*Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)*
*http://hdl.handle.net/11245/2.29173*

| | |
|---|---|
| File ID | uvapub:29173 |
| Filename | 130463y.pdf |
| Version | unknown |

SOURCE (OR PART OF THE FOLLOWING SOURCE):

| | |
|---|---|
| Type | article |
| Title | Dimensional analysis of vowel spectra |
| Author(s) | R. Plomp, L.C.W. Pols, J.P. van der Geer |
| Faculty | UvA: Universiteitsbibliotheek |
| Year | 1967 |

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.426094

*Copyright*

*It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).*

*UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)*
*(pagedate: 2014-11-27)*

Received 6 September 1966 9.3, 9.10

# Dimensional Analysis of Vowel Spectra

R. Plomp, L. C. W. Pols, and J. P. van de Geer

*Institute for Perception RVO-TNO, Soesterberg, The Netherlands*

Traditionally, the formant frequencies are regarded as the most important characteristics of the frequency spectra of vowels. It is possible, however, to approach the differences between vowel spectra in a more general way by means of a dimensional analysis. For a particular vowel, the sound-pressure levels in each of a number of frequency passbands can be considered as coordinates of a point in a multidimensional Euclidean space. Different vowel spectra will result in different points. Frequency spectra of 15 Dutch vowels were determined with 18 bandpass filters (10 speakers). The analysis indicated that the "cloud" of 150 points can be described by four independent dimensions that are linear combinations of the original 18. The percentage of total variance "explained" by these dimensions were 37.2%, 31.2%, 9.0%, and 6.7%, respectively. This approach presents interesting perspectives for the development of vowel-discrimination equipment.

## INTRODUCTION

SINCE von Helmholtz, spectrum analysis has been considered an important technique for studying vowels. The occurrence of characteristic peaks or formants in the different vowel spectra has led to extensive research on the properties of these maxima. Traditionally, they have been assigned an important rôle in the psychophysiological mechanism whereby the ear discriminates between different vowels.

We may wonder, however, whether the differences between the frequency spectra of the various vowels are fully described by taking into account only their formant frequencies. In our opinion, there may be an advantage in approaching the problem in a more general way by means of a multivariate analysis of the vowel spectra.

In this paper, an introduction to this approach of the differences between vowel spectra is given. As we shall see, the results involve interesting possibilities for the development of vowel-discrimination equipment and may also be of value for discriminating between consonants.

## I. METHOD

The purpose of dimensional analysis as a technique for studying the differences between vowel spectra can be elucidated in the following way. The vowel is determined by the sound-pressure levels (SPL's) in a number ($N$) of successive frequency passbands. Thus, the vowel can be represented by a point in an $N$-dimensional Euclidean space with these levels as coordinates. Since different sound spectra will be represented by different points, the sound spectra of $n$ vowels result in a "cloud" of $n$ points. Then, the question may be asked, "Do we indeed need $N$ dimensions to describe this cloud of points?" Because $n$ points can be described always by an $(n-1)$-dimensional space, a reduction of the number of dimensions from $N$ to $n-1$ is always possible for $n-1<N$. This reduction, however, is trivial. The proper question is whether the $n$ points can be described by *less than* $n-1$ dimensions. If this is the case, the implication is that there is some simple structure underlying the various vowel spectra—in other words, that the $n$ spectra are governed by a limited number of basic dimensions. If there were only one dimension, this would mean that the $n$ points are ordered along a straight line; in the case of two dimensions, they would be ordered in a plane, etc. Which case actually applies can be investigated only by computation.

In these computations, the concept of *variance*, equal to the square of the standard deviation, plays a basic rôle. In our geometric model, the total variance of the $n$ points is equal to the sum of the squares, divided by $n$, of the distances between the individual points and their "center of gravity." Since the square of a distance in a multidimensional space is equal to the sum of the squares of the projections of this distance along the

different axes (Pythagoras' theorem), the total variance is equal to the sum of the variances for each dimension. In this way, each dimension is seen to account for a certain portion of the total variance. Most often, this portion is expressed as a percentage of the total variance, which percentage then can be said to be "explained" by that particular dimension. It is obvious that the greater this percentage, the more important the dimension is for the description of the total set of data.

Now the question arises, as stated above, whether we can find a new set of coordinates, by rotation of the first set, in such a way that a small number of new dimensions would explain a large part of the total variance.

The computation of these new dimensions is an eigenvalue problem. From the original data, a variance-covariance matrix is calculated; this is an $N \times N$ matrix. The eigenvalues and eigenvectors of this matrix are then determined. The elements of the eigenvector corresponding to the largest eigenvalue are the direction cosines of the dimension that "explains" most of the variance, the magnitude of the eigenvalue being equal to the variance in that direction. Subsequently, the eigenvector corresponding to the second-largest eigenvalue determines the dimension, perpendicular to the first one, that explains most of the residual variance, and so on.

This technique has some resemblance with one used by Kramer and Mathews[1] for the design of a vocoder system with a minimal number of transmission channels.

## II. MEASUREMENTS

Figure 1 represents a block diagram of the apparatus used for measuring the sound spectra of vowels. The equipment in the upper part was used for recording and that in the lower part for analyzing.

Each subject pronounced in a nonreverberant room successively 15 words of the consonant–vowel–consonant (CVC) type (see below), and by means of a condenser microphone, amplifier, and Recorder 1 (Telefunken M5, tape speed 15 ips), these words were recorded on tape. A 6-dB/oct pre-emphasis circuit was included in order to improve the signal-to-noise (S/N) ratio at high frequencies. The tape was played back and each word was recorded again on one track of an endless loop (repetition time 3 sec) on Recorder 2 (Revox G36, modified, tape speed $7\frac{1}{2}$ ips). On the other track, a 1-kHz sine wave was recorded.

The lower part of Fig. 1 represents the analyzing equipment. The endless loop containing one word was played back repeatedly and the 1-kHz signal was used to operate a preset counter in such a way that the same segment of 100 msec of the word was passed by the gate (Claire relays HGS 1060, closing time about 1 msec) for

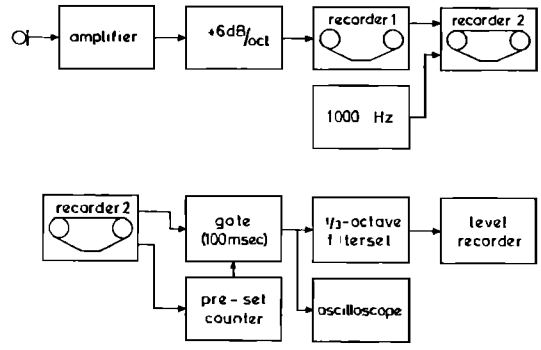


FIG. 1. Block diagram of the apparatus used for recording and analyzing the vowel spectra.

every revolution of the loop. The start moment of the segment was adjusted individually for each loop to its best value for obtaining a 100-msec segment of the vowel. For this, an oscilloscope (Tectronix storage scope 564) was used. The sound was analyzed with a set of $\frac{1}{3}$-oct filters with center frequencies from 100 up to 10 000 Hz (Brüel & Kjær spectrometer 2112). This bandwidth was chosen because it agrees over a large frequency range rather well with the critical bandwidth of the ear's analyzing mechanism.[2] The output signal of the filters was recorded by a level recorder (Brüel & Kjær 2304); the speed of the pen was such that the pen reached its end value within 100 msec.

Calibration of the entire system from microphone to level recorder, without the 6-dB/oct circuit, showed that the frequency-response curve between 80 and 12 000 Hz did not deviate more than 2.5 dB from a flat curve. This calibration was repeated during the measurements. Since the differences between the levels of the various vowels and not the absolute values of the SPL were used in the calculations, we may expect that these deviations were of no influence on the results.

As did Peterson and Barney,[3] we preferred to use words of the type h(vowel)t. Since not all Dutch vowel sounds are covered by using this kind of word, some other words were also used. The list of words is given in Table I. These words were pronounced by 10 young male subjects. They were trained to speak all words equally loudly. In all cases, the duration of the vowel was long enough for taking a 100-msec segment.

The data obtained with the apparatus described above were modified in three different respects. (1) At first, they were corrected for the 6-dB/oct pre-emphasis. (2) After that, the SPL's of the $\frac{1}{3}$-oct bands with center frequencies of 200 and 250 Hz were replaced by one level corresponding to the total intensity in the two bands. The same was done for the bands with center frequencies of 100, 125, and 160 Hz. The main reason

[1] H. P. Kramer and M. V. Mathews, "A Linear Coding for Transmitting a Set of Correlated Signals," IRE Trans. Inform. Theory 2, No. 3, 41–46 (1956).

[2] R. Plomp, "The Ear as a Frequency Analyzer," J. Acoust. Soc. Am. 36, 1628–1636 (1964).

[3] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24, 175–184 (1952).

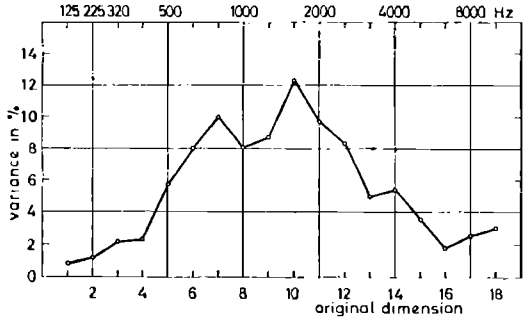


FIG. 2. Percentage of the total variance explained by each of the 18 original dimensions.

for these substitutions was to reduce the influence of differences in voice pitch on the data in the low-frequency range. Moreover, this modification appeared to be attractive in view of our preference to use bandwidths comparable with the ear's critical band (about 90 Hz at low frequencies). (3) Finally, the data were corrected for differences in the over-all sound intensity of the vowels.

In the computations discussed below, the SPL's in decibels *below* the over-all level for each vowel as a function of frequency band were used as basic data.

## III. CALCULATIONS

As discussed in Sec. I, the decibel values measured can be considered as coordinates of points corresponding to the vowels in an 18-dimensional space (18 filters). Since we included 15 vowels and 10 speakers, the total number of points was 150.

A preliminary statistical analysis of these data showed that there were significant differences between the centers of gravity of the vowel points for the different subjects. As we are more interested in differences between vowels than between subjects, the coordinates of the vowel points for each subject were corrected by translation in such a way that the 10 centers of gravity came to coincide.

Figure 2 shows the percentage of the total variance of the 150 points explained by each of the 18 original dimensions. The graph shows that the contributions of the dimensions 1–4 and 16–18 are relatively small; in other words, the frequency bands from 500 up to 5000 Hz are the most important ones for vowel discrimination. The largest contribution of a single band, however, is only 12.3%.

A dimensional analysis of the data was then carried out. The variances explained by the first nine new dimensions are given in Fig. 3. Apparently, the first two dimensions are the most important ones, explaining 37.2% and 31.2% of the total variance, respectively. The third and fourth dimensions are also of some importance (9.0% and 6.7%, respectively). These four dimensions together explain 84.1% of the total variance. There is evidence that the contributions of the fifth and higher dimensions (insofar as they have any statistical significance) are mainly due to differences between the subjects, since a factor analysis based on the vowel

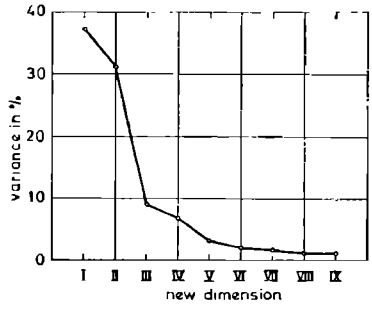


FIG. 3. Percentage of the total variance explained by the first 9 computed dimensions.

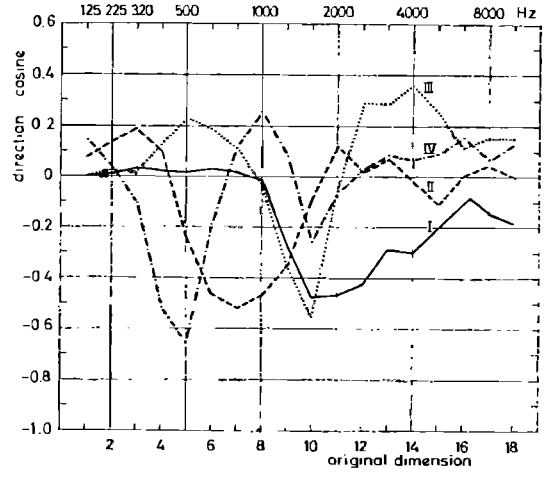


FIG. 4. Cosines of the angles between the computed dimensions I–IV and the original 18.

TABLE I. List of Dutch words used for measuring the sound spectra of the vowels, and approximate equivalents in English. It should be noted that the Dutch vowels are shorter than the corresponding English ones. In a few cases, no English equivalent is available.

| | Dutch word | Vowel pronunciation | Symbol used in this paper |
|---|---|---|---|
| 1 | h*oe*t | f*oo*t | oe |
| 2 | h*aa*t | f*a*st | aa |
| 3 | h*oo*t | n*o*te | oo |
| 4 | h*a*t | h*a*rd | a |
| 5 | h*eu*t | p*eu* (French) | eu |
| 6 | h*ie*t | fr*ee* | ie |
| 7 | h*uu*t | min*u*te (French) | uu |
| 8 | h*ee*t | f*a*ce | ee |
| 9 | h*u*t | h*u*rt | u |
| 10 | h*e*t | h*a*t | e |
| 11 | h*o*t | h*o*t | o |
| 12 | h*i*t | h*i*t | i |
| 13 | d*e* | *a*go | ə |
| 14 | h*ee*r | m*e*hr (German) | ēē |
| 15 | h*oo*r | *O*hr (German) | ōō |

TABLE II. Direction cosines of the computed dimensions I–IV with respect to the original 18.

| | 1<br>125 Hz | 2<br>225 | 3<br>320 | 4<br>400 | 5<br>500 | 6<br>640 | 7<br>800 | 8<br>1000 | 9<br>1250 | 10<br>1600 | 11<br>2000 | 12<br>2500 | 13<br>3200 | 14<br>4000 | 15<br>5000 | 16<br>6400 | 17<br>8000 | 18<br>10 000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 0.001 | 0.012 | 0.030 | 0.024 | 0.005 | 0.032 | 0.151 | −0.013 | −0.278 | −0.475 | −0.468 | −0.422 | −0.288 | −0.301 | −0.195 | −0.078 | −0.148 | −0.189 |
| II | 0.077 | 0.136 | 0.189 | 0.097 | −0.241 | −0.462 | −0.517 | −0.474 | −0.354 | −0.067 | 0.121 | 0.015 | 0.071 | −0.019 | −0.110 | 0.002 | 0.043 | −0.007 |
| III | 0.001 | 0.033 | 0.017 | 0.132 | 0.228 | 0.188 | 0.112 | −0.030 | −0.368 | −0.556 | −0.073 | 0.288 | 0.286 | 0.366 | 0.257 | 0.121 | 0.153 | 0.150 |
| IV | 0.141 | 0.037 | −0.101 | −0.520 | −0.661 | −0.187 | 0.105 | 0.253 | 0.086 | 0.267 | −0.062 | 0.024 | 0.080 | 0.067 | 0.092 | 0.157 | 0.072 | 0.140 |

TABLE III. Coordinate values of the 15 vowels with respect to the center of gravity, averaged over the 10 speakers, on the new dimensions I–IV.

| | 1<br>oe | 2<br>aa | 3<br>oo | 4<br>a | 5<br>eu | 6<br>ie | 7<br>uu | 8<br>ee | 9<br>u | 10<br>e | 11<br>o | 12<br>i | 13<br>e | 14<br>ee | 15<br>oo | Coordinate value of center of gravity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | −34.4 | 18.4 | −22.5 | −1.9 | 7.0 | 2.8 | −2.1 | 13.9 | 10.7 | 24.3 | −21.0 | 13.4 | 7.1 | 14.0 | −29.7 | −83.2 |
| II | −6.9 | 29.2 | 10.7 | 30.4 | −3.4 | −22.7 | −26.6 | −11.3 | −1.1 | 15.3 | 13.3 | −10.1 | −8.1 | −13.3 | 4.9 | −41.4 |
| III | 3.0 | 12.5 | −4.6 | −1.2 | 11.0 | −8.5 | 15.5 | −3.6 | 4.5 | −2.9 | −10.7 | −6.5 | 5.1 | −8.1 | −0.5 | 38.5 |
| IV | −5.8 | −10.2 | 4.6 | −9.0 | 10.3 | −13.9 | −6.9 | 2.8 | 7.6 | 3.8 | 3.2 | 3.2 | 6.8 | −1.7 | 5.1 | 8.9 |

points averaged over the 10 subjects showed that 96.4% of the total variance could be explained by four dimensions. For this reason, we restrict ourselves to the first four dimensions in the remaining calculations.

Table II gives the cosines of the angles between the new dimensions 1–4 and the original 18. These values are reproduced graphically in Fig. 4. By means of these data, the coordinate values of the vowel points along the new dimensions were computed. This calculation, carried out for the average values of the 15 vowels, gave the coordinates represented in Table III. In this Table, the coordinates are given with respect to the center of gravity of all points. The positions of the points are plotted in Fig. 5, in which the graphs (a–c) represent the positions in the I–II, I–III, and I–IV planes, respectively. As the dimensions I and II explain most of the total variance, Fig. 5(a) is the most important graph. It is of interest to notice that the configuration of the points in the I–II plane is similar to that obtained when the frequency of the second formant is plotted against

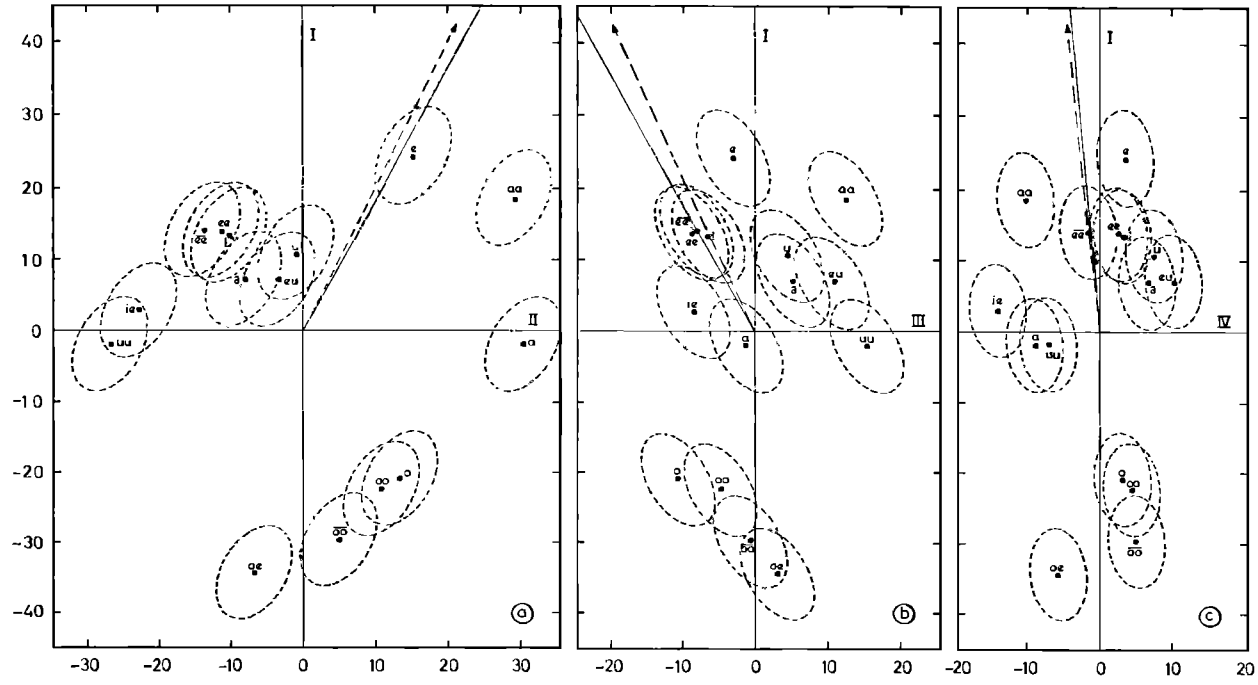


FIG. 5. Positions of the 15 vowels, averaged over the 10 subjects, in the I–II, I–III, and I–IV planes, respectively. The coordinate values are given with respect to the center of gravity of all points. The ellipses indicate the spread of the individual vowel points (see text).

the frequency of the first formant (cf., for instance, Peterson and Barney[3]). Apparently, the first two dimensions are related to the formant frequencies.

It was mentioned above that there were significant differences between the centers of gravity of the vowel points for the different subjects, which was the reason why the analysis was carried out after correcting the data for these differences. Calculation of the 10 centers of gravity in the four new dimensions on the basis of the *un*corrected data resulted in the points plotted in Fig. 6. It appears that the 10 points agree very well with a straight line through the origin of the coordinate system (dashed line). The continuous line represents the best fit in the least-squares sense. The angle between these two lines is only 5.5°.

Plotting the individual vowel points of the 10 subjects, corrected for differences in their centers of gravity, in graphs as in Fig. 5 revealed another interesting property of the variation of the points. It appeared that for all vowels the largest spread of these points occurred in about the same direction. As the number of points for each vowel is only 10, we decided to pool all the points by eliminating the differences between their average vowel values and to study the resulting cloud of 150 points. Variation in this cloud, then, must be looked upon as a sort of residual variation, remaining after the major effects of different subject and vowel have been eliminated. The eigenvectors of this cloud give the directions of the four principal axes of the cloud, the first one explaining most of the variance, the second one explaining most of the residual variance, and so on. The amount of variation along each of these four axes can be represented by the standard deviation in these directions. These values determine an ellipsoid, of which the projections on the I–II, I–III, and I–IV planes are plotted in Fig. 5(a–c), respectively. These ellipses include, on the average, 39% of the projections of the individual vowel points. The direction of the long axis is represented by the continuous line. The dashed line points to the origin of the coordinate system. Again, the angle between the lines is very small.

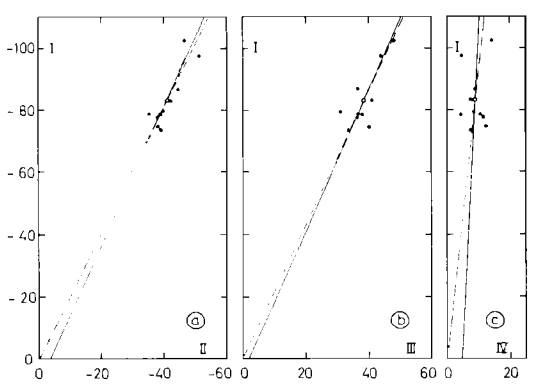


FIG. 6. Positions of the centers of gravity of the vowel points for each subject in the I–II, I–III, and I–IV planes, respectively, calculated from the uncorrected data.

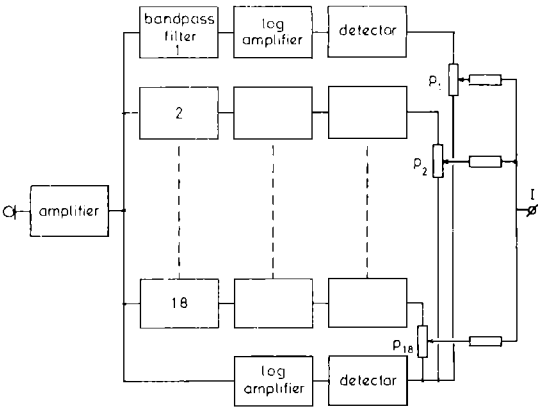


FIG. 7. Block diagram of the apparatus that displays the positions of the vowel points in the I–II plane. A set of potentiometers similar to $p_1, p_2, \cdots, p_8$ was included for dimension II.

There must be a reason why both the variation in the centers of gravity of the vowel points and the variation in the individual vowel points after correction for the former variation are maximal in a direction roughly toward the origin of the coordinate system. To go into more detail in this paper, however, would be premature.

## IV. DISCUSSION

The calculations strongly suggest that the differences between vowel spectra are determined by four independent factors, of which the two most important ones are related to the frequencies of the first and second formant. As it is much easier to derive the coordinate values of vowels along the computed dimensions than to determine the formant frequencies, the approach presented in this paper may have interesting possibilities for the development of vowel-discrimination equipment.

An attempt to investigate whether the vowel points along the new dimensions I–IV, averaged over the 10 speakers, could be used for the identification of the vowels pronounced by each of them individually gave promising results. An individual vowel was considered to be correctly identified when the distance between its vowel point and the average point of the same vowel was shorter than the distance between the point and any other average vowel point. On the assumption that errors between the vowels oo and $\overline{oo}$ and between ee and $\overline{ee}$ may be neglected because these vowels are written in the same manner, and also errors between long and short vowels with similar frequency spectra (o and oo/$\overline{oo}$; i and ee/$\overline{ee}$) because they can be discriminated by their difference in duration, it appeared that about 90% of all individual vowel points were correctly identified on the basis of the criterion mentioned. Omitting Dimension IV, this percentage was 85%; omitting Dimension III also, it was 75%. These calculations are based on the simplifying assumption that the spread of the points is equal in all directions,

whereas Fig. 5 suggests that ellipsoids will give a better result. In future calculations, based on more data, this fact will be taken into account.

Application of these results to the design of speech-recognition equipment is only of interest when the same technique can be used also for the discrimination of consonants. Although in this case particular problems may arise, experiments have been started in which the spectra are measured at short time intervals of running speech, so that spectra representative of the different consonants can be included in calculating a reduced number of new dimensions.

A device, designed for demonstration purposes, which displays the positions of the vowel points in the I–II plane, has been designed. Figure 7 represents a block diagram of this device. The bandpass filters are followed by logarithmic amplifiers and envelope detectors so that output signals proportional to the SPL's in decibels are obtained. Since the new dimensions are linear combinations of these levels, the coordinate value along Dimension I can be obtained by means of the potentiometers $p_1$, $p_2$, $\cdots$, $p_{18}$, each adjusted to the corresponding direction cosine of Dimension I. By connecting all potentiometers to a point at which the potential is proportional to the over-all SPL, variations of the latter level are eliminated automatically. A similar set of potentiometers was used to account for Dimension II. The coordinate values along Dimensions I and II were represented by the vertical and horizontal deflections of the spot on the screen of an oscilloscope, respectively. Pronouncing different vowels results in a position of this spot similar to the points in Fig. 5(a). Such a representation may have some value as a visual feedback system for speech training of the deaf.

## ACKNOWLEDGMENTS

The authors wish to express their thanks to A. M. Mimpen and H. J. M. Steeneken, who participated in the measurements and the building of the apparatus.