Factor analysis of tongue shapes

Richard Harshman,[a] Peter Ladefoged, and Louis Goldstein

Phonetics Laboratory, Linguistics Department, University of California at Los Angeles, Los Angeles, California 90024
(Received 12 October 1976; revised 12 May 1977)

A new analytic procedure PARAFAC has been applied to the description of the shape of the tongue in English vowels. The procedure models the data in terms of a unique set of possibly explanatory factors. It solves in parallel for factors in several data sets, simultaneously describing the differences among data sets in terms of different relative involvement of these common factors. Tracings were made of x-rays taken during the pronunciation of 10 English vowels by five speakers. The positions of the tongue in these 50 vowels were quantized in terms of 13 superimposed grid lines. PARAFAC analysis shows that the data can be described in terms of two factors. One factor generates a forward movement of the root of the tongue accompanied by an upward movement of the front of the tongue. The second factor generates an upward and backward movement of the tongue. Movements from front to back vowels involve decreasing amounts of factor one. Movements from high to low vowels involve decreasing amounts of factor two. Different speakers use the two factors to different degrees which may be associated with their individual anatomy. The correlation between the observed data and that predicted by the model is greater than 0.96.

PACS numbers: 43.70.Bk, 43.70.Gr

INTRODUCTION

A. The nature of the problem

During speech, the surface of the tongue assumes hundreds of subtly different shapes. Each of these shapes is a particular nonuniform curve, usually one having no simple representation in terms of words or equations. Nonetheless it becomes clear by simply examining the curves, that they are related to one another in some kind of orderly way. For example, there appear to be systematic patterns of relationships among the tongue shapes used by a given speaker for producing different vowels, as well as patterns relating the shapes used by different speakers to produce the same vowel. Such patterns of tongue shape variation are of great interest to speech scientists, but, for methodological reasons, attempts to measure and accurately describe these patterns have run into serious difficulties. The basic problem has been one of finding a way to represent the curved surface of the tongue so that the interesting relationships among different tongue shapes can be easily identified and measured. This is the problem to which this article is primarily addressed.

1. Pictorial representations

One of the simplest yet most complete methods of representing a tongue shape is to use a picture of that shape. Such pictures, particularly tracings from x-ray photographs, form an essential starting point for the investigation of tongue-shape variations. But while the information in pictorial representations is relatively complete, it is not sufficiently compact: tracings do not separate the essential information from the nonessential. For this reason, comparison of two or more tongue shapes by means of pictures does not provide a precise analysis of the relationship among the several shapes, i.e., it does not reveal the rule relating the shapes to one another. What is needed is a compact, hopefully quantitative, representation that still can capture the subtle differences among closely related shapes of the tongue.

2. Analysis into features

One potentially effective way of studying the relationships among different tongue shapes would be to reduce each tongue shape to a description in terms of a few common underlying features. Descriptions of this kind would also be useful for studying relationships between tongue shapes and other variables such as formant frequencies, and for studying linguistic questions, such as historical patterns of language change. Once a set of common features for all the tongue shapes has been defined, and each particular shape has been described in terms of its values on these common features, it then becomes possible to compare any two or more shapes by simply comparing their respective feature values.

The next question thus becomes "By what method of analysis can one find the most appropriate set of features for describing tongue shapes and their relationships to one another?" The availability of a good analysis method, appropriate to the data at hand, can make a big difference in the success or failure of the search for an optimal feature set. A case in point is the search for features to describe the acoustic properties of vowels, a search which has been much more successful than the one for features to describe articulatory properties of vowels. Much of the progress with acoustic features was made possible by the use of spectral-analysis techniques. Spectral analysis reduced the complex sounds of different vowels into a common set of underlying spectral components and allowed vowels to be compared in terms of their respective values on these common components. This led, in turn, to the identification of a small number of features—the formant frequencies—which appear to capture most of the information relevant to acoustic relations among vowels.

Since at least in part the ear analyzes sounds into spectral components, this may help account for the success of spectral analysis in the description of acoustic properties. However, there is no similar *a priori* reason to lead us to expect that articulatory properties would be well suited for spectral analysis (but see Ref.

9 discussed below). An alternative analysis technique better suited to such data has not been demonstrated. Lacking an appropriate analytic tool, the search for articulatory features for vowels has proceeded on a more *ad hoc* and intuitive basis, but (as will be shown) with only limited success.

This article presents preliminary results of an attempt to adapt new analytic tools to the problem of describing and comparing tongue shapes. It seeks to develop a systematic alternative to previous *ad hoc* methods of generating articulatory features. The approach consists of (a) a general and uniform method of measuring vocal tract x-rays, and (b) a procedure for analyzing the resulting x-ray measurements into a few underlying components by means of the statistical techniques of factor and principal-component analysis. In particular, it seeks to demonstrate how PARAFAC, a recently developed method of three-way factor and component analysis[1-3] can be used to help identify a compact and descriptively adequate set of articulatory features for the tongue shapes of 10 English vowels, as produced by five speakers.

B. Previous features for tongue shapes

Although the discovery of psychologically real features may be an ultimate goal, one must first grapple with the question of descriptive adequacy. The traditional method of describing tongue shapes in vowels is in terms of the highest point of the tongue. The tongue shape is thought of as being completely specified by two numbers, one representing the position of the highest point of the tongue in the horizontal dimension, and the other representing it in the vertical dimension. No one has ever stated an explicit process that would enable one to draw the shape of the whole tongue, given only the position of the highest point. But Jones[4] and many other authors have diagrams which claim to show the relation between the high point of the tongue and the shape of the rest of the tongue in a set of cardinal vowels. It has long been known, however, that these diagrams contain only a skeleton of the truth.

Stevens and House[5] proposed a system in which the width of the vocal tract at its narrowest point is specified, and the distance of this narrowest point from the glottis is also given. From these two numbers the position of the rest of the tongue can be generated by means of an explicit algorithm. A similar scheme was proposed by Fant[6] using cross-sectional areas of the vocal tract, rather than tongue positions.

A different two-parameter model of the body of the tongue was proposed by Flanagan *et al.*[7] These authors specified tongue shapes of vowels in terms of the coordinates of the center of a circle, an arc of which represented the position of the body of the tongue, with a straight line continuation of the arc representing the position of the tongue in the pharynx. They also graphically compared the shapes that could be produced by their model with data derived from x-ray tracings for three vowels, but they did not quantify the goodness of fit.

Another two-number method of characterizing shapes of the tongue has been suggested by Lindblom and Sundberg.[8] Their technique is to use one number to specify the degree of distortion of the tongue shape from a neutral shape, and another to specify the position of the point of maximum displacement.

Finally, Liljenkrants[9] has shown how a curve representing the shape of the tongue can be described in terms of its Fourier components. Using a partly polar and partly Cartesian coordinate system which enabled him to straighten out the vocal tract, he showed that the profile of the tongue could be fairly well described in terms of the magnitude and phase of a fundamental frequency. He shows quite conclusively that the tongue shapes in vowels can be described in terms of two or three numbers.

Other experimenters[10-12] have constructed more elaborate models of the tongue, which will not be reviewed here. These researchers are all more concerned with modeling the physiological forces moving the tongue. They are not trying to find a minimal set of descriptively adequate features.

C. Comparison of the current and previous approaches

There are certain limitations in the previous attempts to specify features for tongue shape. There is, first of all, an arbitrariness to the system employed. It is the intuition of the researcher that determines whether geometric figures, narrowest points, or specified distortions will be used to describe the shape of the tongue. Secondly, with the exception of the spectral-analysis model, there is no explicit measurement of the fit of the model to the data. Thirdly, (again with the exception of the spectral-analysis model), there is no explicit optimization of the fit of the model under variation of parameters within the model. No mathematical guarantee is present that the features chosen insure optimal fit, subject to particular explicit constraints. Fourthly, patterns of individual differences among speakers are not integrated into the model. Finally some of the models (e.g., the narrowest-point models and the spectral-analysis model) do not lend themselves to interpretation as possible descriptions of psychologically real feature systems.

The factor-analytic approach to be adopted in this study is not subject to these particular limitations. It presents an integrated mathematical model which simultaneously describes (a) the features of tongue shape for English vowels; (b) the relationships among English vowels in terms of these features; and (c) person differences in vowel production (for our five speakers) in terms of these features. A very general specification system is adopted which is transformed into a more specific set of feature shapes by optimization of fit of the model to the data, rather than by intuitive selection of shapes by the researcher. In this sense, the procedure for developing the model might be considered more objective or less arbitrary, i.e., more theoretically neutral and strictly determined by the characteristics of the data, than previously used procedures.

In this study, feature discovery starts with sets of

x-rays of vocal-tract shapes for five people. The x-rays are traced and the tracings of the tongue are characterized in numerical terms by means of an algorithm that is as objective as possible. (We are well aware, however, that experimenter judgment of what is relevant is often involved both in tracing x-rays and in measuring the tracings.) Then certain minimal assumptions are made about the characteristics of possible features of tongue shape. The model assumes that at least our five speakers (and presumably all speakers of English) use some of a small number of these features, that these features add together to produce the observed vocal tract shapes, and that the elementary features can be described in terms of patterns of displacement of the tongue from a reference position of the vocal tract. Given the data base and assumptions just described, an algorithm proceeds to select from the set of possible features the best specific features for vocal tract shapes, along with feature values for vowels and speakers, by using a three-way factor analysis or three-way principal-component analysis procedure as described below. The best vocal-tract features and vowel/person feature values are the ones that do the best job of reproducing the actual x-ray measurements across all vowels and speakers. The fit of the feature-model predictions to the observed data is optimized by making the sum of the squared errors of these reproductions as small as possible.

The intuition or educated judgment of the researchers still enters in, of course. It enters, for example, when the obtained solutions of the model must be interpreted, and in deciding whether some artifacts of measurement or other errors have distorted the solution. Furthermore, the model simply determines the best solution for each given number of factors. The researcher has to decide whether or not the solution is satisfactory, or whether more or fewer factors are required. Nonetheless, the features are determined in a relatively standarized and clearly specifiable fashion. The restricted points at which the researcher's judgment operates can be described, as can the evidence supporting each particular choice, and the choices are generally explicit enough to enable them to be independently evaluated by other interested persons who have access only to the reports on the research.

I. METHOD

A. Construction of the data base

The data that were analyzed have been described elsewhere.[13] Briefly, from a larger body of data, we selected high-quality audio recordings and cine-fluorograms of five speakers saying sentences of the form "Say *h*(vowel)*d* again." The vowels included /i, ɪ, eɪ, ɛ, æ, ɑ, ɔ, o, ʊ, u/ as in "heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd." These vowels cover the range of articulatory movement that occur in English vowels. The vowel /ər/ as in "bird" was not included, because it is known[14] that this vowel can be produced in more than one way.

A single frame in the film was chosen for each of the 10 different vowels as said by each of the five speakers. In the case of the vowels /ɪ, ɛ, æ, ɑ, ɔ/ the most steady-state part of the second formant was selected. Whenever possible this procedure was also followed for the other vowels. But for some speakers at least one of these vowels was diphthongal throughout. In these cases a point approximately 30 msec after the first consonant was chosen. In each case, the corresponding frame in the film was located and traced.

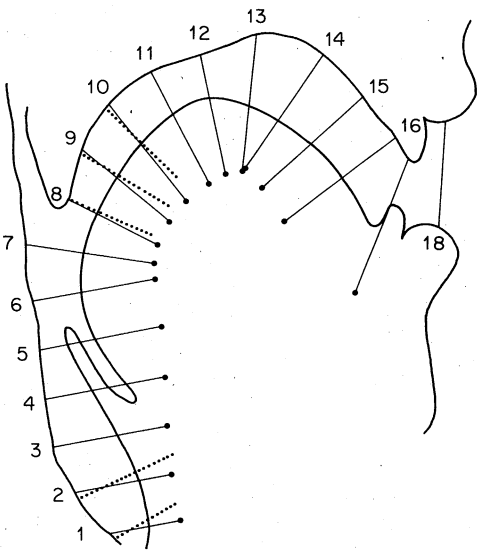

FIG. 1. The reference lines used for characterizing the positions of the tongue for one of the speakers. The vowel is /æ/ as in "had." For reasons given in the text the solid lines were used, rather than the dashed lines that are more nearly perpendicular to the middle of the vocal tract.

1. *Reference grids*

The vocal-tract shapes were characterized in terms of a set of reference lines of the kind shown in Fig. 1. Because head shapes differ, a different reference grid had to be used for each speaker. The procedure used to generate the reference grid for each speaker was as follows: First we estimated the position of the midline of the vocal tract when in position for the vowel /æ/ as in "had." We then measured the length of this line from the estimated position of the glottis to a point midway between the two lips. This length was divided into 18 sections, subject to the following conditions: (a) the middle of section 3 should be below the level of the epiglottis; (b) the middle of section 4 should be above the root of the epiglottis; (c) sections 1–3 should be the same length; (d) sections 4–18 should be the same length; (e) the lengths of sections 1–3 should be as similar to the lengths of sections 4–18 as the constraints allow. If possible within these constraints (as it was for three of our five speakers), *all* the sections were made the same length. For the other two speakers the first three sections were slightly longer. It was thought preferable to anchor the measurements around the epiglottis in this way, since this is the only fixed landmark on the tongue. It was hoped that this procedure would increase within- and cross-speaker comparability. An additional consideration was the fact that data on the sections below

the root of the epiglottis are less reliable because of problems in seeing the position of the tongue on x-rays in this region. We did not want to have four unreliable measurements in the case of some subjects and three for others.

When the midline of the vocal tract had been determined and divided into 18 sections as described above, reference lines were drawn through the center of each section. These lines were as nearly as possible perpendicular to the midline of the vocal tract, subject to two additional constraints: (f) in the pharyngeal region, they were all parallel to each other, with the majority of them being as perpendicular to the back wall of the pharynx as possible; and (g) when the slope of the reference lines changed so as to remain perpendicular to the curving midline, the changes of slope were always as uniform as possible. Thus, for the speaker represented in Fig. 1, the solid lines were used instead of the dotted lines.

In the case of the reference lines that crossed the hard palate, the alveolar ridge, and the teeth (sections 13–17 on Fig. 1) which are fixed structures on the upper surface of the vocal tract, an arbitrary end point was marked on each line at a fixed distance (45 mm on the x-ray tracings) from the upper surface of the vocal tract. End points were then marked on each of the other lines so that the set of end points followed a smooth curve within the body of the tongue (as shown by the series of solid points at the ends of the reference lines in Fig. 1). The precise shape of this curve is irrelevant since we will be considering only deviations of the tongue from its average position for each subject. This will subtract out any constant effects introduced by choice of the shape of this curve. (See Sec. I B.)

2. Measurements

Once a grid had been established for a given speaker, all his tongue shapes were characterized by measuring the distance along the reference lines between the surface of the tongue and each of these 17 points. The distance between the lips was also measured. All the original measurements were in millimeters to the nearest 0.5 mm. They were subsequently converted to centimeter life size by dividing by 16.

The data set was verified by means of a computer program which simplified the comparison of the measurements with the original x-ray tracings. Each measurement, except for the distance between the lips (point 18), was subtracted from the constant representing the distance from the upper surface of the tract to the arbitrary end point. The resulting subtracted values would be equivalent to cross sections (sagittal dimensions) of the vocal tract *if* the tongue were moving in a vocal tract with a smooth, fixed upper surface. (In practice, of course, there is considerable movement of the velum and of the back wall of the pharynx, which we are not concerned with in this study of tongue movements). Using the pseudosagittal dimensions, the computer program diagrammed the vowels of different subjects with reference to a common base line representing the upper surface of an arbitrary vocal tract. An example is given in Fig. 2, which shows a superimposition of the measurements for each speaker saying the vowel [o] as in "hoed."

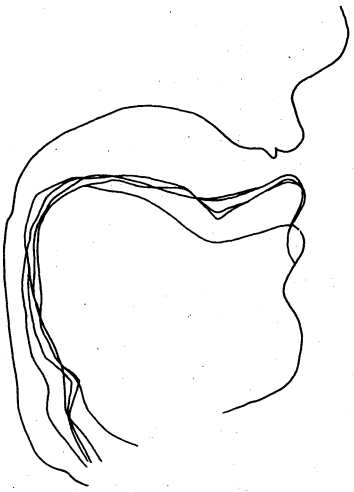

FIG. 2. The tongue positions for the five speakers saying the vowel /o/ as in "hoed" superimposed on one another within a standardized vocal tract.

Our analysis of the data using these techniques showed that there is a great deal of apparently random variation, particularly in the lower part of the pharynx (points 1–3). This may be due to individual anatomical variation, but, as we have already mentioned, our x-ray tracings may not be very reliable in this area. Comparison of the diagrams with the original x-ray tracings also showed that point 17 (just inside the lips) was often beyond the tip of the tongue, and simply indicative of the position of the teeth. The measurement for point 18 (the distance between the lips) is also obviously not relevant to characterizations of tongue shape. Accordingly we selected for analysis the data corresponding to points 4–16.

The pseudosagittal dimensions (i.e., the original measurements with the constants subtracted) for each of the 13 points, 10 vowels, and 5 speakers are shown in Table I. The measurements are presented in this way for the convenience of other research workers who wish to diagram the positions of the tongue. But it should be remembered that, since the upper surface is in constant motion, Table I does not give the true sagittal dimensions of the vocal tract.

B. Factor analysis models and procedures

1. The conceptual model

The original pseudosagittal measurements may be turned into measurements of tongue displacements, by defining an arbitrary reference position for each speaker, and considering each observed x-ray shape as a displacement of the tongue surface from that speaker's reference position. Analyzing these displacements, we seek to uncover a few underlying displacement patterns that combine in various amounts to produce particular observed shapes. These displacement patterns or gestures of the tongue act together to carry the tongue from

TABLE I. Measurements in cm of the pseudosagittal dimensions (see text) corresponding to tongue points 4–16 for five speakers each saying 10 vowels.

Speaker	Vowel	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	2.45	2.40	2.40	2.50	2.45	2.05	1.65	1.00	0.45	0.40	0.55	0.55	0.95
1	2	2.55	2.05	1.95	1.90	1.80	1.60	1.30	0.95	0.55	0.65	0.90	0.90	1.05
1	3	2.35	1.90	1.80	1.80	1.80	1.55	1.30	0.95	0.65	0.70	0.90	0.85	1.05
1	4	2.50	2.05	1.65	1.65	1.55	1.45	1.25	1.05	0.70	1.05	1.20	1.15	1.10
1	5	2.05	1.50	1.35	1.50	1.55	1.45	1.30	1.15	0.95	1.40	1.55	1.40	1.25
1	6	1.55	1.00	0.85	1.00	1.25	1.35	1.50	2.00	2.10	2.55	2.65	2.35	2.00
1	7	1.65	0.90	0.65	0.65	0.75	0.95	1.40	1.90	2.15	2.60	2.70	2.60	2.40
1	8	1.90	1.30	0.95	0.85	0.75	0.75	0.95	1.30	1.65	2.15	2.30	2.25	2.30
1	9	2.40	1.60	1.45	1.25	1.00	0.95	0.80	0.85	1.10	1.50	2.10	2.00	1.65
1	10	2.70	1.95	1.50	1.30	0.90	0.70	0.55	0.55	0.95	1.45	1.80	1.90	2.00
2	1	2.95	2.70	2.75	2.75	2.70	2.60	2.25	1.00	0.35	0.15	0.30	0.60	1.15
2	2	2.40	2.20	2.25	2.20	2.25	2.15	1.85	1.25	0.75	0.75	0.90	1.05	1.10
2	3	2.25	2.45	2.65	2.65	2.40	2.20	2.05	1.55	0.95	0.85	1.10	1.40	1.65
2	4	2.00	1.75	1.90	2.30	2.40	2.20	2.00	1.45	1.00	1.05	1.40	1.75	1.80
2	5	1.25	1.15	1.30	1.65	1.95	1.90	1.80	1.65	1.40	1.70	2.15	2.45	2.60
2	6	0.45	0.25	0.30	0.40	1.15	1.70	1.95	2.30	2.60	2.95	3.30	3.15	2.60
2	7	0.40	0.20	0.20	0.30	0.60	1.05	1.35	1.65	2.60	3.05	3.45	3.60	3.40
2	8	1.00	0.55	0.55	0.45	0.65	0.80	1.15	1.55	2.25	2.75	3.20	3.35	3.25
2	9	1.30	0.70	0.65	0.45	0.65	0.90	1.20	1.45	1.90	2.40	2.85	2.80	2.45
2	10	2.15	1.80	1.50	1.05	0.65	0.55	0.65	0.80	0.95	1.55	2.10	2.35	2.60
3	1	2.10	2.00	2.15	2.05	1.95	1.80	1.45	1.10	0.75	0.65	0.75	0.80	0.90
3	2	2.00	1.70	1.90	1.95	1.90	1.75	1.35	1.15	0.95	1.00	1.10	0.90	0.65
3	3	1.95	1.80	1.80	1.95	1.95	1.95	1.65	1.25	0.90	0.85	1.05	0.95	0.90
3	4	1.55	1.40	1.50	1.70	1.85	1.80	1.90	1.80	1.75	1.70	1.70	1.40	1.10
3	5	1.65	1.25	1.40	1.70	1.90	1.95	2.05	2.10	1.95	1.95	2.15	2.10	1.70
3	6	0.95	0.55	0.70	1.15	1.65	2.20	2.65	2.95	3.05	3.20	3.35	2.95	1.90
3	7	1.20	0.65	0.45	0.65	0.75	1.00	1.45	2.10	2.40	2.65	2.80	2.55	1.95
3	8	1.55	1.45	1.05	1.15	1.05	1.00	1.15	1.45	1.90	2.40	2.70	2.65	1.85
3	9	1.80	1.05	1.05	1.05	1.00	1.00	1.15	1.40	1.65	1.95	2.15	1.85	1.50
3	10	2.00	1.70	1.40	1.20	1.00	0.85	0.95	1.00	1.10	1.55	1.80	1.70	1.25
4	1	2.70	2.60	2.55	2.50	2.45	2.40	1.80	1.35	0.70	0.55	0.75	0.85	1.85
4	2	2.25	1.90	1.85	1.90	2.15	2.05	1.85	1.65	1.35	1.40	1.50	1.90	1.80
4	3	2.25	2.20	2.30	2.25	2.30	2.20	1.70	1.45	0.90	0.90	1.10	1.25	1.85
4	4	1.90	1.50	1.40	1.40	1.65	1.75	1.75	1.85	1.60	1.80	1.90	1.65	1.50
4	5	1.70	1.20	1.05	1.05	1.55	1.70	1.80	1.90	1.85	2.10	2.35	2.40	2.25
4	6	1.05	0.90	0.45	0.60	1.45	2.05	2.90	2.90	3.00	3.20	3.35	2.95	2.15
4	7	0.90	0.40	0.45	0.55	1.30	1.80	2.30	2.80	3.10	3.40	3.45	3.00	2.40
4	8	2.00	1.30	1.05	0.90	0.95	0.90	1.25	1.65	1.80	2.30	2.60	2.60	1.90
4	9	2.15	1.70	1.45	1.30	1.30	1.25	1.20	1.35	1.45	1.95	2.20	2.25	1.95
4	10	2.95	2.30	2.05	1.80	1.70	1.45	1.00	0.80	0.80	1.15	1.55	1.90	1.40
5	1	3.00	2.45	2.30	2.20	2.10	1.45	1.15	0.80	0.40	0.60	0.45	0.40	0.85
5	2	2.40	2.10	1.95	1.90	1.80	1.45	1.10	0.90	0.70	0.95	0.95	0.75	1.10
5	3	2.50	2.40	2.20	2.05	2.05	1.70	1.30	0.95	0.65	0.95	1.00	0.85	1.20
5	4	2.25	2.10	1.95	1.90	1.90	1.55	1.15	1.00	0.90	1.10	1.05	0.90	1.25
5	5	1.70	1.95	2.05	2.10	1.95	1.50	1.15	1.15	1.10	1.30	1.30	1.20	1.45
5	6	1.40	0.85	1.05	1.30	1.55	1.55	1.65	2.00	2.40	2.75	2.80	2.60	2.35
5	7	1.10	0.70	0.70	0.90	1.15	1.00	1.20	1.80	2.40	2.75	2.80	2.35	2.05
5	8	1.80	1.05	0.75	0.70	0.70	0.55	0.60	1.20	1.85	2.40	2.45	2.25	2.40
5	9	1.90	1.25	1.05	0.90	0.95	0.65	0.65	1.25	1.85	2.35	2.35	2.05	2.30
5	10	2.70	2.05	1.65	1.40	1.15	0.60	0.40	0.50	0.60	1.15	1.40	1.60	1.65

its reference position to a given observed position. Each gesture corresponds to a single factor that will be uncovered by the analysis procedure. In the models to be considered here, the basic set of gestures is the same for all vowels, and for all speakers, but differences between vowels and speakers are described in terms of differences in the relative *amounts* of the various gestures employed. The common set of basic tongue gestures is the set of articulatory features discovered by the analysis. The amount of each gesture used to generate the observed displacements for a particular vowel are the feature values of that vowel. In the PARAFAC three-way analysis, different speakers are also distinguished by systematically different emphasis on each of the various basic tongue gestures.

Let x_{ijk} represent the measurement along the ith grid line (ith-tongue coordinate line) of the jth vowel, as measured for the kth speaker (e.g., in Table I, $x_{411} = 2.45$, $x_{16105} = 1.65$, and $x_{511} = 2.40$). We next define the reference measurement for the ith grid line of speaker k as follows:

$$r_{ik} = \frac{1}{10} \sum_{j=1}^{10} x_{ijk} . \quad (1)$$

The displacement of the tongue corresponding to any observed x-ray shape is computed by simply taking the difference between the measurements for that shape and the reference-position measurements, as follows:

$$d_{ijk} = (x_{ijk} - r_{ik}). \quad (2)$$

Table I contains 5 sets of 10 vectors, each vector with 13 components. The problem of decomposing the deviation versions of these 50 vectors into different combinations of a few common underlying vectors is a classical problem for principal-component and factor analysis. While the general properties of these methods are well known,[15,16] it will be useful to clarify their relation to the particular conceptual model described above. This will also facilitate discussion of a newer and less well known three-way generalization of these methods (see Refs. 1 and 2) that will be used as the primary analytic technique of this article.

2. The two-way mathematical model

The classical two-way factor analytic techniques can not handle triply subscripted (three-way) data arrays. Therefore we will initially consider only one of the five speakers (perhaps an average speaker). A given data point for this speaker can be represented by x_{ij}, and a displacement score by d_{ij}. The classical factor and principal-component models would represent the underlying component of this displacement for w factors as follows:

$$d_{ij} = t_{i1}v_{j1} + t_{i2}v_{j2} + \cdots + t_{iw}v_{jw} + e_{ij}. \quad (3)$$

In this equation the t and v values are weights (classically called loadings and scores, respectively) which describe the relation between each underlying gesture and the surface variables of the data set. The t_{i1} and t_{i2} coefficients, for example, describe the relative effect of factors 1 and 2 on displacements along tongue grid line i. The v_{j1} and v_{j2} values, on the other hand, describe the relative contributions of factor (gesture) 1 and 2 to the jth vowel. The e_{ij} term is an error term that represents the difference between the actual measured displacement d_{ij} and the value predicted by the sum of the contributions of the factors. The discrepancy may be thought of as being due to errors of measurement, or to specific variance for that tongue section i not shared by any other tongue section, or some combination of these. Technically, principal-component analysis and factor analysis make different assumptions about the statistical properties of the e_{ij} values (e.g., whether they are uncorrelated), but for the purposes of the present discussion we will ignore these distincions. We will concentrate on the principal-component case where the e values are the least-square residuals of prediction of the displacement values from the factor loadings and scores (the t and v coefficients). Also, for simplicity, we will use the term "factor" to refer either to factors or principal components.

As an example, consider a basic tongue gesture which involves a large displacement of the tongue surface along grid line i, so that the t_i coefficient will be large. Further, assume that this gesture is a major component of the observed surface displacement for vowel j, so that the v_{ij} coefficient is also large. The value of d_{ij} predicted by the model is the sum of the effects of each of the factors (gestures), i.e., the first w terms on the right-hand side of Eq. (3). So, in the example we are considering, the product of the t and v coefficients for this component of the surface displacement would be large, and contribute substantially toward the sum of the underlying components which estimates the final observed displacement, d_{ij}. On the other hand, if either the effect of the gesture along tongue grid line i were small, or the amount of that gesture involved in the production of vowel j were small, the contribution of that component to this particular d_{ij} would be small. The error term e_{ij} is not used to predict tongue displacements, but rather to reconcile such predictions with the actual observed displacements. It expresses the fact that a component for the ith tongue section peculiar to the jth vowel will have to be added.

A traditional factor analysis of a single set of tongue-shape deviation measurements produces two tables. One table is a set of values describing the tongue gestures or articulatory features of tongue shape. This is the table of t values, and for a solution in which two factors are extracted (i.e., $w=2$), it would have two columns (one for each basic gesture), and as many rows as there were tongue grid lines entered into the analysis (in this case 13). The entries in this table describe, for each gesture or articulatory feature, the relative size of tongue displacement occurring along each of the grid lines. Thus, a column of this table describes the shape of a given displacement gesture. The second table, the table of v values, gives the amount that each gesture contributes to the observed displacement for each vowel. The entries in this table can be interpreted as the values of the extracted features for each vowel.

The solution obtained by a two-way factor analysis has a number of desirable characteristics: it gives a description of underlying tongue gestures—our articulatory features—and also provides a description of vowel relationships in terms of these features. However, it suffers from at least two shortcomings: (a) it does not provide a way of incorporating individual speaker differences into the model; (b) it does not provide a unique solution. The first shortcoming is evident from the two-way nature of the required data, but the second shortcoming requires some further explanation.

As a simple example, consider the equation $x+y=20$. This equation has no unique solution, since an infinite number of x, y pairs will satisfy the weak constraints imposed by the specified relationship. Similarly, the values of the factor loadings in the two-way factor model are not sufficiently constrained by the model to uniquely determined the description of the underlying factors. Given one set of t and v values which provide an optimum fit to a particular set of data, we can always find other sets of t and v values (call one such set $\overset{*}{t}$ and $\overset{*}{v}$, for example) which provide exactly the same fit of the model to the data. For example, if we defined our new factor loadings as $\overset{*}{t}_{i1} = t_{i1} + t_{i2}$, $\overset{*}{t}_{i2} = t_{i1} - t_{i2}$, we would obtain a new version of factor 1 which would incorporate

those aspects common to old factors 1 and 2, and a new version of factor 2 which would incorporate those aspects which differentiated old factors 1 and 2. Given the $\overset{*}{t}$ values, it is a straightforward matter to determine the new $\overset{*}{v}_{j1}$ and $\overset{*}{v}_{j2}$ values so that $\overset{*}{t}_{i1}\overset{*}{v}_{j1}+\overset{*}{t}_{i2}\overset{*}{v}_{j2}$ does just as good a job of reproducing the data as the old t and v values.

Selecting the preferred solution in the face of this nonuniqueness is the classical "rotation" problem of two-way factor analysis (called "rotation" because a change of coordinates, such as the one described above, can be considered a rotation of axes in the factor space). Psychologists and others who have used two-way factor analysis have devised special procedures for rotation of factors into preferred form. Foremost among these is the well known principle of "simple structure," which (loosely speaking) selects the particular solution in which as many as possible of the t values are near zero. (Kaiser's Varimax roughly approximates this objective). While this simplifies the mathematical description of any factor, there is no reason, *with our tongue data*, to presume that this procedure will lead to patterns of factor loadings which are empirically more basic or "real" (e.g., which might correspond to the empirical sources of variation which underlie these data—see Refs. 1 and 2 for further discussion of "explanatory" validity).

Because of these shortcomings of two-way factor analysis and principal-component analysis, a three-way analytic procedure called PARAFAC was adopted as the primary method of analyzing this tongue data. Not only does this procedure allow explicit incorporation of subject differences, but, if the data are adequate, it can provide a unique solution for the underlying factors.

3. The three-way (PARAFAC) mathematical model

The model discussed so far has considered variations in only two modes—tongue points and vowels. The third mode of our data set—variations in tongue shape across speakers—is explicitly represented in the three-way principal-component and factor-analytic model of our data. Here the deviation of the original three-way data set is represented as follows:

$$d_{ijk}=t_{i1}v_{j1}s_{k1}+t_{i2}v_{j2}s_{k2}+\cdots+t_{iw}v_{jw}s_{kw}+e_{ijk}\,. \qquad (4)$$

This model is the same as the traditional model, except that each term is multiplied by an additional coefficient. The first of these coefficients, the s_{k1} value, represents the relative emphasis of the first feature by person k. Similarly s_{kw} gives the relative degree of use of feature w by person k. The PARAFAC analysis will provide, in addition to the tables of t and v values described previously, a third table of s values describing the relative degree to which each speaker emphasizes a given feature across all vowels.

The incorporation of this third mode has an important property besides facilitating analysis of three-way data. It provides one possible solution to the problem of nonuniqueness present in two-way factor analysis. We can best see how this occurs by referring again to our example of the simple equation $x+y=20$. One solution to the nonuniqueness of x and y for such equations consists of providing additional constraints on the solution by bringing in additional information about the unknown variables. In terms of our original analogy, unique values of x and y can be determined if we consider several related equations in parallel, such as $x+y=20$ and $2x+3y=55$. If we require the same values of x and y to satisfy both equations, a unique solution is obtained. Similarly, PARAFAC obtains a unique solution for the set of factor loadings by considering several *sets* of equations in parallel (in our case, one set of equations for each subject).

But not all additional equations will do. Note that the coefficients of x and y must not have the same ratio to one another in the two simultaneous equations, if the second equation is to provide additional constraints and thus guarantee a unique solution. The equation $2x+2y=40$, for example, would not add any new constraints. So, too, with PARAFAC: the ratio or relative emphasis of the two (or more) factors must differ across at least some of the subjects in order to provide the additional constraints necessary to obtain unique factors. (For further discussion and mathematical proofs, see Refs. 1 and 2). This point will become important in our discussion of the obtained tongue factors, below.

The PARAFAC model will provide a solution which is meaningful or explanatory only to the extent that the data conform to the mathematical model described above. This model assumes that there is a particular kind of systematic difference among speakers. It presumes that each speaker can be characterized by coefficients indicating his degree of use of each feature. Thus if speaker A uses more of factor 1 than does speaker B for a particular vowel, then speaker A must use more of factor 1 than speaker B in all other vowels. The ratio of any two speakers' usage of a given factor must be the same for all vowels. If this assumption of systematic variation is valid, then a unique and explanatory set of features (factors) can be obtained from the data. (See Ref. 1, pp. 19–22 for further discussion of types of three-mode data variation in relation to the PARAFAC model).

4. Analysis procedure

a. Criterion for convergence. Each PARAFAC solution is computed by starting with a random set of values for the factor loadings (i.e., random t, v, and s tables) and incrementally improving on this first guess by means of an iterative process. This process changes these values in small steps improving the fit of the solution to the data at each step until a final optimal solution is reached. The stepwise changes in factor loadings are large at first, and become smaller as the solution approaches its final form at which point no more stepwise improvement is possible for that solution. Convergence of this iterative process onto a stable set of values is determined by comparing each value of t and v with their values earlier in the iterative process (in this study we used the values 30 iterations previously). Convergence is considered adequate when the largest change in any value is less than 0.001. This criterion corresponds to a cumulative change over 30 iterations of less than 0.1% in the mean observed val-

ues. Experience has shown that when changes in the values become this small over that many iterations no number of further iterations is going to appreciably alter the solution.

b. Testing uniqueness of solutions. Before settling on a final solution for a given data set it is necessary to determine whether there are any alternative solutions that might be found if different random starting values are used in the iterative procedure. With error-free appropriate data, PARAFAC will provide a unique solution whenever the necessary independent systematic variation in factors across speakers, vowels and tongue points has occurred and whenever the number of factors being extracted matches the number underlying the data. With noisy data, however, both pseudo-non-uniqueness and pseudo-uniqueness can occur. Therefore it is necessary to investigate the uniqueness properties of the data in question at each dimensionality of interest.

To do this an analysis is performed with three different random starting positions for each dimensionality being investigated. These three converged solutions are then compared to one another to determine whether all three solutions represent the same factors. If all solutions agree, the common solution is tentatively assumed to be unique for that dimensionality. If all three do not agree, the several versions are compared to determine their relation to one another and to solutions at lower and higher dimensionality.

Non-uniqueness may be due to a true multiplicity of possible solutions, which may be the result of extracting too many factors, or of insufficient independent variation of the effects of the factors across speakers, vowels, or tongue points. It may also be due to presence of a local optimum. If one of the competing solutions has a substantially worse fit to the data, it is not considered to be evidence of a true multiple solution at that dimensionality, but rather as the result of a local optimum which has trapped the iterative procedure before it had reached the true optimum. Alternative solutions of approximately equal fit value (where r^2 differs by less than 0.005) indicate the true multiplicity of underlying solutions, and are suggestive of the extraction of too many factors. Because of noise in the data, however, attempts to extract too many factors will often fail to produce non-uniqueness. On the other hand, non-uniqueness can occasionally occur when two few factors have been extracted. For example, one may have a stable, unique, four-dimensional solution but two competing solutions at three dimensions. On closer examination it may become evident that factors one, two, and four of the four-dimensional solution appear in one of the three-dimensional solutions, and factors one, two, and three appear in the other. This example demonstrates how comparison of the content of alternative competing solutions with each other and with solutions at different dimensionalities is sometimes necessary to adequately interpret the nature of the observed non-uniqueness.

c. Examining solutions across the range of possible dimensionalities. In order to determine the number of dimensions that will be used to describe the data, three solutions are computed at each dimensionality from one dimension up to at least three dimensions above the maximum number of dimensions probably present in the data. In this study, since it was anticipated that only a few factors would be involved, solutions were computed for up to seven dimensions. It was expected that the higher-dimensional solutions (beyond four) would not contribute useful new information. The analysis of these higher dimensionalities was performed mainly to confirm that the fit value of the solution does not substantially improve for these dimensionalities. A plot of fit versus dimensionality is then constructed (see, e.g., Fig. 12) and examined to determine if there is a sharp bend or elbow in the curve indicating the highest dimensionality where substantial improvement in fit is obtained by allowing for additional dimensions.

In addition to this plot of fit versus dimensionality of solutions (which is often quite ambiguous in that it does not have a clear elbow) the actual form of the solution at different dimensionalities is compared, and the nature of the changes from one dimensionality to the next is used as a guide to help determine the appropriate dimensionality. Interpretability of individual solutions and of the changes from one solution to another are used to help guide choice of dimensionality.

The best means of testing the "genuineness" (statistical stability) of additional factors was unfortunately not feasible for the data analyzed here. This technique relies on the comparison of two solutions, each of which is based on one-half of the total subject sample randomly divided. Spurious factors or those too weak to be well determined by the data will not crossvalidate in these split-half solutions.

d. Additional analysis at likely dimensionalities. Once a likely dimensionality is selected on the basis of the criteria described above, three additional solutions are run at this dimensionality as well as at the dimensionality just below and just above this dimensionality. If the preferred dimensionality is n, this provides six different random starting positions for the n-dimensional, the $n-1$-dimensional and the $n+1$-dimensional solutions. With six solutions a statistical claim about uniqueness can then be made. If all six random starting positions give the same final version of the solution, the probability is less than 0.05 that another equally likely solution exists at that dimensionality. If alternative solutions exist this process gives a better idea of whether there are two or more than two possibilities.

The preferred solution, based on all the criteria mentioned above, is then selected for study and interpretation in more detail.

II. RESULTS

As we have noted, the first task in considering the results of a PARAFAC analysis is to decide how many factors underlie the variations in the data. For expository reasons this point will be considered after we have discussed the nature of the factors that are present. To begin with we will assume that there are only two identifiable factors in our data. We can then show how these factors can be used to describe the variations in the data in each of the three modes of the analysis. We will first discuss how the movements of individual points

TABLE II. Factor loadings t_{i1} and t_{i2} for each point on the tongue, and the contribution of each point to the mean-squared error in reproducing the data, for the two-dimensional PARAFAC solution.

Point (i)	Factor loadings Factor 1 (t_{i1})	Factor 2 (t_{i2})	Contribution to error of fit
4	−0.05949	−1.1900	0.00249
5	−0.8817	−1.0940	0.00235
6	−1.0500	−0.9461	0.00215
7	−1.1640	−0.5893	0.00274
8	−1.1370	−0.0056	0.00222
9	−0.9540	0.4594	0.00222
10	−0.5536	0.9281	0.00205
11	0.1349	1.2440	0.00194
12	0.8719	1.4180	0.00207
13	1.2590	1.3670	0.00202
14	1.4060	1.2420	0.00150
15	1.3190	0.9281	0.00236
16	0.9063	0.4905	0.00444

Total mean-squared error = 0.03055

on the tongue can be described, then how variations between vowels can be specified, and then how the individual speakers differ from one another.

A. Tongue points

The PARAFAC procedure shows that the shape of the tongue can be described in terms of the values shown in Table II, which are the values for the tongue-point coefficients t_{i1} and t_{i2} in Eq. (4). These two sets of coefficients are represented graphically in Fig. 3. Each may be thought of as determining a possible movement of the tongue. Each movement consists of a deviation of the tongue from its reference position (shown by solid points) either in the direction of the arrow or away from it. The possible size of the movement for each point is indicated by the length of the arrow through that point. For example, consider a vowel in which the first factor has a large negative value, and the second factor has a zero value. In this vowel the tongue shape will be as shown by the solid line in the left-hand diagram.

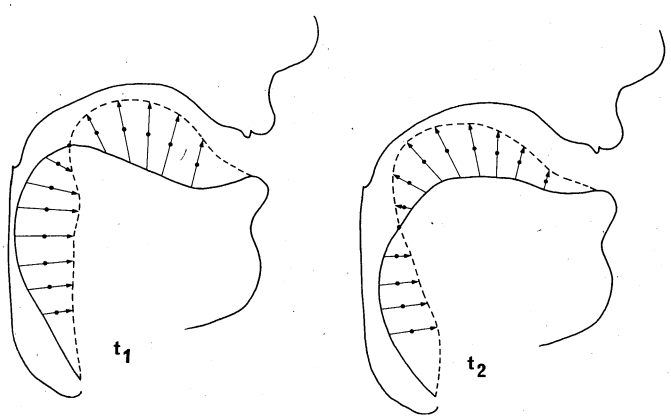

FIG. 3. A graphical representation of the two sets of constants, t_{i1} and t_{i2}, proportions of which can be combined to form different vowels. The solid points represent a neutral position of the vocal tract. The solid lines indicate tongue positions corresponding to large negative values of each factor. The dashed lines indicate large positive values.

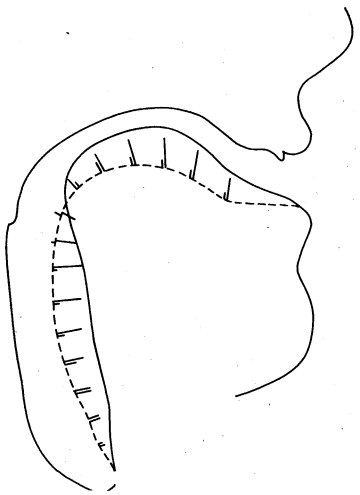

FIG. 4. The proportions of t_{i1} and t_{i2} required for the vowel /i/ as in "heed." The solid lines represent the tongue shape for the vowel. The dashed line indicates an arbitrary reference position. Lines standing out from the dashed line indicate deviations associated with t_{i1} and t_{i2}. The deviations associated with t_{i2} are to the left or below those associated with t_{i1}.

Each of the English vowels can be considered to be a combination of certain proportions of these two sets of constants. Figure 4 shows the proportions of these two sets of constants that are required to produce the vowel /i/ as in "heed." In this vowel the proportion of factor 1, t_{i1} is fairly large, and the proportion of the second factor, t_{i2}, is somewhat smaller. The contribution of each factor is shown as a set of lines deviating from the mean tongue position. The contributions of t_{i2} are to the left or below those of t_{i1}. The line representing the vowel is the sum of the two deviations. The position of the lips is estimated from other data,[17] and the position of the point immediately behind the lips is the mean of the position of the lips and the point on the tongue nearest to the lips. The positions of the first three points just above the glottis are extrapolated from the points determined by the PARAFAC solution.

Figure 5 shows in a similar way the combinations of proportions of t_{i1} and t_{i2} that occur in /ɑ/ as in "hod." This vowel is mainly associated with large negative proportions of t_{i2}. Figure 6 shows the combinations of proportions of t_{i1} and t_{i2} that occur in /u/ as in "who'd." In this case both factors are involved, often resulting in opposite deviations. In order to produce the upward and backward movement of the tongue, a positive value of t_{i2} is required; but in order to keep the front of the tongue down a negative value of t_{i1} must occur.

There is a point below the soft palate that is virtually unaffected by the first factor. Similarly, there is a point in the upper part of the pharynx that is very little affected by variations in the second factor. The first factor is associated with comparatively large movements of points in the front of the mouth, whereas the second factor is associated with larger movements in the back

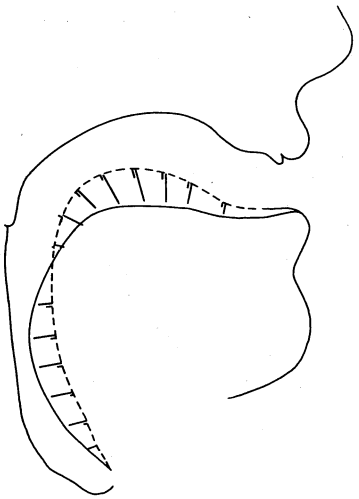

FIG. 5. The proportions of t_{i1} and t_{i2} that occur in /ɑ/ as in "hod." (Compare Fig. 4).

TABLE III. Factor loadings v_{j1} and v_{j2} for each vowel, and the contribution of each vowel to the mean-squared error in reproducing the data, for the two-dimensional PARAFAC solution.

Vowel (j)		Factor loadings Factor 1 (v_{j1})	Factor 2 (v_{j2})	Contribution to error of fit
1	i	1.5220	0.6931	0.00341
2	ɪ	0.9730	0.3293	0.00237
3	e	1.0430	0.4190	0.00292
4	ɛ	0.7739	−0.1704	0.00225
5	æ	0.3306	−0.5261	0.00495
6	ɑ	−0.3107	−2.0370	0.00400
7	ɔ	−1.0720	−1.2360	0.00322
8	o	−1.3780	0.2591	0.00195
9	ʊ	−1.0490	0.5826	0.00271
10	u	−0.8333	1.6870	0.00274

of the mouth. In general it appears that t_{i1} determines a forward movement of the root of the tongue together with a raising of the front of the tongue. This movement is very similar to the action of the genioglossus muscle. On the other hand t_{i2} determines an upward and backward movement of the tongue. Loosely speaking this factor can be thought of as corresponding to the pull of the styloglossus muscles. We hasten to add, however, that we do not think it entirely appropriate to associate either factor with particular muscles or muscle groups. The factors are simply sets of constants that enable us to characterize tongue shapes in terms of two variables, one representing the proportion of t_{i1} that is present in a given vowel, and the other representing the proportion of t_{i2}.

The PARAFAC solution has some similarities to the characterization of tongue positions proposed by Lindblom and Sundberg (Ref. 8). They characterize tongue positions in terms of a two-dimensional space in which the vowels [i, a, u] are each considered to be maximal distortions of the neutral tract, with [i] having this distortion in the palatal region, [u] in the velar region, and [a] in the pharyngeal region. Other vowels are then said to be intermediate in the place of maximum distortion, and in the degree of this distortion. Their notion is that vowels can be specified in terms of a mechanism for making a hump (and a corresponding hollow) in the neutral shape of the tongue, the two parameters being the position and size of the hump. The PARAFAC solution shows that tongue shapes can indeed be characterized in terms of distortions from a reference shape. But in this solution the tongue position is the sum of two distortions, one having minimum and maximum values for [o] and [i], and the other having minimum and maximum values for [ɑ] and [u].

We should also note that there is a good fit between the original data and the reconstruction of the data using a PARAFAC model. The correlation between these two is very high ($r = 0.9626$). The mean-squared error is 0.0305 cm. Table II shows the proportion of this error that is contributed by each point on the tongue. It may be seen that with the exception of point 16, which is near the tip of the tongue, each point contributes a similar proportion of the error. The greater error associated with the tip of the tongue may be due to the fact that its exact position has very little effect on the quality of most vowels. However, it must also be remembered that this point is often difficult to locate reliably on x-rays. The shadow of the tongue in this region is often obscured by the teeth.

B. Vowel contrasts

Table III shows the values v_{j1} and v_{j2} that are required for reconstructing each vowel. These values are represented graphically in Fig. 7. It may be seen that vowel 1, the vowel /i/ as in "heed," requires a maximum value of v_{j1} and a near maximum value of v_{j2}. Vowels 2–5, which are the other front vowels /ɪ, e, ɛ, æ/ as in "hid, hayed, head, had," also require positive values of v_{j1}, and decreasing values of v_{j2}. Vowels 6 and 7, the low back vowels /ɑ, ɔ/ as in "hod, hawed," require both components to have negative values. Vowels 8–10, the mid and high back vowels /o, ʊ, u/ as in "hoed, hood, who'd," have positive values of factor 2, but negative values of factor 1.

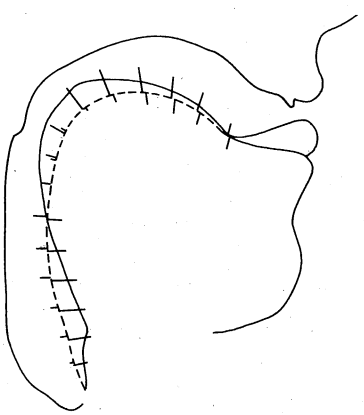

FIG. 6. The proportions of t_{i1} and t_{i2} that occur in /u/ as in "who'd." (Compare Fig. 4.)

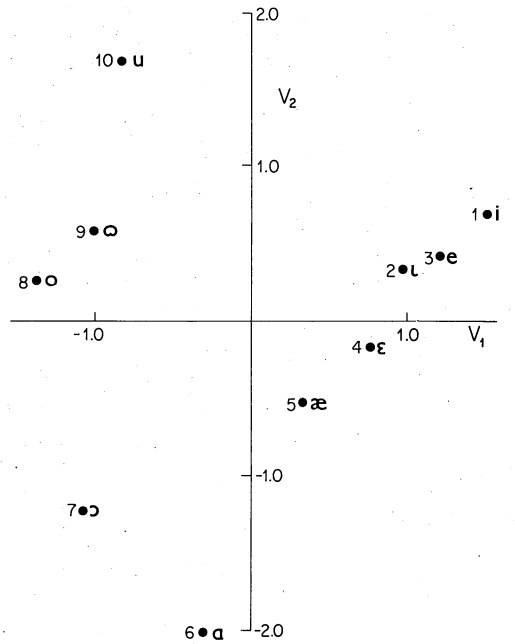

FIG. 7. The values of v_{j1} and v_{j2} for the 10 vowels in Table III.

There is a strong similarity between the arrangement of the vowels in Fig. 7, and their arrangement in more traditional vowel charts. In fact v_{j1} and v_{j2} determine a somewhat distorted formant space. Variations in the first factor are positively correlated with the second formant, and the second factor is inversely proportional to the first formant. In comparing Fig. 7 with a traditional formant chart it must also be remembered that formant frequencies are considerably affected by the lip opening, which will be continually increasing for the vowels /u/–/ɑ/ on the left of the figure.

Another way of considering the values shown in Table III is in terms of the reconstructed tongue shapes for the mean speaker. These shapes are shown in Fig. 8. They may be compared with the means of the observed displacements of the tongue for each vowel, which are also shown in Fig. 8. Whenever there is a noticeable difference between the original data and the reconstructed shapes, the curve for the original data is indicated by *o*. It is apparent that the PARAFAC analysis allows us to reconstruct mean tongue shapes that are very much in accord with the original data, and do not differ substantially in any of the regions that are important from the point of view of the function of the vocal tract.

TABLE IV. Factor loadings s_{k1} and s_{k2} for each person, and the contribution of each person to the mean-squared error in reproducing the data.

Speaker (k)	Factor loadings Factor 1 (s_{k1})	Factor 2 (s_{k2})	Contribution to error of fit
1	−0.4253	−0.2705	0.00412
2	−0.6959	−0.3590	0.00591
3	−0.3708	−0.3682	0.00680
4	−0.3383	−0.4872	0.00785
5	−0.4654	−0.3111	0.00584

C. Differences between speakers

The five speakers had different coefficients for the two factors, as shown in Table IV. It should be noted that all these coefficients have negative signs. As may be seen from Eq. (4), estimates of the data are the sums of two triple products. The sign of the products depend on the signs of each of the three coefficients. If two of the signs are arbitrarily determined, then the correct sign for the product must be determined by the sign of the third component—the speaker loadings. The signs for the values for the points on the tongue were arranged so that they could be given a plausible physiological interpretation. Those for the vowels were arranged so that the graph of v_{j1} versus v_{j2} corresponded as much

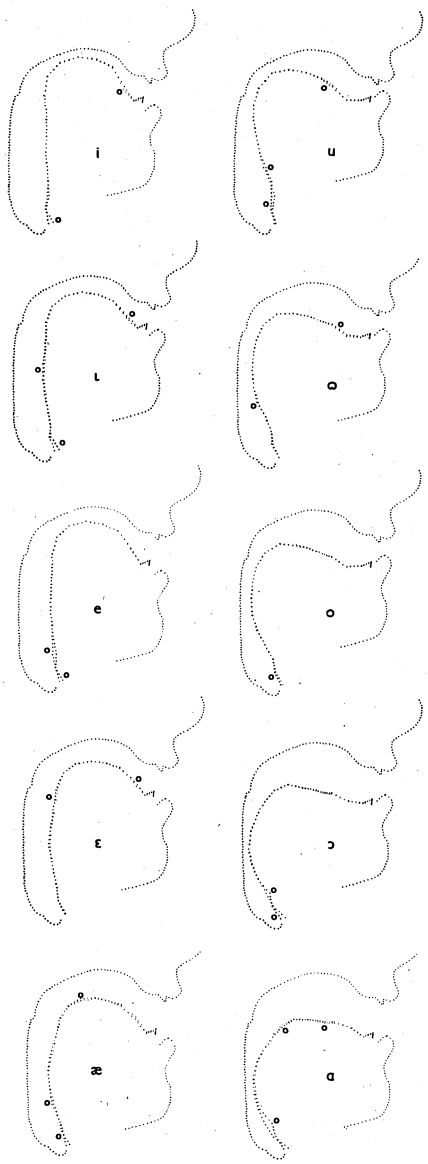

FIG. 8. Reconstructed tongue positions and mean observed displacements of the tongue for the five speakers for each of the 10 vowels. Whenever there is a noticeable difference between the reconstructed and the original data, the original data is indicated by *o* alongside the tongue position.

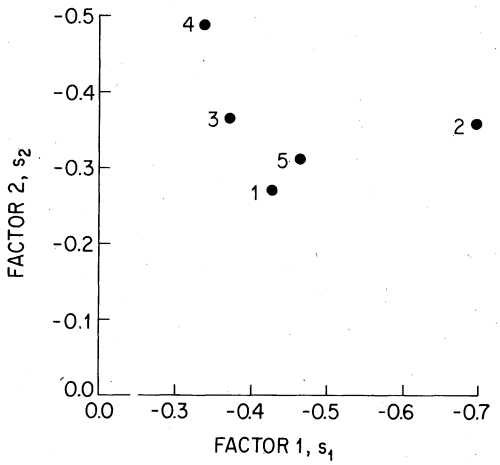

FIG. 9. The relations between the two factors for each of the five speakers.

as possible to a traditional vowel chart. But the measurement convention for tongue position measured downward from the roof of the mouth, and so a "raised" tongue (closer to the roof of the mouth) produced negative tongue deviations. As a result the signs on the speaker values all have to be negative.

As shown in Fig. 9, for four of the five speakers, the higher the value of factor 1, the lower the value of factor 2. Speaker 2 does not follow this pattern in that the weight of factor 1 is usually high while the value of factor 2 is also considerable.

There may be physiological reasons why one factor has a higher weight than another for some speakers. We do not want to give much emphasis to this point, but it seems that, within our limited data, the longer the pharynx in comparison with the oral cavity, the less[18] the influence of factor 2. The correlation between the absolute value of s_{k2} and the ratio between the pharynx-cavity length (from grid line 1 to grid line 9), and the oral-cavity length (grid lines 9–16) is -0.903 ($p < 0.05$). This particular relationship, which is shown in Fig. 10, is largely determined by one speaker. On the other hand, a higher absolute value of factor 1 is correlated with a longer oral cavity. The length of the pharynx or of the tract as a whole has less effect on the use of factor 1. For our five speakers the correlation between the absolute values of s_{k2} and the length of the oral cavity (grid lines 9–16) was 0.900 ($p < 0.05$).[19] But as may be seen from Fig. 11, which displays this relation graphically, this result should be regarded with caution, since it depends on one speaker.

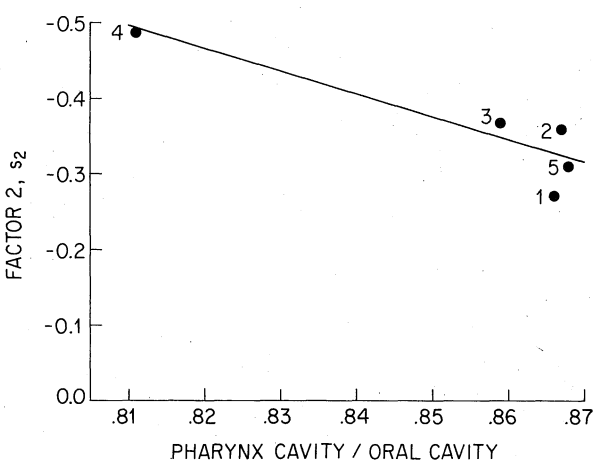

FIG. 10. The relation between factor 2, s_{k2}, and the ratio of pharynx-cavity to oral-cavity length.

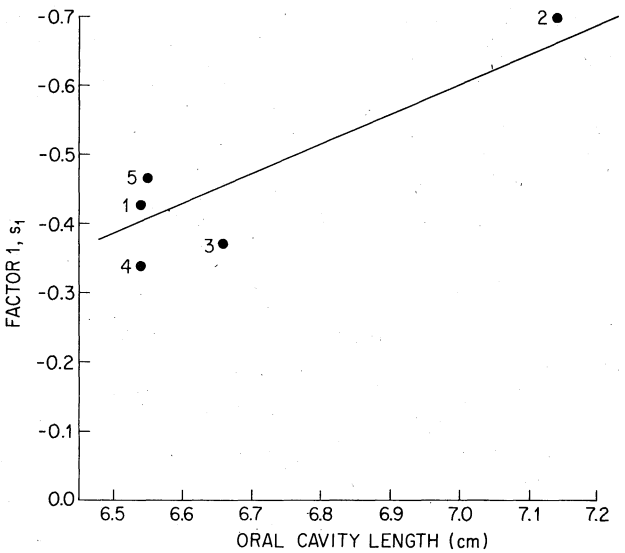

FIG. 11. The relation between factor 1, s_{k1}, and the oral-cavity length.

D. Number of factors

The PARAFAC analysis procedure finds the best analysis of the data for a given number of factors. From our results we can tell how much error there is in attempting to describe the data in terms of any number of factors from one to seven. Figure 12, shows the mean-squared error for each of these conditions, together with the variance of the data, which is taken to be equivalent to the mean-squared error when no factors have been extracted. It may be seen that a single factor ac-

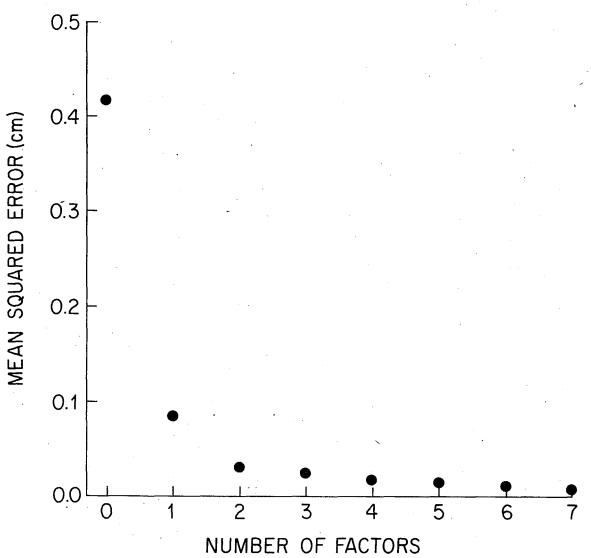

FIG. 12. The relation between the mean-squared error and the number of factors in the analysis. For comparison, r^2 values were 1D = 0.793, 2D = 0.927, 3D = 0.942, 4D = 0.956, 5D = 0.966, 6D = 0.975, 7D = 0.982.

TABLE V. A comparison of one-, two-, and three-dimensional solutions.

One factor	Two factors		Three factor (a)			Three factor (b)		
1	1	2	1	2	3	1	2	3
Tongue points								
4 −0.873	−0.595	−1.190	−0.584	−1.244	−0.306	0.267	−0.435	−1.169
5 −1.038	−0.881	−1.094	−0.875	−1.128	−0.147	0.776	−0.840	−0.962
6 −1.095	−1.050	−0.946	−1.054	−0.996	0.393	1.108	−1.092	−0.724
7 −1.032	−1.164	−0.589	−1.178	−0.644	0.941	1.393	−1.291	−0.270
8 −0.788	−1.137	−0.006	−1.162	−0.024	1.246	1.295	−1.223	0.423
9 −0.489	−0.954	0.459	−0.987	−0.457	1.472	1.077	−1.025	0.933
10 −0.039	−0.554	0.928	−0.590	0.970	1.260	0.371	−0.473	1.389
11 0.565	0.135	1.244	0.102	1.261	1.283	−0.200	0.163	1.492
12 1.148	0.872	1.418	0.845	1.417	1.059	−0.888	0.876	1.422
13 1.405	1.259	1.367	1.241	1.343	0.889	−1.332	1.296	1.209
14 1.460	1.406	1.224	1.392	1.182	0.937	−1.420	1.412	0.998
15 1.283	1.319	0.928	1.308	0.818	1.151	−1.176	1.250	0.624
16 0.804	0.906	0.491	0.900	0.394	0.883	−0.427	0.670	0.205
Vowels								
1 1.573	1.522	0.693	1.553	0.715	−0.892	−1.840	1.770	0.664
2 0.945	0.973	0.329	1.032	0.221	−1.247	−0.853	0.935	0.378
3 1.046	1.043	0.419	1.051	0.467	−0.220	−1.267	1.214	0.371
4 0.519	0.774	−0.170	0.748	−0.127	0.326	−0.185	0.422	−0.085
5 −0.024	0.331	−0.526	0.268	−0.436	1.102	0.593	−0.240	−0.416
6 −1.327	−0.311	−2.037	−0.474	−1.914	2.217	0.815	−0.790	−2.119
7 −1.501	−1.072	−1.236	−1.067	−1.450	−0.620	1.459	−1.410	−1.223
8 −0.959	−1.378	0.259	−1.364	0.268	0.444	1.017	−1.152	0.250
9 −0.523	−1.049	0.583	−1.010	0.568	−0.259	0.262	−0.551	0.514
10 0.251	−0.833	1.687	−0.738	1.689	−0.847	−0.001	−0.200	1.668
Persons								
1 −0.491	−0.425	−0.271	−0.423	−0.255	0.002	0.452	−0.874	−0.212
2 −0.763	−0.696	−0.359	−0.698	−0.313	0.042	0.983	−1.658	−0.270
3 −0.487	−0.371	−0.368	−0.374	−0.271	0.182	0.384	−0.757	−0.322
4 −0.543	−0.338	−0.487	−0.331	−0.505	−0.059	−0.320	−0.101	−0.381
5 −0.543	−0.465	−0.311	−0.462	−0.265	0.077	1.022	−1.435	−0.289

counts for a considerable part of the error. There is a further large decrease in the error if we assume that there are two factors instead of one; but there are not comparable decreases in the error when additional factors are taken into account. Each factor beyond two produces only a small, fairly constant, increase in the goodness of fit. It seems probable that each of these additional factors simply accounts for some of the random effects due to measurement error and other irregularities. The sharp bend in the fit curve at two factors and smoothness of descent thereafter suggest that there are only two identifiable factors underlying the nonrandom variations in the data. The correlation between the original set of measurements and the values generated by the two-factor analysis is 0.9626.

The notion that there are only two factors is substantiated by the data in Table V, which shows the factor loadings on the assumption that there are one, two, and three factors. Two different three-factor solutions are presented, since different random starting positions produced different results. In the first alternative solution the first two factors are much the same as in the two-factor solution, and the third factor has only a small effect since it has very low values for all speakers except for speaker three. In this solution the third factor may be accounting for noise in the data, or for some pattern specific to speaker three. In the other solution the first factor in the two-factor solution has simply split into two very similar factors. Non-uniquenesses of this kind are very good indications that the analysis is trying to extract more factors from the data than can be detected. In addition, this solution did not converge to a stable set of values. All the other solutions in Table V took less than 90 iterations to converge to values which did not vary by more than 0.001 over the last 30 iterations. But the two similar factors in this third solution did not converge to stable values even by 300 iterations. All these indications suggest that the current data set is best described in terms of two factors. This does not necessarily mean that English vowels can always be described in terms of two factors. It is possible that a weak potential third factor is confounded with noise in the somewhat limited data set at hand.

III. DISCUSSION

A. Descriptive adequacy

The PARAFAC analysis described above selects a model which provides a description of the tongue shapes of English vowels for five speakers. The description is

based on a fully explicit (and general) x-ray measurement procedure, and the model is optimally fitted to the data with a high correlation ($r > 0.96$). Tongue shapes are described both compactly and quantitatively in terms of two features, each of which represents the pattern of the tongue displacement from a reference position, at a fixed number of points along its length. Patterns of variation in tongue shape among vowels and among speakers, are also described in a systematic, quantitative manner. Each English vowel is assigned a unique representation in terms of the two features of tongue shape (output by PARAFAC as the vowel coefficients). The relationships among the feature values for different vowels is consistent with the similarity among the vowels' tongue shapes, as judged by a phonetician's intuition, or by casually examining the x-ray tracings. Finally, variations in patterns of tongue shape among different speakers are described quantitatively by the variation in the relative degree to which the two features are associated with them. The value of a given feature for a particular speaker was shown to be correlated with some gross anatomical properties of the speaker's vocal tract.

B. Explanatory value of the features

Given that the features extracted by PARAFAC in the current experiment meet the criteria for descriptive adequacy, one may go on to ask whether these features represent some underlying physiological or linguistic reality, or whether they simply provide for a convenient mathematical description of the data. The answer in part depends on whether the data conform to what Harshman (Refs. 1 and 2) refers to as a system-variation model. In such a model, a single set of underlying factors is viewed as influencing all the objects under consideration. The system influence is such that any aspects of the data-collection situation that alter a given factor's influence on the objects, alters its effects on all of the objects in the same proportion. With reference to the current experiment, the system-variation assumption translates as follows: while there may be some differences among speakers in the degree to which they use the extracted features in producing vowels, for a given pair of speakers, those differences must be in a constant proportion for all vowels. PARAFAC claims that if the condition of system variation is met, then the factors it extracts describe the entities that underlie the patterns of variation in the system. We need to consider, then, whether the assumption of system variation can be shown to be a reasonable approximation to the structure of the data at hand.

There is *prima facie* evidence for thinking that the hypothesis of system variation could be applicable to the present data set. This evidence lies in the fact that the speaker coefficients in the PARAFAC solution do show a fair amount of variability for both of the two features. Thus, speakers differ in their weightings for the features. This variability was shown to be correlated with anatomical differences between the speakers. This correlation makes it appear unlikely that the differences in speaker coefficients are solely due to the model's accounting for noise in the data. Rather, it seems to indicate that there is some systematic variation among speakers. However, whether or not that pattern of variation is proportional across all vowels is not yet known. Various ways of assessing whether this assumption is met will be discussed in a further paper.

An affirmative answer to the question of system variation implies that the factors extracted by the PARAFAC analysis correspond to some of the real sources of variation underlying the data. Moreover, the system-variation model—if appropriate—has other interesting implications. In particular, it allows us to construct a model characterizing the articulatory vowel space employed by a given dialect, while still accounting for speaker variation within that dialect. It should be recalled that the PARAFAC analysis yields a single representation of the articulatory space of vowels, in terms of their values on the two features of tongue shape. This vowel space can be taken as the single underlying vowel space for the speakers under consideration. Variation between the speakers is limited to stretching or shrinking of the vowel space along its axes. This limit of individual variation to shrinking and stretching along the various factor axes amounts to a fairly strong constraint on how the individuals within a dialect can differ from one another. It rules out the possibility of the vowel space differences between two speakers of the same dialect being specific to individual vowels. Harshman and Papçun[20] have proposed a similar use of PARAFAC analysis for normalizing differences in formant frequencies of sets of vowels spoken by different speakers.

Further important questions remain to be answered before theoretical questions of interpretation should be considered in detail: (a) is the sample large enough for a stable solution; (b) how well will these factors generalize to other sets of English speakers; (c) how well will they generalize to speakers of other languages. Once such questions of reliability and generalizability have been answered, a number of interesting more detailed psychological and linguistic questions can be framed in terms of the dimensions of tongue-shape variation uncovered by this type of analysis.

ACKNOWLEDGMENT

This research was supported in part by a grant from the National Institute of Neurological and Communicative Disorders and Stroke.

[a]Richard Harshman, is now at the Department of Psychology, University of Western Ontario, London, 72, Canada.

[1]Richard Harshman, "Foundations of the PARAFAC procedure: Models and procedures for an 'explanatory' multi-modal factor analysis," UCLA Working papers in Phonetics 16 (1970): University microfilms No. 10,085.

[2]Richard Harshman, "PARAFAC: Methods of three-way factor analysis and multidimensional scaling according to the principle of proportional profiles," Ph.D. thesis (UCLA, 1976): (University microfilms, Ann Arbor, MI, Order No. 76-25,210).

[3]The same model is developed from the point of view of multidimensional scaling in J. D. Carroll and J. J. Chang, "Analysis of individual differences in multi-dimension scaling via

n-way generalization of 'Eckart–Young' decomposition," Psychometrika **35**, 283–319 (1970).

[4]Daniel Jones, *An Outline of English Phonetics* (Heffer, Cambridge, 1956).

[5]K. N. Stevens and A. House, "Development of a quantitative description of vowel articulation," J. Acoust. Soc. Am. **27**, 484–493 (1955).

[6]C. G. M. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).

[7]J. Flanagan, C. Coker, L. Rabiner, R. Schaffer, and N. Umeda, "Synthetic voices for computers," IEEE Spectrum **7.10**, 22–45 (1970).

[8]B. Lindblom and J. Sundberg, "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179 (1971).

[9]J. Liljenkrants, " A Fourier series description of the tongue profile," Speech Transmission Lab (Stockholm) **QPSR 4**, 9–18 (1971).

[10]P. Mermelstein, "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am. **41**, 1283-1294 (1967).

[11]S. Kiritani, K. Miyawaki, O. Fujimura, and J. E. Miller, "A computational model of the tongue," J. Acoust. Soc. Am. **57**, S3 (A) (1975).

[12]J. Perkell, "Physiologically-oriented model of tongue activity in speech production," Ph.D. dissertation (MIT, 1974) (unpublished).

[13]Peter Ladefoged, Joseph DeClerk, and Richard Harshman, "Control of the tongue in vowels," in *Proceedings* 7th *International Congress of Phonetic Sciences*, edited by Andre Rigault and Rene Charbonneau (Mouton, The Hague, 1972), pp. 349–354.

[14]E. T. Uldall, "American 'molar' r and 'flapped' t," Revista do Laboratorio de Fonetica Experimental, Coimbra **4**, 103–106 (1958).

[15]P. Horst, *Factor Analysis of Data Matrices* (Rinehart and Winston, New York, 1965).

[16]H. H. Harman, *Modern Factor Analysis* (University of Chicago Press, Chicago, 1960).

[17]Victoria Fromkin, "Lip positions in American English vowels," Language and Speech **7**, 215–225 (1964).

[18]Since the negative speaker weights are the result of arbitrary conventions (e.g., which direction of tongue displacement is considered negative), and since the weights for all speakers have the same sign, it is the *size* and not the sign of these weights which is crucial in evaluating the relative importance of a given factor across speakers. Consequently, we used the absolute values of these weights when correlating them with physiological variables.

[19]The significance levels reported for these correlations should be interpreted as only suggestive, at best, since it is not clear how well the required independence and normality assumptions are met by these data.

[20]Richard Harshman and George Papçun, "Vowel normalization by linear transformation for each speaker's acoustic space," J. Acoust. Soc. Am. **59**, S71 (A) (1976).