane of the second formant F_2 *versus* the first formant F_1. These areas enclose points for several repetitions of the sustained vowels by one of the writers. It is clear that here the vowels may be separated readily, simply by plotting F_2 against F_1; that is, on the F_2-F_1 plane, points for each vowel lie in isolated areas, with no overlapping of adjacent areas, even though there exists the variation of the measured values which we have discussed above.

The variation of the measured data for a group of speakers is much larger than the variation encountered in repetitions with the same speaker, however, as may be shown by the data for F_1 and F_2 for the 76 speakers. In Fig. 8 are plotted the points for the second calling by each speaker, with the points identified according to the speaker's word list. The closed loops for each vowel have been drawn arbitrarily to enclose most of the points; the more extreme and isolated points were disregarded so that in general these loops include about 90 percent of the values. The frequency scales on this and Fig. 9 are spaced according to the approximation to an aural scale described by Koenig, which is linear to 1000 cps and logarithmic above.[17]

Considerable overlapping of areas is indicated, particularly between [ɝ] and [ɛ], [ɝ] and [ʊ], [ʊ] and [u], and [ɑ] and [ɔ]. In the case of the [ɝ] sound, it may be easily distinguished from all the others if the third formant frequency is used, as the position of the third formant is very close in frequency to that of the second.

The data of Fig. 8 show that the distribution of points in the F_1-F_2 plane is continuous in going from sound to sound; these distributions doubtless represent

steady state period of the vowel. When corrected, these 88 points were within the ±3 σ limits. Of the remaining 30 points which were still outside the limits, 20 were the result of the individuals' having produced pairs of sounds which were unlike phonetically, as shown by the results of the listening tests.

The duplicate measurements may also be used to show that the difference between successive utterances of the same sound by the same individual is much less significant statistically than the difference between utterances of the same sound by different individuals. An analysis of variance of the data in Fig. 7 shows that the differences between callings of pairs are not significant. However, the value for the variance ratio when comparing speakers is much larger than that corresponding to a 0.1 percent probability. In other words, if the measurements shown in Fig. 7 for all callings by all speakers were assumed to constitute a body of statistically random data, the probability of having a variance ratio as high as that found when comparing speakers would be less than one in a thousand. There-

TABLE I. Classifications of vowels by speakers and by listeners. Vowels as classified by listeners.

		i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɝ
	i	10267	4	6	...	...	3	...	...	...	...
	ɪ	6	9549	694	2	1	1	...	...	...	26
	ɛ	...	257	9014	949	1	3	...	...	2	51
	æ	...	1	300	9919	2	2	...	...	15	39
Vowels intended by speakers	ɑ	...	1	...	19	8936	1013	69	...	228	7
	ɔ	...	...	1	2	590	9534	71	5	62	14
	ʊ	...	...	1	1	16	51	9924	96	171	19
	u	...	...	1	...	2	...	78	10196	...	2
	ʌ	...	1	1	8	540	127	103	...	9476	21
	ɝ	...	...	23	6	2	3	...	...	2	10243

[17] W. Koenig, Bell Labs. Record 27, (August, 1949), pp. 299–301.

TABLE II. Averages of fundamental and formant frequencies and formant amplitudes of vowels by 76 speakers.

		i	I	ε	æ	a	ɔ	U	u	ʌ	ɝ
Fundamental frequencies (cps)	M	136	135	130	127	124	129	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
Formant frequencies (cps)											
F₁	M	270	390	530	660	730	570	440	300	640	490
	W	310	430	610	860	850	590	470	370	760	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
F₂	M	2290	1990	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2480	2330	2050	1220	920	1160	950	1400	1640
	Ch	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F₃	M	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
Formant amplitudes (db)	L₁	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
	L₂	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
	L₃	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

large differences in the way individuals speak the sounds. The values for F_3 and the relative amplitudes of the formants also have correspondingly large variations between individuals. Part of the variations are because of the differences between classes of speakers, that is, men, women and children. In general, the children's formants are highest in frequency, the women's intermediate, and the men's formants are lowest in frequency.

These differences may be observed in the averaged formant frequencies given on Table II. The first formants for the children are seen to be about half an octave higher than those of the men, and the second and third formants are also appreciably higher. The measurements of amplitudes of the formants did not show decided differences between classes of speakers, and so have been averaged all together. The formant amplitudes are all referred to the amplitude of the first formant in [ɔ], when the total phonetic powers of the vowels are corrected so as to be related to each other by the ratios of powers given by Fletcher.[18]

Various methods of correlating the results of the listening tests with the formant measurements have been studied. In terms of the first two formants the nature of the relationship is illustrated in Fig. 9. In this figure measurements for all vowels of both callings are plotted in which all members of the listening group agreed with the speaker. Since the values for the men and the children generally lie at the two ends of the distributions for each vowel, the confusion between vowels is well illustrated by their data; thus the measurements for the women speakers have been omitted.

· The lines on Fig. 9 are the same as the boundaries drawn in Fig. 8. As indicated previously, some vowels received 100 percent agreement much more frequently than others.

The plot has also been simplified by the omission of [ɝ]. The [ɝ] produces extensive overlap in the [u] region in a graph involving only the first two formants. As explained previously, however, the [ɝ] may be isolated from the other vowels readily by means of the third formant.

. When only vowels which received 100 percent recognition are plotted, the scatter and overlap are somewhat reduced over that for all callings. The scatter is greater, however, than might be expected.

If the first and second formant parameters measured from these words well defined their phonetic values; and if the listening tests were an exact means of classifying the words, then the points for each vowel of

---

[18] H. Fletcher, Speech and Hearing (D. Van Nostrand Company, Inc., New York, 1929), p. 74.
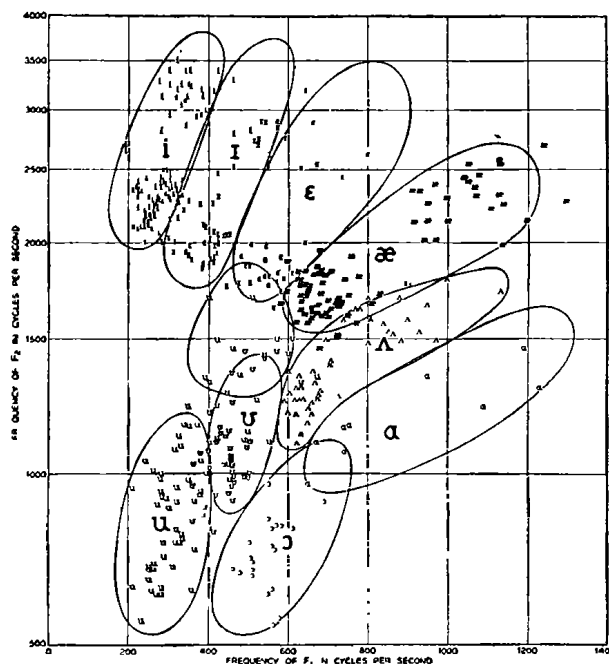

FIG. 9. Frequency of second formant versus frequency of first formant for vowels spoken by men and children, which were classified unanimously by all listeners.

Fig. 9 should be well separated. Words judged intermediate in phonetic position should fall at intermediate positions in such a plot. In other words, the distributions of measured formant values in these plots do not correspond closely to the distributions of phonetic values.

It is the present belief that the complex acoustical patterns represented by the words are not adequately represented by a single section, but require a more complex portrayal. The initial and final influences often shown in the bar movements of the spectrograms are of importance here.[19] The evaluation of these changing bar patterns of normal conversational speech is, of course, a problem of major importance in the study of the fundamental information bearing elements of speech.

A further study of the vowel formants is now nearing completion. This study employs sustained vowels, without influences, obtained and measured under controlled conditions. The general objectives are to determine further the most fundamental means of evaluating the formants, and to obtain the relations among the various formants for each of the vowels as produced by difference speakers. When this information has been obtained it is anticipated that it will serve as a basis for determining methods of evaluating and relating the changing formants within words as produced by various speakers.

SUMMARY

The results of our work to date on the development of methods for making acoustic and aural measurements on vowel sounds may be summarized as follows.

1. Calibration and measurement techniques have been developed with the sound spectrograph which make possible its use in a detailed study of the variations that appear in a broad sample of speech.

2. Repeated utterances, repeated measurements at various stages in the vowel study, and randomization in test procedures have made possible the application of powerful statistical methods in the analysis of the data.

3. The data, when so analyzed, reveal that both the production and the identification of vowel sounds by an individual depend on his previous language experience.

4. It is also found that the production of vowel sounds by an individual is not a random process, i.e., the values of the acoustic measurements of the sounds are not distributed in random order. This is probably true of many other processes involving individuals' subjective responses.

5. Finally, the data show that certain of the vowels are generally better understood than others, possibly because they represent "limit" positions of the articulatory mechanisms.

ACKNOWLEDGMENTS

---

[19] Potter, Kopp, and Green, *Visible Speech* (D. Van Nostrand Company, Inc., New York, 1947).