Title
WPP, No. 22

Publication Date
1972-03-01

Working Papers in Phonetics

22

U.C.L.A. March 1972

UCLA Working Papers in Phonetics 22

March, 1972

The UCLA Phonetics Laboratory Group

Research	Technical and Secretarial
Victoria A. Fromkin	Ron Carlson
Richard Harshman	Julie Haaker
Leon Jacobson	Willie Martin
Stephen Krashen	Kathy Saracino
Peter Ladefoged	Sarah Spitz
Mona Lindau	Renee Wellin
George Papçun	
Lloyd Rice	
Dale Terbeek	
Diana Van Lancker	

As on previous occasions, the material which is presented here is simply a record for our own use, a report as required by the funding agencies, and a preliminary account of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS grant NS 09780
ONR contract NR 049-226
NSF grant GS 2859
and the UCLA Department of Linguistics

Correspondence concerning this series should be addressed to:

Professor Peter Ladefoged
Phonetics Laboratory
UCLA
Los Angeles, California 90024

Recent Publications

Peter Ladefoged, Ruth Glick, and Clive Criper. (1971) *Language in Uganda*. Nairobi, Kenya: Oxford University Press.

Peter Ladefoged. (1971) *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Analysis of Tongue Parameters for Vowels

Lloyd Rice

[Paper to be presented at the 83rd Meeting of the

Acoustical Society of America]

A computer system is described which will accept a spoken vowel as input and then adjust the area function of a vocal tract synthesizer to duplicate the input vowel quality as nearly as possible. The synthesizer output is matched against the input vowel spectrum using the Markel inverse filtering system. The program allows the definition of alternative parameter models for controlling the vocal tract area function and represents a preliminary attempt to simulate the major vocal articulators. Each of three to six parameters is defined in terms of its contribution to the movement of each of the eighteen sections used in the line analog synthesizer. Thus, the action of parameter Pi is defined by the 18-element vector of weights W_{i1}, W_{i2}, - - - , W_{i18}. Any vocal tract shape within the range of the model is specified by assigning the appropriate value Pi to each of the parameters. The deflection of the jth section away from the neutral tract position is then given by the sum of products $P_1W_{1j}+P_2W_{2j}+P_3W_{3j}+$ - - - . Various parameter models will be explored to determine their usefulness in providing unique and reasonable tongue shape descriptions.

An "Unbiased" Procedure for Comparing Degree of Lateralization of Dichotically Presented Stimuli

Richard Harshman and Stephen Krashen

[Paper to be presented at the 83rd Meeting of the Acoustical Society of America]

In some experimental designs, it is desirable to compare "degree of lateralization" for different stimuli or different groups. Current measures of ear difference make such comparisons difficult, because scores are influenced by subjects' relative accuracy in identifying the stimuli. One reason for this "bias" is demonstrated using some simple mathematical models of a subjects' behavior in dichotic listening. A scoring method is deduced which is not biased by different amounts of guessing by subjects. This procedure, the percent of errors method (POE), consists of computing each ear's percentage of the total number of errors. The utility of the POE method is confirmed by reanalysis of previously published data. Standard methods for measuring ear differences are found to be negatively correlated with changes in accuracy. POE scores, however, do not correlate significantly with accuracy scores. This is true whether accuracy differences are due to age, first versus second channel reported, or differences in stimulus difficulty. Re-analysis of consonants versus vowels, younger versus older subjects, and other comparisons of degree of lateralization suggest some alternatives to previous conclusions.

Lateralization and the Critical Period

Stephen Krashen and Richard Harshman

[Paper to be presented at the 83rd Meeting of the
Acoustical Society of America]

The percent of errors method, described in Harshman and Krashen, above, was applied to dichotic listening data for children between five and ten. No significant change in degree of lateralization was found. This supports the hypothesis that the development of lateralization is complete by age five, rather than by puberty, as argued by Lenneberg. Data cited by Lenneberg as well as additional cases were examined. Analysis of cases of acquired aphasia in children resulting from localized lesions revealed that right-sided lesions do not cause more cases of speech disturbance after five than in adults, indicating that the involvement of the right hemisphere in the language function in children older than five is the same as in adults; this also supports the age five hypothesis. Not enough crucial data is available from hemispherectomies to decide between the two hypotheses. The completion of lateralization by five implies that lateralization reflects the maturation of some essential component of the language faculty, rather than the termination of "organization plasticity," as suggested by Lenneberg.

PARAFAC2: Extensions of a Procedure for "Explanatory" Factor-Analysis and Multidimensional Scaling

Richard Harshman

[Abstract of paper presented at the 82nd Meeting of the Acoustical Society of America]

Recently developed techniques for factor analysis and multidimensional scaling (PARAFAC-INDSCAL) allow discovery of a unique orientation of axes and thus "more explanatory" factors or dimensions. A certain price is paid for these advances, however. In multidimensional scaling applications, the price is a restriction of the possible form of the solution to orthogonal "nonrelated" as opposed to oblique "perceptually related" dimensions. In factor analytic applications, the price is a limitation to those three-way data sets showing "system-variation" [Harshman, UCLA W.P.P. No. 16 (1970)]. An extension of PARAFAC to summed-cross-product matrices (e.g., covariance matrices) overcomes both of these limitations. For factor analysis with PARAFAC2 it is only necessary that the average effect of a given factor change, relative to the other factors, from one covariance matrix to the next. For multidimensional scaling, PARAFAC2 describes the perceptual space in terms of orthogonal dimensions only when this best fits the data. Otherwise, it will recover the stimulus projections on oblique dimensions and also give the angles between these dimensions. Advantages of PARAFAC2 are explored by comparing its results with those of PARAFAC. Both factor-analytic and multidimensional scaling applications are demonstrated using data from speech physiology and perception.

PARAFAC: Parallel-Factor Models Yielding Mathematically Unique, Empirically Meaningful Solutions with Certain Classes of Three-Mode Data

Richard Harshman

[Paper to be presented at the spring 1972 meeting
of the Psychometric Society]

PARAFAC is a three-mode extension of traditional factor analysis which can uniquely determine an empirically meaningful or "explanatory" orientation of factors by using the additional information present in certain three-way data sets.

Our starting point is Cattell's principle of seeking "parallel proportional profiles" for the loadings on a factor extracted from two related data sets. The mathematical definition of "proportional profiles" is generalised from a rotation criterion into both a three-mode factor model of *data*, and another three-mode model for *cross-products*.

The simpler model, PARAFAC1, describes a set of related two-way data tables in terms of common factors in the following manner:

(1) $$X_i = A\,D_i\,B'$$

where X_i is the ith two-way set of data consisting of n measurements on each of m things. A is an n by l matrix of factor loadings for the n variables, and B is an m by l matrix giving factor scores for the m things. D_i is the ith l by l diagonal matrix giving the changes in the size of the l factors for the ith occasion. This decomposition is unique (except for column order and simple scale changes) when the X, A, B and D matrices are all at least of rank l, and $(D_i D_j^{-1})$ has distinct diagonal elements for some i,j. It is *not* necessary that A and B have orthogonal columns.

This model describes "system-variation" type of data. Such a variation pattern can be expected when a number of variables are measured on each of a number of different components of some *unified system* (e.g. a biological, economic, or ecological system). PARAFAC1 is too restricted to describe "object-variation", the data pattern obtained when a number of variables are measured on each of a number of *independent objects* (e.g. personality variables measured on the same set of individuals on several occasions). (See Harshman 1970, UCLA W.P.P. # 16)

Several examples of the successful application of PARAFAC1 will be discussed.

When PARAFAC1 is applied to scalar product data, it is equivalent to INDSCAL, the three-mode multidimensional scaling model developed by Carroll and Chang.

In multidimensional scaling applications, the uniqueness provided by PARAFAC1 or INDSCAL is bought at the price of restricting the solution to orthogonal perceptual axes.

In order to overcome both the limitation to "system-variation" in factor analysis and the restriction to orthogonal axes in multidimensional scaling, PARAFAC2 extends the proportional profiles approach to analysis of summed-cross-product matrices. If C_i is our ith cross-product matrix, then

(2) $C_i = X_i X_i' = A D_i \Phi D_i A'$

which is the factor model of PARAFAC2. Here X_i, A and D_i are defined as before, and Φ is the matrix of cross-products between factors. As long as the angles between factors do not change across occasions, this model allows analysis of "object-variation." In fact, each C_i matrix can be derived from measurements taken on a different population.

For multidimensional scaling, PARAFAC2 describes the perceptual space in terms of orthogonal dimensions when this best fits the data. Otherwise, it will recover the stimulis projections on oblique dimensions and also give the angles between these dimensions.

Comparison of PARAFAC1 and PARAFAC2 analysis of vowel similarity data (derived from triadic comparisons) suggests the presence of some oblique perceptual dimensions. Several possible interpretations of this obliqueness will be discussed.

A useful research strategy might be to use PARAFAC2 in conjunction with Carroll's more general (but non-unique) model IDIOSCAL.

PARAFAC provides a new and powerful additional criterion for determining the number of factors underlying a set of data. When too many factors are extracted from a data set, the uniqueness breaks down.

PARAFAC can also aid in the discovery and accurate representation of non-linear factor relationships.

Both of these features are illustrated by a series of PARAFAC analyses which revealed an apparent non-euclidian curvature in the perceptual space for vowels.

Language Lateralization and Grammars

Diana Van Lancker

[Abstract of a forthcoming paper]

The grammar of a language is inferred from observed performance data, according to theoretical claims as to what generalities belong to the competence of a speaker underlying, or generating, his performance. The important question, what aspects of actual language use (parole, speech, performance) constitute appropriate data for grammatical analysis, has not been satisfactorily answered, but remains controversial. It is argued in this paper that neurophysiological and experimental evidence can contribute a principled basis for evaluating certain facts of performance for their relevance to writing grammars.

Evidence from various sources (aphasia, hemispherectomy cases, epileptic ictal speech and split-brain tests) is presented to suggest that 1) actual language use constitutes a spectrum of identifiable subsets, from an extreme of "most propositional language" to the other extreme of "most automatic speech"; and 2) "propositional language" is the specialized capacity of the left "dominant" hemisphere in man, while modes of language use called "automatic speech" are a capacity of both hemispheres. The "automatic" modes may be considered less language-like, insofar as they do not need to be processed according to the unique language-processing mechanisms known to be lateralized in man.

The subsets referred to as "more automatic speech" are emotional expressions, overlearned phrases including songs and poems, lists and serials such as the alphabet and days of the week, social gestures of speech such as greetings, memorized phrases, pause fillers, exclamations and expletives, some idioms, and others yet to be determined. Should a grammar of English generate "How are you" despite its automatic character? How does the grammarian approach embedded "ya know" and the ubiquitously generated "like" of everyday English? Questions of this kind are addressed in this paper.

Such expressions are often retained in aphasic speech. Clinical observations in expressive aphasia are unanimous that "propositional" or "intellectual" use of language is impaired, but that other modes are less impaired. (Jackson, 1876; Head, 1926; Goldstein, 1948; Ombredane, 1951; Wepman, 1956; Brain, 1961; Bay, 1964; Luria, 1964; Lenneberg, 1967.) For example, Alajouanine, the French aphasiologist, noted that for the aphasic patient, "situations leading to emotional, expletive, interjectional, automatic, habitual language are always more effective that those requiring volitional or propositional speech." (1956)

The first to distinguish "automatic" use of language from what he called "voluntary" language use was Hughlings Jackson (1876, reprinted 1932), who made the stronger claim that the "right cerebral hemisphere is the one for the most automatic use of words, and the left the one in which the automatic use of words merges into the voluntary use of words -- into speech." Jackson's views of a century ago are remembered by many people currently doing research on lateralization and the role of the non-dominant, right hemisphere. The extreme view of total lateralization of language processing, standard a few decades ago, is being qualified on the basis of clinical and experimental data, in favor of such concepts as degrees of dominance, degrees of lateralization, and graded functional lateralization of language. (Conrad, 1949; Ettlinger, Jackson and Zangwill, 1955; Hecaen and Piercy, 1956; Penfield and Roberts, 1959; Zangwill, 1960; Bay, 1964) Studies specifically on the possible linguistic capabilities of the right hemisphere have claimed to demonstrate certain levels of comprehension (Eisenson, 1962; Gazzaniga, 1970) and of expression (Gazzaniga and Sperry, 1967; Butler and Norrsall, 1968; Levy, Nebes and Sperry, 1971; Gazzaniga and Hillyard, 1971)

There is further clinical evidence that residual modes of language behavior in aphasia are retained in the unimpaired right hemisphere. In four reported cases of left hemispherectomy (Zollinger, 1935; Crockett and Estridge, 1951; Hillier, 1954; Smith, 1966), such language modes as swearing, "yes" and "no", the whole phrases "I don't know" and "I don't want any" and singing familiar songs were observed in the post-operative patient. Furthermore, studies on speech behavior in epileptics lends strong support to the hypothesis that the right hemisphere is capable of types of "automatic" language. (Bingley, 1958; Hecaen and Angelergues, 1960; Serafetinides and Falconer, 1963; Falconer, 1967).

The neurophysiological and experimental evidence reviewed points to distinctly different processes underlying the generation of such expressions as "good morning,""dammit," and "I don't know." We can assume that everyday language behavior is made up of all types of the subsets from propositional language to automatic speech, which obviously require non-uniform modes of grammatical analysis to accommodate them all. It is suggested that because "automatic" modes of language are processed differently, they need be therefore neither described in terms appropriate to propositional language, nor represented uniformly with other aspects of language production in a performance model. It is possible that to include automatic modes in a model of competence, provisions for specialized treatment will be necessary, such as unique constraints on the generation of such items, selective immunity to classes of transformations, unique storage in the lexicon, or special bypassing of certain phonological rules. Speech error data, for example, which may exhibit different tendencies in different modes of speech, would permit hypotheses about stages of phonological rules bypassed. It is also possible that a performance model must differentiate between language processing, and generation of language-like behavior, having the characteristics of the automatic modes. In the Fromkin model, (1971), the initiation stage may not be relevant, or valid, for the generation of automatic language, in which something less complex than the semantic-syntactic pairing operation of propositional speech seems to be the case. Thus, such data leads to interesting hypotheses regarding controversies in what constitutes the speakers' competence and how to build a performance model.

Is Vowel Perception Non-Euclidean?

Dale Terbeek and Richard Harshman

[Paper presented at the 82nd Meeting of the Acoustical Society of America]

In studying the structure of vowel perception, it is customary to represent vowels as points in a space, where perceived dissimilarities between vowels are represented as distances between the points in the space. With a traditional multidimensional scaling technique, an investigator can produce a configuration of points whose relative distances reflect more or less adequately an experimentally derived set of dissimilarities between vowels. But in order to do this, he must estimate the true number of dimensions which underlie his subjects' behavior. Once he has obtained an adequate configuration, he must try to rotate it so that the axes of the space correspond to the fundamental properties of the stimuli used by the listeners in making their judgments.

A recent improvement on this simple model is the PARAFAC -- INDSCAL approach (Harshman 1970, Carroll and Chang, 1970). This model has two powerful properties of interest here. First, the axes in the derived space theoretically correspond to the true, or psychologically real, dimensions used in the perception of the stimuli. [1] In this model, the orientation of axes is uniquely determined, and thus not subject to rotation. Second, the program provides a new criterion for estimating the true number of dimensions. For each dimensionality tested, PARAFAC (or INDSCAL) will converge to the same configuration of points and the same orientation of axes no matter what the starting position -- unless more dimensions are extracted than are systematically present in the data.

---

1 It is worthwhile to qualify claims of psychological reality which are based on PARAFAC analyses. There may be real psychological dimensions which are simply not invoked in a given experiment; for example, if the stimulus set does not include both nasal and non-nasal vowels, there will be no nasality dimension in the resulting perceptual space, even if a nasal/non-nasal contrast in the subjects' language makes nasality psychologically real for him. On the other hand, the experimental design may influence the subjects' behavior to produce dimensions that are valid only for the particular experiment. This is possible if listeners are asked to respond, say, to vowels which have been systematically filtered or otherwise distorted. Conclusions as to psychological reality, then, must take into account the stimuli used and the experimental design.

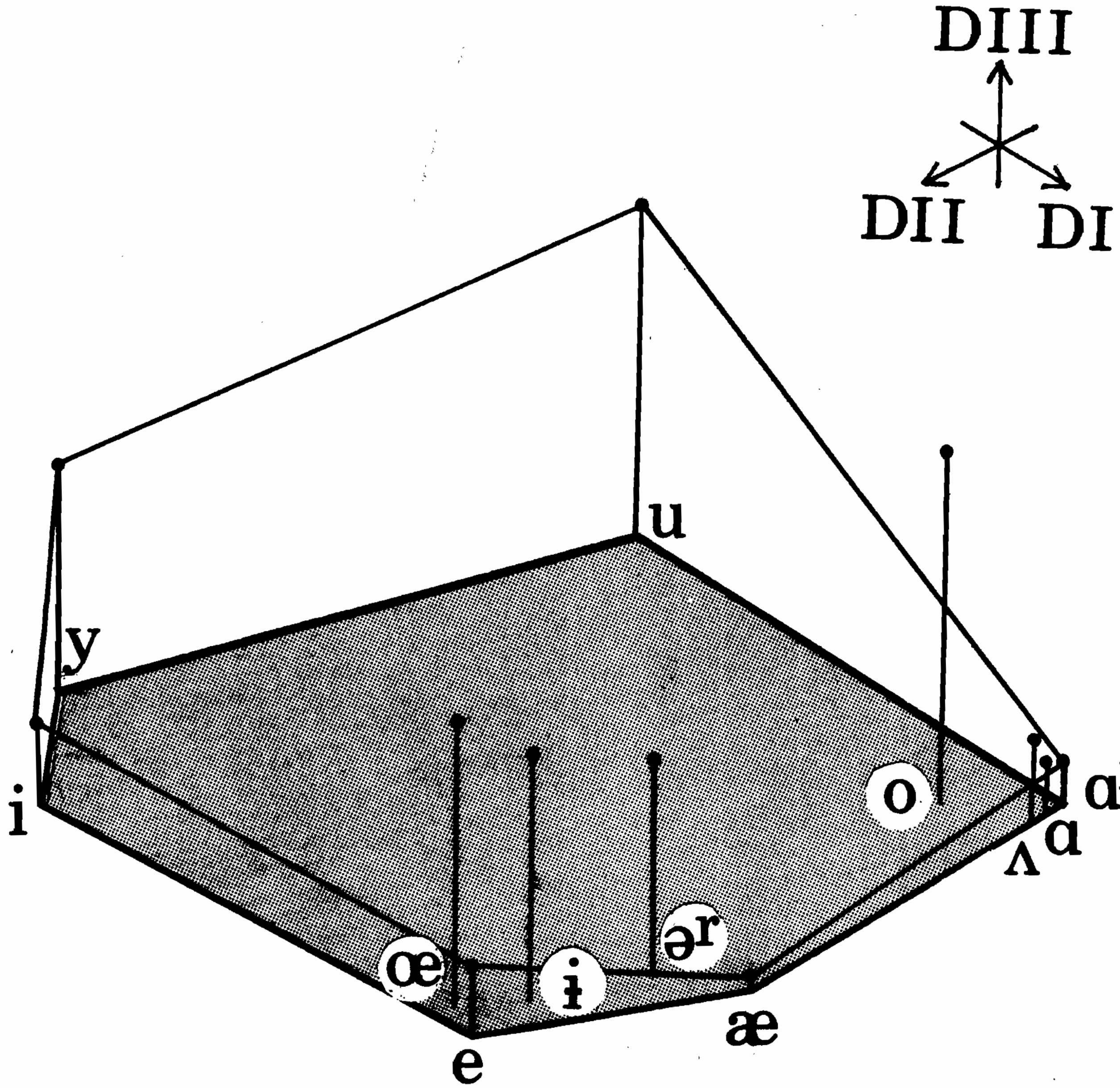


Figure 1. Three-dimensional perceptual vowel space for German.

When too many dimensions are extracted, PARAFAC will converge to different solutions from different starting points. In theory, then, the true dimensionality is the highest dimensionality before uniqueness of solution breaks down.

It was the use of PARAFAC in analyzing data from vowel perception experiments that led to the question posed by the title. Some analyses were easily interpretable in traditional terms, but others were not. We have found that many difficulties in interpretation can be traced to two assumptions made in the analysis model (traditional multidimensional scaling and PARAFAC as well): 1) The space must be Euclidean. It is assumed that all pairs of dimensions intersect to form flat planes; the model cannot express results in terms of curved spaces. We will show how this assumption can be bypassed by simply reinterpreting solutions that are obtained. 2) The space must be Cartesian; the dimensions must be orthogonal to each other. To counter this assumption, the PARAFAC model itself was revised to allow solutions in which the angles between the dimensions were not fixed at 90°. [2]

In discussing these two assumptions and their relationship to our work, we will refer to data collected in triadic comparison experiments using listeners from five different native language backgrounds: English, German, Thai, Turkish, and Swedish. A more thorough description of the experimental procedure and some results of the analysis of judgments made by English, Thai, and German listeners can be found in a previous paper (Terbeek and Harshman, 1971).

At this point it will be useful to briefly describe the difference between results which are easy to interpret and results which pose problems. An example of a PARAFAC analysis that caused little difficulty in interpretation is shown in figure 1. This is a derived perceptual space for five of the seven German listeners taken together. Dimension F1 represents the linguistic scale of front/back, dimension F2 can be interpreted as dividing the space into rounded and unrounded vowels, and dimension F3 clearly separates high vowels from low and mid vowels. Not only is the configuration of points satisfactory, but the orientation of axes is such that each axis can be equated with a known linguistic criterion. Also, the number of dimensions matches our expectations.

Many of our analyses, though, have not been so straightforward. There are three basic troublesome situations. 1) A PARAFAC space may be obtained that is easy to interpret except for one extra "real" dimension ("real" according to the criterion of uniqueness of solution) in which the vowels are not ordered in any expectable sequence. If this occurs, one is obliged to search for an interpretation of the extra dimension. 2) The uniqueness criterion suggests many more dimensions than seem reasonable, and neither the configuration nor the orientation of axes is satisfactory.

---

2 A stodgier, though more accurate title for this article would have been, of course, "Is Vowel Perception Non-Euclidean and Non-Cartesian?"

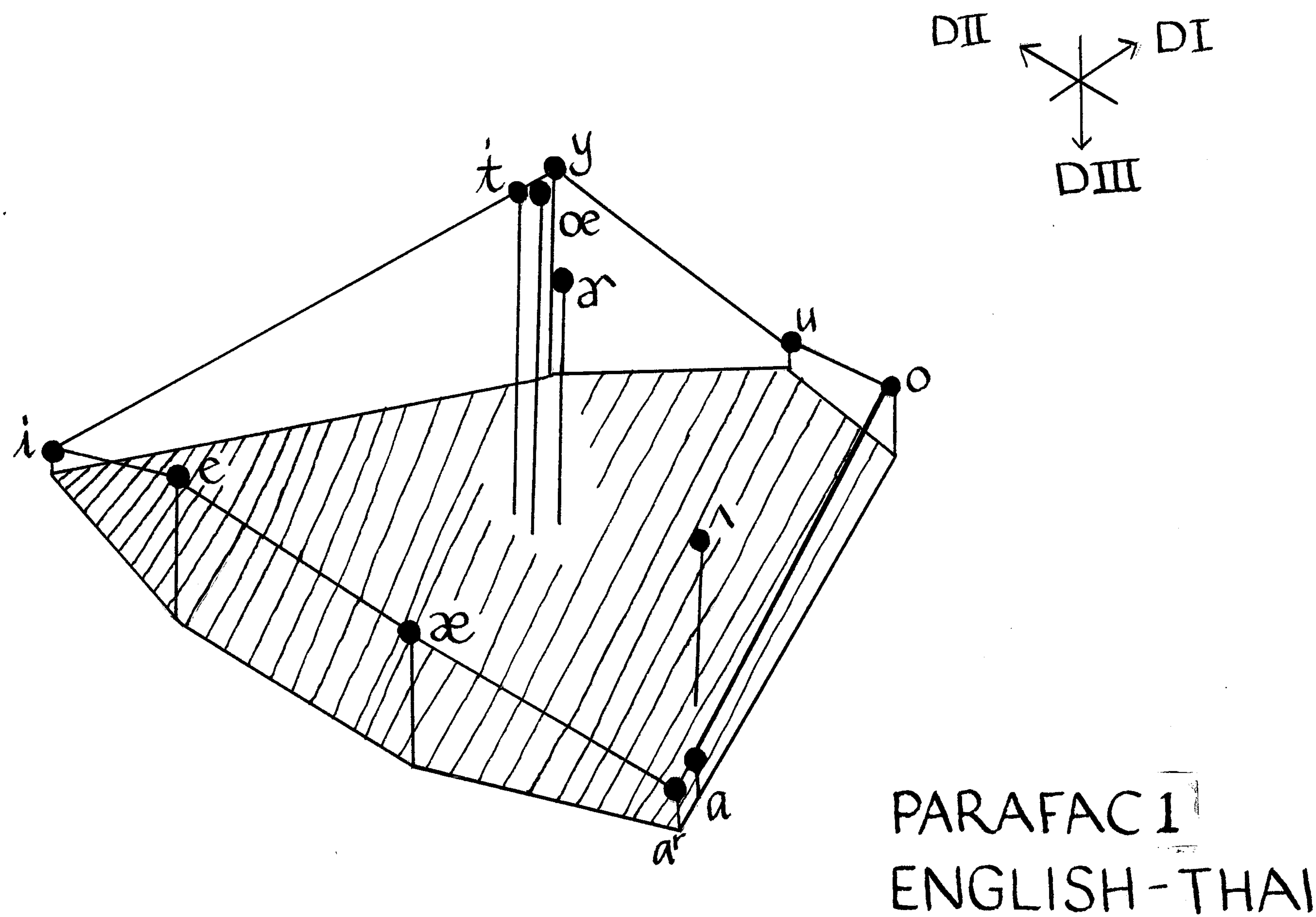


Figure 2. PARAFAC analysis for English and Thai subjects

In solutions of this type, analyses of lower dimensionality are also extremely difficult to interpret. 3) A configuration may appear which is understandable enough, but whose axes are difficult to interpret; a rotation of the space would result in new, more satisfying dimensions. But since the PARAFAC model uniquely determines the orientation of the axes, one is again obliged to interpret the axes in the original orientation.

As with any kind of analysis of experimental data, solutions that baffle rather than enlighten may be due to the use of inadequate or inappropriate data, [3] or to aspects of the experimental procedure, or to the investigator's inadequate understanding of the properties of the data. But if there is no justification for doubting the data or the method, and if there is simply no satisfying interpretation to be found for some PARAFAC solutions, we must reconsider the appropriateness of the analysis model. By not requiring that perceptual criteria be represented as directions in a Euclidean space, orthogonal to one another, we can provide more satisfying explanations for the troublesome situations listed above.

The first assumption to be discussed is that each derived dimension reflects a real perceptual dimension. Consider a language which uses two criteria to distinguish vowels. Calling each of those criteria a dimension in the perceptual space, we can place the vowels of that language on a two-dimensional surface. But why should that surface be a flat plane? To describe the surface as a flat plane implies that perceived distances between vowels conform to the Pythagorean theorem, and there is no evidence that vowel perception should show this property. Imagine that distances between vowels on the periphery of the surface are consistently underestimated, as compared to what the Pythagorean theorem would predict. The surface would then be slightly warped into a bowl-like shape. If perceived distances in vowel perception were distorted in this way, a model based on Euclidean geometry, such as PARAFAC or INDSCAL, could not be expected to provide a solution in two dimensions. It would require a third dimension to show how the configuration of points deviates from a flat plane. It would then follow that the third dimension value for each vowel would be predictable from the position of that vowel on the first two dimensions. To search for a third perceptual criterion which corresponded to the third dimension would be misguided.

Bearing this in mind, consider figure 2, which shows a PARAFAC analysis of the English and Thai listeners taken as a group. The shaded plane formed by the dimension D1 and D2 looks very much like a conventional vowel chart, ignoring for a moment the exact orientation of axes. Front vowels are to the left, back vowels are to the right. At the top of the shaded plane are the high vowels, and at the bottom are the low vowels. But what sense is there to be made of the third dimension D3? No linguistic

---

3 Conditions for data adequacy and appropriateness for PARAFAC are discussed in detail in Harshman 1970.

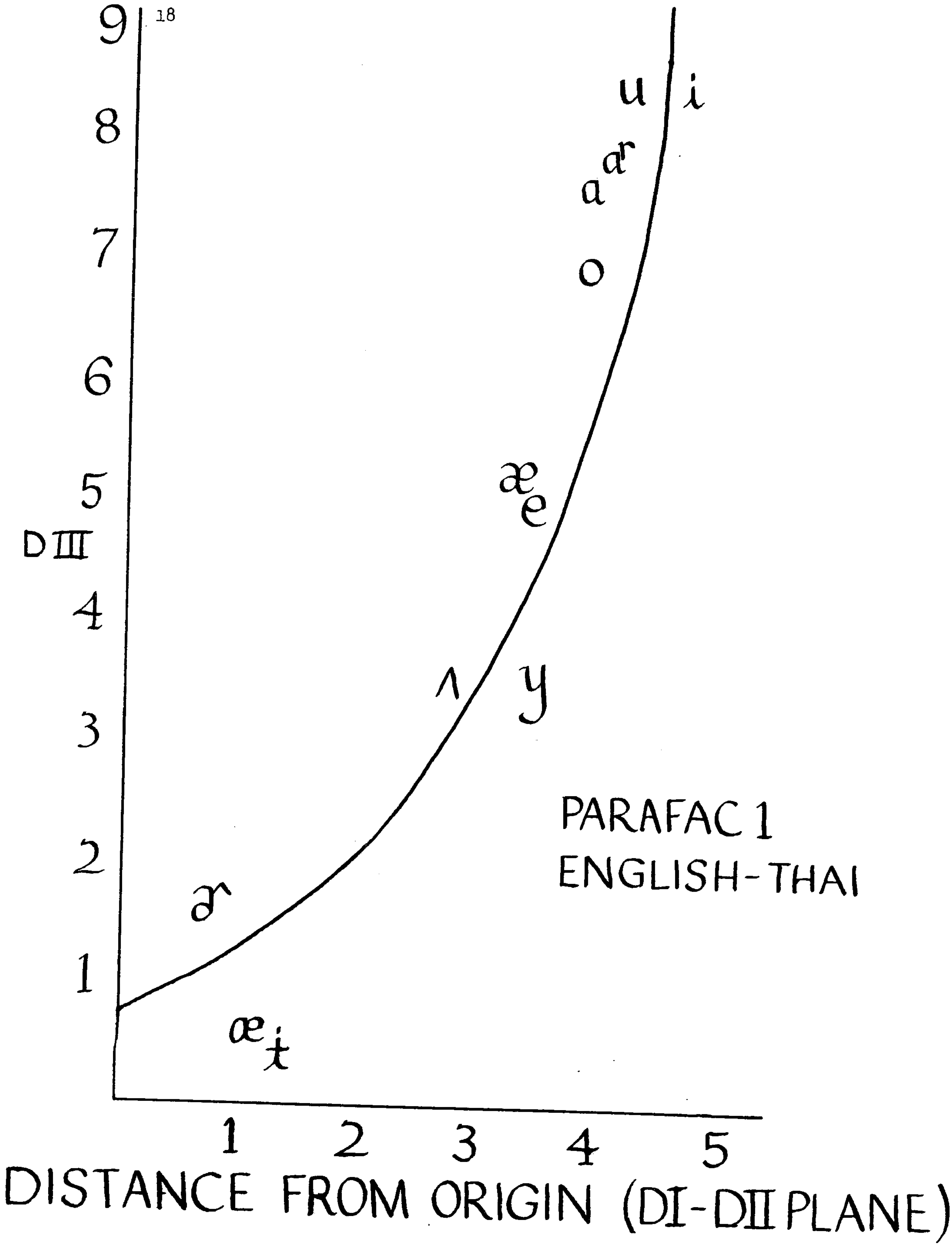


Figure 3. Scatter diagram for D3 values plotted against that function of D1 and D2 values which reflects the distance of a point from the center of the D1-D2 plane ($D1^2 + D2^2$).

or articulatory or acoustic scale groups the vowels [iuoa] together in opposition to [ɚœɨ]. Notice, however, that the position of a vowel on the third dimension is related to its position on the shaded plane. Vowels near the center of the plane have large D3 values, which vowels near the periphery have low D3 values.

This relationship can be seen more clearly on figure 3. Here the values along dimension D3 are plotted against the distance of each vowel from the center of the D1-D2 plane $(D1^2 + D2^2)^{1/2}$. By squaring the horizontal axis, that is, plotting D3 against $D1^2 + D2^2$, we obtain a correlation coefficient of .95, significant below the .002 level. Since the value of a vowel on dimension D3 is predictable from a function of its value on D1 and D2, the vowels are necessarily restricted to a curved surface embedded in the three-dimensional space. One might suspect, then, that the third dimension is not a third perceptual criterion, but an artefact of the model's restriction to linear dimensions. [4]

Before going on, we must stress the arbitrary nature of the particular function $D1^2 + D2^2$. Our only justification for choosing this function is that it correlates strikingly well with values on the third dimension. There may well be others, perhaps more complex functions of D1 and D2 that predict D3 values even better. Furthermore, the above description of the vowel surface as "bowl-like," with distances between vowels on the periphery of the surface being smaller than expected, was intended only as an introduction to the phenomenon we have observed in the interpretation of our data. A parabolic-type deviation from a flat plane is not at all the only possible kind of deviation; it is rather the only one we have found. Far from claiming to have found "the" non-linear relationship which characterizes the dimensions of vowel perception, we claim simply to have found a closer approximation to the "true" vowel space.

Nevertheless, the specific function we have described is useful in interpreting not only our English-Thai data but data from other analyses as well. In order to test this curved surface hypothesis for other analyses, including traditional analyses with no unique orientation of axes, a computer program was written capable of rotating a perceptual space of three or more dimensions to find the set of axes a,b,...y,z such that z-projections correlate most highly with values of the function $a^2 + b^2 \ldots + y^2$ for each vowel. Table 1 shows the results of this search procedure

---

4 Stated another way, the third dimension is a reflection of a non-linear relationship between two other dimensions. Notice, though, that this kind of non-linearity is entirely different from non-linearity as discussed by Kruskal (1964). Kruskal's non-metric multidimensional scaling technique adjusts for a non-linear function relating "true" distances to dissimilarity scores found in the raw data matrix. The resulting solution, however, is expressed in terms of Euclidean co-ordinates.

FIT OF NON-LINEARITY HYPOTHESIS

	Pearson's r	p
English-Thai	.97	<.002
German	.95	<.002
Turkish	.85	<.002
Swedish	.90	<.002
Swedish (Hanson 1967)	.94	<.002
Dutch (Pols et al 1969)	.78	<.004
English (Singh & Woods 1971)	.45	ns

Table 1. The correlation between one dimension and the function of two other dimensions ($x^2 + y^2$) for seven vowel perception studies.

as applied to seven different perceptual spaces. For six out of the seven analyses listed, there is a significant correlation between one dimension and the above function of the other dimensions.

The first four correlations listed in table 1 are based on our own experiments. The four perceptual spaces are different from one another, in some cases strikingly so, but in each case there is one artefact dimension. In the three-dimensional Thai-English analysis, the correlation is improved from the .95 mentioned above to .97 by choosing a direction a few degrees away from the third (non-linearity) dimension. For the analysis of German, when all seven listeners are taken together, there are four dimensions: three as described earlier, and an extra dimension which reflects the curvature of the space. The Turkish solution, also in four dimensions, is similar to the German one. The Swedish data proved unusually difficult to analyze, as will be seen shortly, but the very reasonable solution finally obtained using PARAFAC2 (discussed below) showed a .90 correlation coefficient, supporting our curved space hypothesis.

The last three items on the list represent our re-analysis of published summed distance matrices. It is interesting to note that Hanson (1967) presented a three dimensional solution for Swedish vowels, one dimension of which he labelled as a "perceptual contrast" factor but could not explain. This dimension can now be re-interpreted as an artefact of the analysis model. Hanson's third dimension correlates at .88 with the non-linearity dimensions which we located in his data. The confirmation of our non-linearity hypotheses using a re-analysis of the data of Pols et al (1969) is based on a correlation somewhat lower than the previous five solutions, but is nevertheless highly significant. For the sake of completeness, we include the re-analysis of the Singh and Woods (1971) data on American vowels, for which we could find no direction giving a non-linear correlation of this type. The first six of the analyses in table 1 then, provide strong support for the hypothesis that some aspects of vowel perception are better described by a curved two-dimensional space than by a linear (Euclidean) three-dimensional one.[5]

---

5 We have already mentioned the fact that there may be other non-linear relationships between dimensions in perceptual space besides the one described here. But whatever relationships one discovers, there is a further point to be made in their interpretation. The following discussion is worded in terms of the function $D1^2 + D2^2$, but will apply no matter what function is chosen.

We have tacitly assumed in the discussion that the curved space is a positively curved surface embedded in a Euclidean space. But in this representation, distances between vowels are measured as straight lines through the three-dimensional space, not as curved lines along the embedded vowel surface. Distances between vowels are therefore measured through a region where(according to our non-Euclidian interpretation) no vowels can be placed. What is required is a surface *along which* the experimental distances are measured. The distance function for that surface is easy to derive. We know that the distance between two vowels

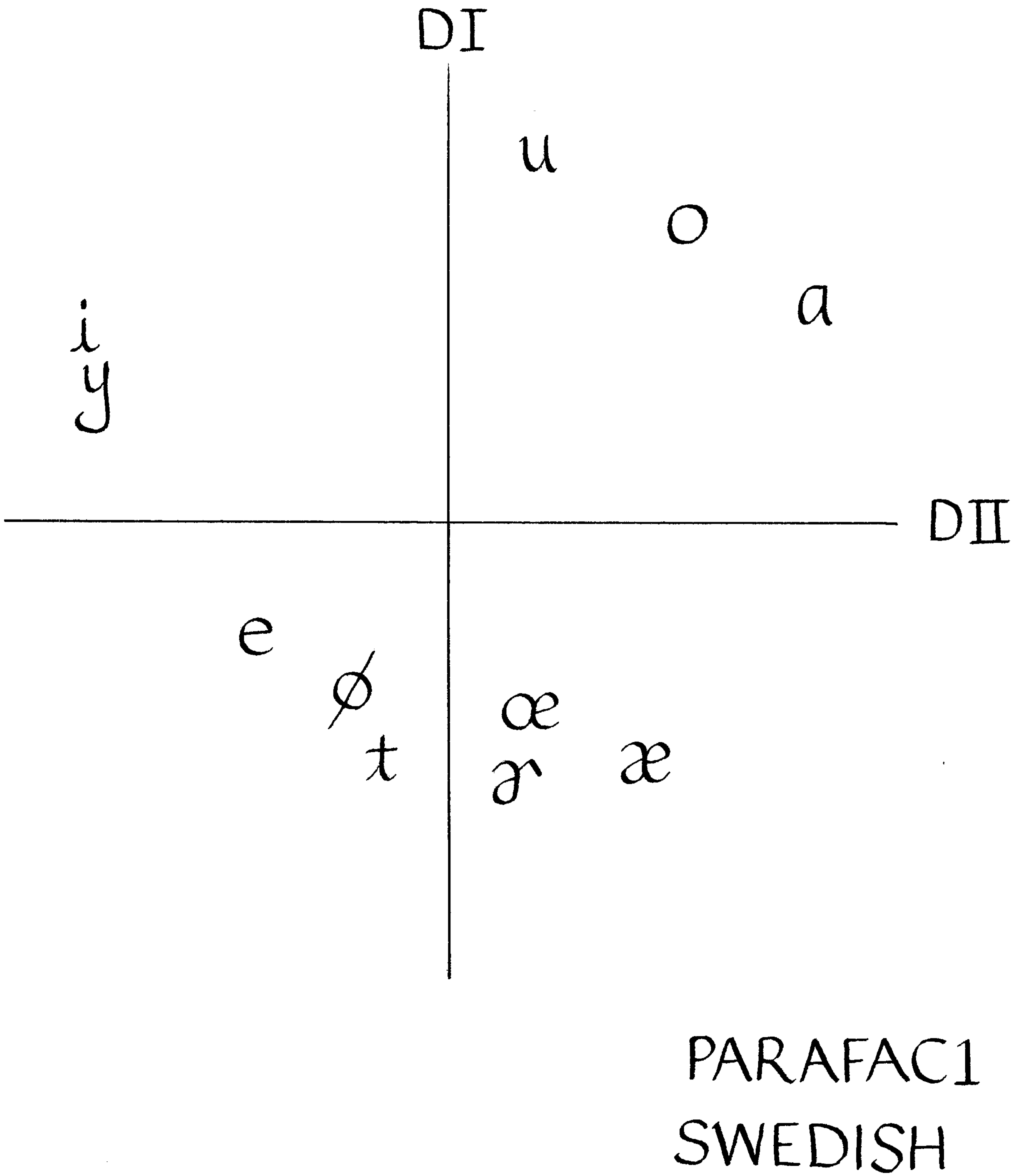


Figure 4. PARAFAC1 analysis for seven Swedish listeners, [ʉ] vowel excluded.

The second assumption in the PARAFAC-INDSCAL model is that dimensions in the perceptual space are orthogonal to one another. But then perhaps there are such things as perceptual dimensions that lie at oblique angles to each other. If there were, a PARAFAC solution could only give an orthogonal approximation to the underlying dimensions; it could not give an adequate, "explanatory" picture of the data. In order to discover exactly how PARAFAC treats data based on oblique dimensions, synthetic data was constructed using oblique axes, and subjected to a series of PARAFAC analysis. The solutions obtained had the following three properties; first, unique solutions were obtained for higher dimensionalities than the number of dimensions actually built into the data. For a two-dimensional problem, PARAFAC returned unique solutions in at least three and four dimensions. Second, the configuration of points in what should have been the correct solution was slightly distorted. Third, the axes recovered by PARAFAC were not oriented in the right directions.

In our analyses of real data, we have seen just these characteristic problems. The best example is our analysis of Swedish. PARAFAC recovered unique solutions for seven dimensions, which is a totally implausible number of perceptual criteria for the vowel stimuli presented to the listeners. The two-and three-dimensional solutions were largely uninterpretable, and as the number of dimensions increased, the resulting spaces became even less interpretable. When the traditionally troublesome and therefore interesting [ʉ] vowel was removed from the data, only three unique dimensions were recovered.

Two dimensions of the three-dimensional Swedish solution (minus the [ʉ]) are shown in Figure 4. Here the configuration of vowels is a reasonable one. Back vowels cluster at the top right of the figure, high front

---

_a_, _b_ in three dimensions _x_, _y_, _z_ is

$$d_{a,b} = \left\{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2\right\}^{1/2}$$

from the Pythagorean theorem. We have also shown empirically that _z_ dimension values can be predicted from _x_ and _y_ values. The equation for this relationship is

$$z = f(x,y) = k(x^2 + y^2) + c$$

where _k_ and _c_ are constants determined empirically by a regression of D3 values against $D1^2 + D2^2$. Our new distance function in two dimensions, then, is

$$d_{a,b} = \left\{(x_a - x_b)^2 + (y_a - y_b)^2 + [f(x_a,y_a) - f(x_b,y_b)]^2\right\}^{1/2}$$

However, describing the characteristics of the surface for which this distance function is valid is not an easy matter and we will not discuss it further in this article. We do feel it is important to point out that the surface embedded in a Euclidean space which contains the vowels is not the same as the "true" non-Euclidian curved vowel surface along which distances are to be measured. This fact, of course, does not alter any of the arguments or conclusions presented.

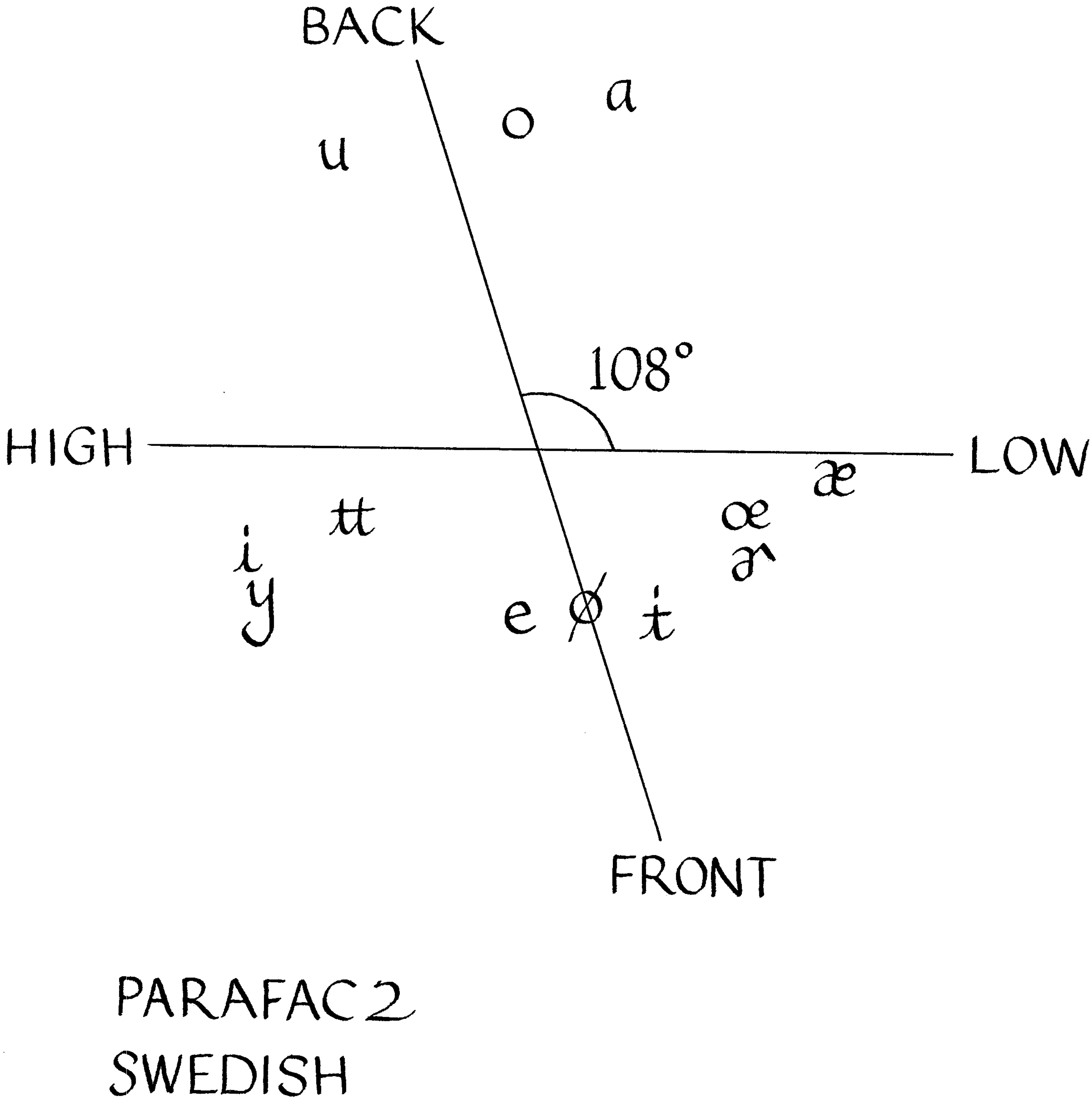


Figure 5. PARAFAC2 analysis for seven Swedish listeners, [ʉ] vowel included.

vowels are at the bottom left, and so on. But these two PARAFAC axes are not interpretable in any meaningful way. In light of our experiments with synthetic data, it is plausible to suggest that there is an oblique orientation of these two axes in the Swedish perceptual space. (The third PARAFAC dimension was an interpretable rounding dimension, suggesting that it is not oblique to either of the other two axes and therefore was correctly recovered.) [6]

An extension of the PARAFAC model which is not limited to orthogonal axes has been developed by Harshman (1972). [7] This new model, PARAFAC2, was applied to our Swedish data. A number of problems were encountered with difficult to interpret solutions, which had poorer fit. But as we became more familiar with the program, we were able to obtain a "good" solution. In this good solution the computed fit of the data to the model was better than the other PARAFAC2 solution, and better than the original PARAFAC analysis. The two previously troublesome dimensions of the three-dimensional solution are shown in Figure 5. With PARAFAC2, there is no difficulty caused by including the [ʉ] vowel; we did not need to reduce the data to arrive at a reasonable solution. Most important of all, the oblique axes of the PARAFAC2 solution are much more directly interpretable. The verticle axis in Figure 5 unambiguously

---

6 We must be cautious, however, in claiming that there are only three dimensions for Swedish. We have found that if there is an insufficient number of subjects used in an analysis, a solution with too few dimensions may be recovered as the correct one, according to the uniqueness criterion. For example, a three-dimensional solution was obtained when five German listeners were taken together, and four dimensions were obtained when all seven German listeners were used. The three-dimensional solution was interpretable in terms of the linguistic scales of front/back, high/low, and round/unrounded. Yet when the non-linearity search program was applied to this solution, the correlation coefficient obtained was .90 (not shown in Table 1), suggesting that a direction in the space could be expressed as a function of two other directions. The point here is that the artefact dimension and the criterion of rounding seemed to collapse with the two other dimensions into a slightly distorted, lower-dimension solution when too few subjects were used.

This may well be the case for Swedish. The non-linearity correlation coefficient for the eleven vowel solution with PARAFAC1 was .86 (also not in Table 1), strongly suggesting that there are non-linearly related directions in the Swedish perceptual space. It is possible,then, that for the PARAFAC1 Swedish solution, as for the German, the dimension which results from the interaction of two other dimensions may not have appeared independently; it might have done so if a larger number of listeners had been used.

7 The restriction to orthogonal axes with PARAFAC1 is true only for multidimensional scaling or the analysis of covariance matrices. For general factor-analytic applications, involving analysis of data matrices rather than cross products or scalar products, there is no constraint on the way factor loading patterns can be oriented with respect to each other. For more information on this point, see Harshman (1970) and Harshman (1972).

sets back vowels off from all the others. The horizontal axis orders vowels according to the parameter called tongue height. The rounding dimension was recovered as before, and was, as predicted, found to be orthogonal to the other two axes.[8] (See table 2 for loadings.)

Since the current version of PARAFAC2 is expensive to operate(and sometimes gives results subject to difficulties not yet fully understood) we have not made complete PARAFAC2 analyses of all our data. But preliminary analyses of the English and Thai data using PARAFAC2 have shown promise of the same kinds of improvement over the original PARAFAC results as we have seen for the analysis of Swedish. There are also hints of some interesting and surprising dimensions.

Having demonstrated the usefulness of allowing oblique solutions, we are beginning to explore hypotheses as to what psychological meaning obliqueness of perceptual axes might have. One certain effect that oblique axes have on perceptual spaces is a compression of distances in two quadrants and an expansion of distances in the other two quadrants. In Figure 5, for example, the distance between the vowels [a] and [o] is greater than it would be if the axes were orthogonal. The same holds true for vowels in quadrant three, and the reverse applies to quadrants two and four.[9] But what could cause the listeners' perceptual space to be altered in this way?

At the present, we are working with two interpretations of obliqueness, which we call the *compensation* and *expectation* hypotheses. According to the expectation hypothesis, the subject expects some values to co-occur based on long past experience; if high vowels, say, were more likely to be front vowels, then the listener would come to automatically expect a larger value on the front-back dimension, given a large value on the high-low dimension. The relative predictability of one criterion based on the other causes the listener to attach less importance to high and front values as separate criteria when they occur in the expected direction.

---

8 Again, an analysis of more listeners might affect the dimensionality of the result, adding a separate fourth dimension which reflects a non-linear relation between the front/back and high/low dimensions.

9 It must be pointed out that the values for the angles are the least stable part of a PARAFAC2 analysis, and we make no strong claims for the angle in Figure 5. The PARAFAC2 algorithm is able to establish a set of vowel loadings relatively quickly, but reaches an optimum angle much more slowly. The present expense of operating PARAFAC2 has prevented us from continuing this solution a sufficient amount further to know whether the angles have stabilized. Nevertheless, it was only with PARAFAC2 that a satisfactory set of dimensions could be obtained at all, suggesting that two underlying dimensions for Swedish are indeed not orthogonal, although the angle between them may not yet be precisely known.

This is reflected in a compression of distances in that quadrant.

If the expectation hypothesis were the proper explanation, we would expect to see, on Figure 5, an acute angle in the high, front quadrant three (and quadrant one) and an obtuse angle for quadrants two and four. This is just the reverse of the way the axes actually lie in Figure 5. In an attempt to explain the results observed in this Swedish example, we propose the compensation hypothesis. With this point of view, we observe that the listener is required to make somewhat finer discriminations for the high front vowels, since there are more vowels in that region to be identified. He therefore has developed a keener sense of distance in that region and perceives it as a somewhat larger region than the adjacent quadrants. The high back quadrant, for example, contains only one vowel, the [u], and it is much less important for the listener to be able to make discriminations in that region of the space.

Our interpretation of obliqueness of axes is plausible, but of course tentative. Although we have not decided upon any single interpretation of obliqueness of axes, it should nevertheless be clear that oblique solutions can provide more meaningful dimensions than solutions restricted to orthogonal axes. Also, simpler and more explanatory solutions may be found if we allow the possibility of non-linear, or curved perceptual dimensions. The mathematics of Euclidean and Cartesian geometry is extremely simple and convenient, but our results suggest that it is insufficient to describe the psychological structure of vowel perception.

References

Carroll, J.D. and J.J. Chang (1970), "Analysis of Individual differences in Multidimensional Scaling via an n-way Generalization of Echert-Young Decomposition," *Psychometrika* 35, 283-319

Hanson, (1967), "Dimensions in speech sound perception: An experimental study of vowel perception," *Ericsson Technics* 23, 3-175

Harshman, Richard (1970), "Foundations of PARAFAC procedure: Models & conditions for an 'explanatory' multi-modal factor analysis," *Working Papers in Phonetics* 16, 84p.

Harshman, Richard (1972a), "PARAFAC2: Extensions of a procedure for explanatory factor analysis and multidimensional scaling", *Working Papers in Phonetics* 22. (abstract)

Harshman, Richard (1972b), "PARAFAC2: Mathematical and technical notes", *Working Papers in Phonetics* 22

Kruskal, J. (1964), "Multidimensional Scaling by Optimizing Goodness of Fit to a Non-Metric Hypothesis", *Psychometrika* 29, 1-27.

Pols, L.C.W., van der Kamp, L.J.Th., and Plomp, R. (1969), "Perceptual & physical space of vowel sounds," *J. Acoust. Soc. Amer.* 46, 458-467

Singh, S. and G. Woods (1971), "Perceptual Structure of American English Vowels", *J. Acoust. Soc. Amer.* 49, 1861-1866.

Terbeek, D., and R. Harshman (1971), "Cross-Language Differences in the Perception of Natural Vowel Sounds", *Working Papers in Phonetics* 19, pp. 26-38.

PARAFAC2 SWEDISH SOLUTION LOADINGS

H/L = High/Low, F/B = Front/Back, R/U = Round/Unrounded. For a discussion of indeterminacy in the signs of person loadings, see Harshman (1972b).

I. Vowel loadings

Vowel	H/L	F/B	R/U
ʉ	-3.010	-1.140	-3.115
ɨ	1.017	-2.521	-1.704
y	-4.705	-1.890	0.547
ø	0.081	-2.384	-1.663
ə˞	2.711	-1.559	-3.367
æ	4.286	-0.206	3.334
ɶ	2.667	-0.906	-2.544
u	-1.582	4.274	-2.815
o	1.183	4.951	-0.400
a	2.938	5.372	2.919
e	-0.950	-2.304	4.311
i	-4.636	-1.687	4.497

II. Cosines of angles between dimensions

	H/L	F/B	R/U
H/L	1.000	-0.319	-0.038
F/B	-0.319	1.000	0.072
R/U	-0.038	0.072	1.000

III. Person loadings (relative salience of each dimension for each person.)

Person	H/L	F/B	R/U
1	1.288	1.173	1.920
2	1.107	1.412	1.392
3	-2.249	-2.274	-2.118
4	1.734	1.325	1.252
5	1.312	1.164	2.112
6	-1.597	-2.123	-2.236
7	-1.630	1.545	-1.654

Table 2. Factor loadings for PARAFAC2 Swedish Solution

PARAFAC2: Mathematical and Technical Notes

Richard A. Harshman*

0. Abstract

The mathematical model of PARAFAC is reviewed, and an examination is made of its application to cross-product matrices (e.g. covariance matrices, scalar product matrices, etc.). It is shown that PARAFAC1 can correctly describe both orthogonal (uncorrelated) and oblique (correlated) factors in system-variation *data* matrices, but that it can only correctly describe orthogonal factors when applied to the *cross-product* matrices computed from that data.

A similar examination is made of the INDSCAL-PARAFAC1 model for three-mode multidimensional scaling. It is shown that this model is restricted to descriptions of stimuli relationships in terms of orthogonal perceptual dimensions, (whereas traditional two-mode multidimensional scaling has no such restriction).

A three-mode model is developed to deal specifically with sets of cross-product matrices. This model, called PARAFAC2, can describe cross-product matrices in terms of orthogonal or oblique factors, whichever best fits the data. Yet it retains PARAFAC1's highly desirable characteristics of providing unique "explanatory" solutions. Along with factor loadings on variables, it provides factor correlations, and loadings for each occasion of measurement. For multidimensional scaling, PARAFAC2 will recover the projections of the stimuli on oblique (or orthogonal) dimensions, give the angles between these dimensions, and provide a set of dimension weights for each person.

The derivation and interpretation of the PARAFAC2 model is discussed, and a precise definition is given of the type of uniqueness which it provides. No formal proof of this uniqueness has as yet been discovered, but the uniqueness has been demonstrated empirically by computer analysis of synthetic data.

An additional important advantage of PARAFAC2 for factor analytic applications is its greater generality: it is not restricted to analysis of system-variation data. This is shown by re-deriving the PARAFAC2 description of cross-products without invoking the restricted system-variation model for data from which the cross products are computed. Instead, a more general model of the data is used. This model applies to almost all three mode cases which might be approached factor-analytically. System-variation is one special case of this more general model.

Current work on algorithms for fitting the PARAFAC2 model to data is reviewed. A "hybrid" approach is under development which combines general optimization techniques with estimation procedures based on the mathematical

---

*The author wishes to express his indebtedness to Robert Jennrich, Joseph Kruskal, and Dale Terbeek, for the important part that they played in developing the ideas underlying PARAFAC2.

properties of the PARAFAC2 model.

A method for circumventing the "communalities" problem is proposed. It consists of a technique for ignoring the diagonal values of the cross-product matrices and estimating the parameters of PARAFAC2 by optimizing the least square fit to the off-diagonal elements.

The differences between analysis of covariance and correlation matrices is made explicit by developing the appropriate modifications of the PARAFAC2 model which would be necessary to analyse correlations.

PARAFAC2 is compared with more general models which lack unique solutions, but which allow different correlations between a given pair of factors from one cross-product matrix to the next. The possible use of a routine such as Carroll's IDIOSCAL program in conjunction with PARAFAC2 is discussed.

1. The factor model of PARAFAC1 (PARAFAC)

Conventional factor analysis describes a two-way set of data in terms of a small number of underlying factors. Let us think of this two-way set of data as a set of values for m persons on n tests. x_{jk} is the value for the jth test on the kth person. Factor analysis writes

(1a) $$x_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \ldots + a_{jl}b_{lk} + E_{jk}$$

where a_{j1} is the loading of the first factor on test j and b_{1k} is the loading of the first factor on person k, and such products are formed for the l factors presumed to underlie the data. E_{jk} is an error term, which is necessary as the model should not be expected to fit real data exactly. (E_{jk} is often written U_{jk} and called "uniqueness" term. It is thought of as consisting of components unique to that person and that test, plus error.)

Let us neglect error terms until we consider estimation procedures. Using summation notation, we can write

(1b) $$x_{jk} = \sum_{r=1}^{l} a_{jr}b_{rk} \quad .$$

In matrix notation, we can consider A, an n by l matrix giving the loadings of n tests on l factors; and B an m by l matrix giving the loadings of m persons on l factors. We can then write the classical factor model (neglecting error) as

(1c) $$X = AB'$$

where X is an n by m matrix of values for n tests obtained by m persons.

The PARAFAC (hereafter called PARAFAC1) generalization of this model (Harshman, 1970) can be expressed by considering a number of two-way data matrices simultaneously (thus obtaining a three-way data set). The values for the factor loadings of the tests (or alternatively, the persons, it does not matter which) are altered from one occasion to the next by a proportional change as follows:

(1d) $$x_{jk}^{(i)} = a_{j1}c_1^{(i)}b_{1k} + a_{j2}c_2^{(i)}b_{2k} + \ldots + a_{jl}c_l^{(i)}b_{lk}$$

where $x_{jk}^{(i)}$ represents the *ith* two-way matrix x_{jk}. We will here consider it the matrix for the *ith* "occasion", just as we considered x_{jk} to be values for "tests" obtained by "persons", although all these names are only for convenience in visualizing a particular application of the model.

In summation notation we can write

(1e) $$x_{jk}^{(i)} = \sum_{r=1}^{l} a_{jr}c_r^{(i)}b_{rk}$$

and in matrix notation this becomes

(1f) $$X_i = AD_iB'$$

where X_i is the two-way matrix x_{jk} for the *ith* occasion, A and B are the matrices of factor loadings for tests and persons, as before, and D_i is an l by l diagonal matrix, whose diagonal cells give proportional changes in the loadings for the factors on the *ith* occasion.*

In this model of data, there are no restrictions on the patterns of loadings which occur in A and B. In particular, no assumptions are made that the columns of A or B are orthogonal to one another.

If there are a sufficient number of different X_i in a given set of data matrices, the decomposition described by (1f) is unique for a small enough number of factors (l), except for trivial differences which do not affect interpretation. Conditions of data adequacy and of uniqueness of the extracted factors, have been described in Harshman, 1970 (pp. 20-23, 35-44, 61-62).

Because of this uniqueness, and the minimal assumptions under which it is obtained, an argument can be made that factors extracted in this way from "system-variation" data (Harshman 1970, pp. 20-23) have a greater likelihood of explanatory validity than those selected by traditional factor analysis using some rotation principle such as Varimax or simple structure. (For the argument, see Harshman 1970, pp. 1-26.)

2. Representation of summed-cross-product matrices

Now let us consider the problem of representing, in terms of the factors of X, values derived from X by taking sums of cross-products. If the means of the columns of X are zero, the summed-cross-products for two variables will be proportional to their covariances. If, in addition, the variances of the columns of X are unity, then the cross-products for two variables will be equal to the correlation coefficient for those two variables.

For tests j and j', the covariance across the m persons is

(2a) $$c_{jj'} = \frac{1}{m}\sum_{k=1}^{m}(x_{jk})(x_{j'k}) \quad . \qquad (\text{if } \bar{x}_j=\bar{x}_{j'}=0)$$

*In other words, if $d_{rs}^{(i)}$ is an entry of D_i, then $d_{rs}^{(i)} = 0$ if $r \neq s$, and $d_{rs}^{(i)} = c_r^{(i)}$ if $r = s$ where $c_r^{(i)}$ is used as in equation (1e).

This is simply $1/m$ times our general summed-cross-product. The full matrix of summed-cross-products C, can be represented in matrix notation by

(2b) $C = XX'$

To represent these summed-cross-products (e.g. unscaled covariances) in terms of factors, rather than data values, we can substitute our theoretical factor model of the data matrix (from equation (1f)) into our description of the summed-cross-products (2b). This will give us, for the cross-product matrix derived from the *ith* data matrix

(2c) $C_i = (AD_iB')(AD_iB')'$

Since the transpose of (AD_iB') is the transpose of the components taken in reverse order we can rewrite $(AD_iB')' = B''D_i'A'$. Further, since $B'' = B$ and $D_i' = D_i$ (since D_i is diagonal), we can rewrite (2c) as

(2d) $C_i = (AD_iB')(BD_iA')$

regrouping gives

(2e) $C_i = AD_i(B'B)D_iA'$

or, combining terms

(2f) $C_i = AD_iWD_iA'$ (where $W = B'B$).

This describes our *ith* summed-cross-product or covariance matrix in terms of A, our n by l matrix of loadings of the n tests on the l factors; D_i, our diagonal matrix giving the weights of the l factors on the *ith* occasion, and W an l by l matrix related to the *factor* cross-products. If the columns of B have zero means, then $(1/l)W$ is the matrix of covariances of the factors. If the cells of W are scaled so that the diagonals are equal to one*, then w_{rs}, a cell of W after appropriate scaling, gives the correlation between factor r and factor s. This is interpretable geometrically as the cosine of the angle between factor r and s in the space spanned by the factors.

In summation notation, this model (2f) is written

(2g) $$c_{jk}^{(i)} = \sum_{s=1}^{l} \sum_{r=1}^{l} a_{jr} d_{rr}^{(i)} w_{rs} d_{ss}^{(i)} a_{ks} .$$

3. <u>Orthogonal factors</u>

Now if the factors are uncorrelated with one another in their pattern of loadings across people, i.e. if the columns of B are uncorrelated (have average cross-products of zero) then

(3a) $W = B'B = \Delta$

*(i.e. w_{rs} is divided by $\sqrt{(w_{rr})(w_{ss})}$, and the inverse scaling is made on all the D matrices to compensate)

where Δ is some diagonal matrix. This gives zero cosines for the angles between different factors, and thus the factors are said to be orthogonal. Now in this case, the summed cross-products can be represented

(3b) $\quad C_i = AD_i \Delta D_i A'$.

This can be rewritten

(3c) $\quad C_i = A\tilde{D}_i A'$

where $\tilde{d}_r^i$, the *rth* diagonal element of $\tilde{D}_i$, is equal to $(d_r^i)^2(k_r)$, where k_r is the *rth* diagonal element of Δ, and relates to the "size" of factor r (i.e. the sum of the squared loadings of factor r in B). If the mean loading for factor r were zero, k_r would be proportional to the variance of factor r in B.

Suppose we took a set of summed-cross-product or covariance* matrices C_i and analysed them by the PARAFAC1 procedure. Then we would be decomposing them according to the model of (1f), namely

(3d) $\quad C_i = \overset{*}{A}\overset{*}{D}_i\overset{*}{B}'$.

Now this would yield a perfect fit and a "unique" solution (see Harshman 1970), when

(i) model (1f), or the more general model of section 7, is appropriate for the data from which the C_i were computed;
(ii) we had guessed the number of factors correctly (guessed the correct value for l);
(iii) $B'B = \Delta$, that is, the factors were uncorrelated or "orthogonal" across B in the original data from which the cross-products were computed.

Further, we would discover the $\overset{*}{B}$ was proportional to $\overset{*}{A}$. Thus the solution could be represented (by choosing the right $\overset{*}{D}_i$) as

(3e) $\quad C_i = \overset{*}{A}\overset{*}{D}_i\overset{*}{A}'$.

By comparing the PARAFAC1 solution in this form with our equation (3c), we can interpret it as follows. $\overset{*}{A}$ would equal A, the matrix of loadings of the tests on factors which underlie the original data from which the covariances were computed. $\overset{*}{D}_i$ would equal $(D_i)^2\Delta$, the squares of the weights of the factors for the *ith* occasion, multiplied by coefficients proportional to the "size" of the factors in B. We note that $\overset{*}{D}_i$ should therefore have all elements greater than or equal to zero. If PARAFAC gave a negative value in $\overset{*}{D}_i$ it would indicate a failure of the data to correspond to the model.

4. Generalized model for oblique factors

We have seen that PARAFAC1 provides an adequate model for analysis of covariance or other summed-cross-product matrices only when the factors are orthogonal in the data. This restricts the application of the model, since

*Covariance matrices can be analysed by PARAFAC2 only if the means of the variables do not change appreciably from one C_i matrix to the next. If they do, cross products or average cross products must be used.

it is not usually possible to know whether or not this would be the case for any given set of covariance matrices.

Therefore, we adopt the more general expression derived in (2f) as a general model for three-mode factor analysis of matrices composed of summed-cross-products:

(4a) $C_i = AD_iWD_iA'$

We shall call this model PARAFAC2. This model describes a set of C_i cross-product matrices in terms of a common set of factors A, and a common set of angles between the factors, W. The occasions differ only in the weights given the factors, described for the *ith* occasion (the *ith* matrix of summed-cross-products) by D_i.

5. Derivation of PARAFAC2 for scalar product matrices derived from distance matrices

In multidimensional scaling, and in particular in the Carroll and Chang three-mode model (Carroll and Chang 1970) we start with a set of matrices giving the distances between stimuli. These are then converted to the scalar products of vectors by application of a trigonometric relationship which holds for oblique or orthogonal coordinates (see Torgerson 1958, pp. 255, 258). The resulting scalar products represent something similar to covariances between the stimuli. Each scalar product gives the "common strength" of two stimuli by multiplying the length of the first stimulus vector from some origin by the size of the projection of the second stimulus vector upon the first.

Now scalar products can also be represented by the sum of products of the projections of the two stimuli onto a set of *orthogonal* axes as follows: Let A be an n by l matrix of projections of n stimuli onto l orthogonal axes, then

(5a) $S = AA'$,

where S is an n by n matrix of scalar products of the stimulus vectors.

Now Carroll and Chang generalized this model to a set of parallel scalar product matrices for the same n stimuli, allowing individual axes to be expanded or contracted in each set (Carroll and Chang 1970, pp. 284-5). This gives

(5b) $S_i = (AD_i)(AD_i)'$

which can be rewritten

(5c) $S_i = AD_i^2A'$.

This is obviously closely related to our model (3e) for summed-cross-products with orthogonal factors. Now the scalar products as derived from the distances between stimuli can only be represented by $S_i = AD_i^2A'$ when the axes onto which the stimuli are projected are orthogonal.

The problem with oblique axes does not arise in traditional two-mode multidimensional scaling. Since the orthogonal axes extracted by two-mode scaling are arbitrary, they can always be rotated, without loss of fit, into a "true" oblique position. But the INDSCAL - PARAFAC1 unique three-mode solution will be the best fitting pair of orthogonal axes "common to all persons." If the true axes are oblique, experiments with synthetic data show that two different distortions will occur. Not only will the orientations of the INDSCAL - PARAFAC1 axes be incorrect, but the recovered stimulus configuration will be distorted as well. (But it is interesting to note that the distorted stimulus configuration with the orthogonal axes will predict the data almost as well as the correct configuration with oblique axes.)

Let us derive a model for three-mode multidimensional scaling which allows oblique, as well as orthogonal axes. We can express the scalar products in terms of *oblique* axis projections of the stimuli as follows.

> Let A be an n by l matrix giving the projections of n stimuli onto a set of l oblique (or orthogonal) axes, the "underlying perceptual dimensions".
>
> Let D_i be a diagonal l by l matrix giving the expansions or contractions of the axes of A for the *ith* person.
>
> Let T be an l by l matrix describing the projections of the axes of A onto a set of orthogonal axes (so that AT describes the projections of the stimuli onto a set of orthogonal axes). Then:

(5d) $S_i = (AD_iT)(AD_iT)'$

(5e) $S_i = (AD_iT)(T'D_iA')$

(5f) $S_i = AD_i(TT')D_iA'$

(5g) $S_i = AD_iWD_iA'$

Thus we get the same model as developed earlier for summed-cross-products, with the same interpretation for all the constituent matrices.

A gives the stimulus projections, D_i the relative scale or importance of the dimensions for the *ith* person, and W the angles between the dimensions.

6. Uniqueness of PARAFAC2 Solutions

No proof of the uniqueness or of the conditions for non-uniqueness has yet been discovered for PARAFAC2. Nonetheless, computer experiments with PARAFAC2 and synthetic data have established empirically that it will give unique solutions, if the number of factors extracted is not greater than the number "actually in the data" (used to create the synthetic data), and if there are a sufficient number of independent C_i matrices. When too many factors are extracted however, the solution is not unique. The behavior of the model and the conditions of uniqueness seem to be similar to those of PARAFAC1, which are described in detail in Harshman (1970).

The term "unique" needs some clarification. There is a trivial non-uniqueness of scale for PARAFAC2, in that all the loadings on a given factor can be doubled or otherwise changed in scale, as long as a compensatory inverse scaling is performed on all the person weights for that factor (or, on the W matrix). Such changes do not affect the *pattern* of loadings across measures or across persons, and since this *pattern* is what is interpreted, such changes do not alter the interpretation that would be given to a particular factor. The columnar order of the factors is also arbitrary. Factor I on a given analysis may come out as factor III on a different analysis, but the same total set of factors will be found to be present in both cases.

Mathematically, these two trivial types of indeterminacy can be expressed by the following equalities. Let S be an l by l diagonal matrix changing the scale of a set of l factors. Let $\overset{*}{S}$ be another such diagonal scale-changing matrix. Let P be an l by l permutation matrix changing the order of the l factors. We can see that if

(6a) $$C_i = AD_iWD_iA'$$

then we can insert l by l identity matrices into this equation without altering it, as follows

(6b) $$C_i = A(I)D_i(I)W(I)D_i(I)A'$$

now since $(SS^{-1})=I$, and $(\overset{*}{S}\overset{*}{S}^{-1})=I$, we can substitute

(6c) $$C_i = A(SS^{-1})D_i(\overset{*}{S}\overset{*}{S}^{-1})W(\overset{*}{S}^{-1}\overset{*}{S})D_i(S^{-1}S)A' \quad .$$

In a similar fashion, since $(PP^{-1})=(I)$, we can write

(6d) $$C_i = A(PP^{-1})D_i(PP^{-1})W(PP^{-1})D_i(PP^{-1})A'$$

combining these substitutions, we can write

(6e) $$C_i = A(P[SS^{-1}]P^{-1})D_i(P(\overset{*}{S}\overset{*}{S}^{-1})P^{-1})W(P[\overset{*}{S}^{-1}\overset{*}{S}]P^{-1})D_i(P[S^{-1}S]P^{-1})A' \quad .$$

By regrouping parentheses we can display the indeterminacies of both scale and permutation that the A, D and W matrices are subject to, and the compensatory inverse transformations that are necessary to preserve the equality:

(6f) $$C_i = (APS)(S^{-1}P^{-1}D_iP\overset{*}{S})(\overset{*}{S}^{-1}P^{-1}WP\overset{*}{S}^{-1})(\overset{*}{S}P^{-1}D_iPS^{-1})(SP^{-1}\ A')$$

However, there are other special indeterminacies of the *signs* in the D_i matrices which aren't as obvious. These indeterminacies can be expressed by the following rule:

(Rule 6a) The sign of a single person's loading on any given factor can be reversed, provided that his signs

> are also reversed for loadings on all factors that are oblique to (i.e. have a non-zero cross-product with) the given factor.

This rule can be understood by examining equation 2g. The individual person loadings are always taken two at a time. If, whenever a sign is reversed for one element of such a pair of d_{rr}, d_{ss}, it is also reversed for the other, then the value of the product is not changed. Now if two factors are orthogonal to one another, then the w_{rs} value for that pair is zero, and those products in which that pair of d_{rr}, d_{ss} values would have occurred will all vanish. Therefore reversal of the sign of d_{rr} to compensate for a reversal of d_{ss}, (or vice versa) is not necessary when factors r and s are orthogonal. A consequence of these rules is that for an orthogonal solution, the signs of all the "d" values are arbitrary. Only the squares of these values would enter into the solution.

In many applications of PARAFAC2, it seems reasonable to put certain constraints on the D_i matrices. It seems reasonable in multidimensional scaling, for example, to demand that all elements of D_i, (i.e. all the person weights for the dimensions), be non-negative. This can be interpreted as requiring that all persons see a given dimension as going in the same direction through the space. It prohibits a description in which a given individual uses a dimension as if it is "reversed in direction" compared to the way others use it. An option for constraining the elements of the D_i to non-negitive values has been implemented in the current PARAFAC2 computer program.

7. Analysis of Non-System-Variation Data with PARAFAC2

Only certain types of three-mode data are analysable with PARAFAC1. The data model underlying PARAFAC1 requires certain strict correspondences between the data matrices from one occasion to the next. First, it requires that the data consist of measurements of the same objects (or persons) on the same variables, on every occasion. Second, it requires that the patterns of variation of the factor influences from one occasion to the next must have a special, restricted form. Specifically, it requires that the changes in one person's loadings on any given factor be proportional to the changes in every other person's loadings on that factor. (See equation 1d or 1f. For a more detailed discussion of system-variation, and other types of data variation, see Harshman, 1970, pp. 18-26.) The requirement of system-variation data places a substantial undesirable restriction on the areas of application of the original PARAFAC1 model. It is therefore of great practical, as well as theoretical significance to note that PARAFAC2 is not restricted to analysis of system-variation data. It is possible to derive the same PARAFAC2 model for cross-products, using a much more general model for data than was used in section 2.

Let us consider the case where we have no particular correspondence between the columns of our data matrices from one occasion to the next. We shall study a set of n common measures, as was done in section 2. But instead of repeatedly measuring the same set of persons, we shall measure a different set of persons on each occasion. The number of persons will vary from occasion to occasion. On the *ith* occasion, our data matrix X_i will have the dimensions n by m_i, where m_i is the number of persons measured on that occasion.

Since the same set of n measures is used in each occasion, we hypothesize that the same set of common factors is used by the m_i persons to respond to the n measures. Let A represent the n by l matrix of factor loadings on measures. Let B_i be an m_i by l matrix giving the loadings of the m_i persons in the *ith* subpopulation on the l common factors underlying the measures. We can then describe our i data matrices as follows

(7a) $X_i = AB_i'$

Now if the correlations between the factors are caused by some basic relationship between the influences which those factors represent, then it will often be reasonable to assume that these correlations remain constant from one subpopulation to another, and that only the size of a given factor's influence will vary across subpopulations. If B_i and B_j are the factor loading matrices for any two of our subpopulations, we can express the consistent "correlations" (cross-products) of the factors by writing

(7b) $B_j'B_j = D_iB_i'B_iD_i$

where D_i is a diagonal matrix with the same interpretation as in the system-variation model. It describes the changes in "size" of the factors from one occasion to the next.

Now let us derive a representation of the cross-products derived from data which is describable by (7a) and (7b). If C_i is the cross-product matrix derived from X_i, then

(7c) $C_i = (X_i)(X_i)'$.

But by substitution of our definition (or model) for X_i given earlier in (7a) we get

(7d) $C_i = (AB_i')(AB_i')'$.

By applying the rule that the transpose of a product is the product of the transposes in reverse order, we get

(7e) $C_i = (AB_i')(B_i''A') = (AB_i')(B_iA')$

or, regrouping terms,

(7f) $C_i = A(B_i'B_i)A'$.

Now by applying the second part of our model, (7b), which postulates the same "correlations" among the factors in the different subpopulations, but differing sizes as described by the D_i diagonal matrices, we can write

(7g) $C_i = A(D_i(B_j'B_j)D_i)A'$

for some other *jth* subpopulation. In general if we write

(7h) $W = B_j'B_j$

we can then describe the cross-products in terms of the model of PARAFAC2, namely

(7i) $C_i = AD_iWD_iA'$

with the same interpretations of the constituent matrices as in the derivation from a system variation model, in section 2.

Two special cases of this general model should be noted. In the first special case, all the B_i could be taken from the same people. The size of the B_i matrices would all be the same, but the columns of B_i would not necessarily be proportional to the corresponding columns of B_j. This case corresponds to the example of personality test "object-variation" type data described elsewhere (Harshman, 1970, pp.22-23).

An even more restricted special case is where $B_j = B_iD_i$, the columns of B_i are proportional to the columns of B_j for all i,j, under study. This would be a case of system-variation. This reveals how the system-variation data model is a special case of the more general PARAFAC2 model.

8. Estimation Algorithms

Work is currently in progress to find the most efficient way of fitting the PARAFAC2 model to a set of data. A detailed report will be forthcoming in a later issue of W.P.P., but a brief glimpse of current developments is appropriate here.

An algorithm for PARAFAC2 has been devised by Robert Jennrich (UCLA Dept. of Mathematics) and implemented by the author as a FORTRAN IV program on the IBM 360/91 computer. The algorithm uses an approach similar to the "quick algorithm" for PARAFAC1. It estimates values for one of the three basic sets of parameters (A, D_i or W) while holding the other two sets of parameters fixed. A complete iteration consists of first estimating the D_i's, then W, and finally A. The program iterates this process until the values converge on a stable least-squares solution.

This procedure generally works, but it is expensive, and sometimes slow to converge. Even worse, from certain bad starting positions local optima or extremely slow convergence can be encountered. We are accumulating experience, however, which would show how, by putting certain constraints on starting positions and values of the parameters, these deadlocks can largely be avoided.

A basic modification to this special algorithm is currently under study. An attempt is being made to combine Jennrich's algorithm with a generalized optimization routine called TBU (Reichenbach, 1962, 1969). The combined program alternates between two phases: first, it observes how the parameters change during an iteration of Jennrich's algorithm. It computes or updates trends for the direction and amount of change in each parameter. In the second phase, it moves all the parameters simultaneously an optimal (hopefully large) distance in the direction of their respective trends. The author has been working with Hans Reichenbach on this approach and has found that it can rapidly accelerate the convergence of PARAFAC1 in certain difficult cases. The application of this technique to PARAFAC2 is the next step to be implemented. Preliminary results with PARAFAC2 have been encouraging.

9. Avoiding the "Communalities" Problem

In deriving our model of the factor structure underlying data and cross-products, we have ignored the fact that there would be a certain amount of error or "noise" in the data. For PARAFAC1, and its estimation procedure, this approach was justified. We are looking for a least squares best estimate of the common-factor loadings, or the A, D_i and B values. The process of directly optimizing the fit of these parameters to the raw data provides the least square estimates that we desire. The problem becomes more complicated, however, when we consider cross-product matrices.

Let us take the following model of "noisy" data:

(9a) $$\tilde{X}_i = X_i+E_i = AD_iB_i'+E_i$$

where E_i is a matrix of the same size as X_i, containing random "error terms". (In traditional E_i actually represents "unique variance" composed of both "specific" and "error" variance.) Since X_i represents *all* the common-factor part of the data matrix $\tilde{X}_i$, then what remains when we subtract it from $\tilde{X}_i$ should be a matrix of uncorrelated random errors. The rows of E_i should be uncorrelated both with the rows of X_i and with each other. The means of these random errors should be zero.

We can develop a model for the cross-products from such "noisy" data, as follows:

(9b) $$\tilde{C}_i = \tilde{X}_i\tilde{X}_i'$$

By substituting from (9a) into (9b) we get

(9c) $$\tilde{C}_i = (X_i+E_i)(X_i+E_i)'$$

(9d) $$\tilde{C}_i = (X_i+E_i)(X_i'+E_i')$$

(9e) $$\tilde{C}_i = X_iX_i'+E_iX_i'+ X_iE_i'+E_iE_i'$$

now, since the rows of E_i are uncorrelated with the rows of X_i, we can write

(9f) $$\tilde{C}_i = XX' +0+0+E_iE_i'$$

(9g) $$\tilde{C}_i = C_i+E_iE_i'$$

Although the rows of E_i are uncorrelated with one another, each row is of course correlated with itself. Thus $(E_i)(E_i')$ will yield a diagonal matrix whose diagonal elements represent the sum of the squared errors for that row of the data matrix (i.e. a measure of the amount of noise in that variable). This diagonal matrix is added to the matrix predicted by the common factors, and therefore increases the size of the diagonal elements of the cross-product matrix beyond what would be predicted by the factors.

An attempt to achieve a least squares best fit for all the values of the covariance matrix, including the diagonals, would distort the solution, since the diagonals are larger than would be predicted by the factor loading matrices.

This problem is analogous to the standard "communalities" problem of traditional two-mode factor analysis. We solve this problem in the following way. In the algorithm which performs the fitting of the model to the data (see section 8 of this article) we perform an extra step at the end of each iteration. In this step, we replace the diagonal entries of the cross-product matrices with the values for those diagonal entries which are predicted by the model at that iteration. This will cause the values in the diagonals to "float" one iteration behind the predictions of the model. As the algorithm proceeds, the difference between these diagonal values (from the last iteration) and the models prediction (from this iteration) will soon become very small, and therefore will not influence the process of convergence. The diagonal values will, in effect, be ignored and only the off diagonal values will be 'fit' by the program. This should provide the best possible least squares estimate of the A, D_i and W parameter matrices.

(This same technique for ignoring cells in the data matrices can be used to ignore missing values in PARAFAC1 or PARAFAC2.)

10. Analysis of Correlation Matrices

The technique for analysis of cross-product matrices developed in this article applies directly to covariance matrices, as was pointed out in section 2, but it does not apply directly to the analysis of correlation matrices.

In a correlation matrix, the rows and columns of the covariance matrix have been rescaled so that the maximum value for any entry is 1.0. This is accomplished by dividing each entry by the standard deviation of its column and row variable. In other words,

(10a) $$\text{Corr}_{ij} = \text{Cov}_{ij}/s_i s_j$$

where s_i, s_j represent the standard deviation of the *ith* and *jth* variables respectively.

In matrix notation, we can represent the relation between the covariance matrix C and the correlation matrix R, as follows

(10b) $$C = \underline{D}R\underline{D}$$

where $\underline{D}$ is a diagonal matrix whose *ith* diagonal element is the standard deviation of the *ith* variable. Therefore

(10c) $\quad R = \underline{D}^{-1}C\underline{D}^{-1}$

and if we substitute our PARAFAC2 model for the covariance into equation (10c) we get

(10d) $\quad R_i = \underline{D}_i^{-\frac{1}{2}}AD_iWD_iA'\underline{D}_i^{-\frac{1}{2}}$.

This expression is more complicated than PARAFAC2, and its mathematical properties have not been investigated. It serves to demonstrate, however, why correlation matrices cannot be directly analysed by PARAFAC2. It would seem inefficient to derive a further algorithm for this model (which may not have the same degree of uniqueness) when one can simply compute covariance matrices instead of correlation matrices and analyse them with PARAFAC2.

11. Relation of PARAFAC2 to Jennrich's multidimensional scaling model

Jennrich has developed a generalization of Carroll and Chang's INDSCAL multidimensional scaling model which leads to a decomposition of matrices of summed-cross-products which is even less restricted than PARAFAC2. His derivation is based on a generalized Euclidean distance metric, and allows each person to use his own metric in judging inter-stimulus distances. His derivation is presented in an accompanying article (Jennrich 1972).

It is possible to reach Jennrich's model by a slightly different path, namely generalizing PARAFAC2 so that the correlations among the factors underlying the data can vary from one occasion to the next.

Start with the model of PARAFAC2, namely

(11a) $\quad C_i = AD_iWD_iA'$.

Now let each occasion have its factors $\overset{*}{A}$ related to A by some oblique or orthogonal rotation. Let T_i be the l by l matrix describing this rotation by giving the projections of the axes of A onto the axes of $\overset{*}{A}$.

If we let $\overset{*}{A}_i$ be the axes which are expanded or contracted on each occasion (or by each individual), then

(11b) $\quad \overset{*}{A}_i = AT_i$

(11c) $\quad C_i = (AT_i)D_iWD_i(T_i'A')$

(11d) $\quad C_i = A(T_iD_iWD_iT_i')A'$

(11e) $\quad C_i = AW_iA'$

This model would not give a unique solution since, for example, we can always let $\tilde{A} = AR_i$ and $W_i = (R_i^{-1}W(R_i')^{-1})$. It might be useful, however, to choose R_i such that $\tilde{A}$ approaches A of PARAFAC2 (2f) as closely as possible. Then the following research strategy might seem reasonable. First, analyze with PARAFAC2. Then use Jennrich's model (11e) to analyze the same data. If the improvement of fit gained with the more general model is significant enough (by some criterion) to lead the investigator to suspect that there were "real" changes in obliqueness across occasions, then rotate the solution of (11e) into maximal agreement with (2f) as just described above. The model of (2f) (PARAFAC2) describes the common space, and (11e) describes how each occasion (or individual) distorts that common space by rotating the axes, in addition to simply expanding or contracting them.

Carroll (personal communication) has recently pointed out that this (11e) model is mathematically equivalent to one which he discusses in conjunction with INDSCAL (Carroll, 1970). The equivalence is easily missed, because Carroll interprets his model in terms of a common set of axes which each person subjects to a different *orthogonal rotation* in addition to contraction or expansion. The interpretation developed above by Jennrich and the author is one of each person imposing an *oblique transformation* on the common axes in addition to his expansion or contraction of the different axes.

Carroll and some others have raised questions about the interpretation of oblique perceptual axes. This interesting question will not be treated here. The author believes that interesting and plausible interpretations do exist, and notes, more importantly, that PARAFAC2 analysis of real perceptual data does in fact sometimes reveal oblique perceptual dimensions. (see Terbeek and Harshman, 1972.)

Carroll is developing a program to analyse data in terms of the more general model of (11e). It will be called IDIOSCAL (for IDIOsyncratic SCALing). If we applied to IDIOSCAL the interpretation developed above for (11e) and rotated it into maximum aggreement with PARAFAC2, it might by a very useful partner to PARAFAC2 in multidimensional scaling.

References

Carroll, J.D. and Chang, J.J. (1970), "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika* 35, 283-319.

Harshman, R.A. (1970), "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-mode factor analysis," *Working Papers in Phonetics No. 16,* 84pp.

Jennrich, R. (1972), "A generalization of the multidimensional scaling model of Carroll and Chang," *Working Papers in Phonetics No. 22*

Reichenbach, H.G. (1962), Routine TBU: routine for optimizing a set of variables. Report 62-21, Institute of Transportation and Traffic Engineering, University of California at Los Angeles, 1962. (unpublished expanded version dated 1969).

Terbeek, D. and Harshman, R.A. (1972), Is Vowel Perception Non-Euclidian? *Working Papers in Phonetics No. 22*

Torgerson, W.S. (1958), *Theory and Methods of Scaling,* New York: John Wiley and Sons, Inc.

A Generalization of the Multidimensional Scaling Model of Carroll and Chang

Robert Jennrich

(Mathematics Department, UCLA)

1. Carroll and Chang model

Carroll and Chang (1970) assume each stimulus may be represented by a point

$$p_j = (x_{j1}, \ldots, x_{jr}) \quad .$$

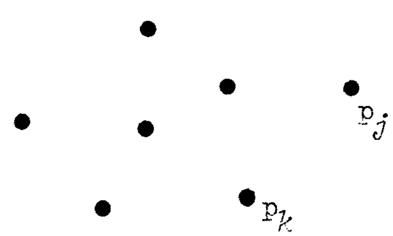
The theoretical distance between the j and kth stimulus as judged by the ith individual is given by

$$d_{jk}^{(i)} = \left(\sum_{t=1}^{r} w_{it}(x_{jt} - x_{kt})^2\right)^{1/2} \quad .$$

For each i this defines a norm on the r-dimensional space which is the home of the stimulus points p_j. Each individual views the points differently, the differences being reflected in the dependence of the weight w_{it} on individual index i.

2. Generalization

In pictures, Carroll and Chang view the stimuli as points in an N-dimensional space.



The most general Euclidean distance formula in such a space is

$$d^2(p_j, p_k) = (p_j - p_k)\ W\ (p_j - p_k)'$$

where W is a symmetric positive definite matrix. Carroll and Chang have demanded that W be diagonal. We simply suggest relaxing this restriction. Each individual i then has his own completely arbitrary Euclidean metric and the distance between the jth and kth stimulus as seen by the ith individual becomes

$$d_{jk}^{(i)} = \left(\sum_{s=1}^{r} \sum_{t=1}^{r} w_{ist}(x_{js} - x_{ks})(x_{jt} - x_{kt}) \right)$$

3. Estimation

For purposes of estimation the distances $d_{jk}^{(i)}$ are converted to inner products (Torgerson 1958)

$$b_{jk}^{(i)} = - 1/2 \left((d^{(i)})^2_{jk} - (d^{(i)})^2_{j\cdot} - (d^{(i)})^2_{\cdot k} + (d^{(i)})^2_{\cdot\cdot} \right)$$

where "·" means average as usual. We now have, assuming no errors,

$$b_{jk}^{(i)} = \sum_{s=1}^{r} \sum_{t=1}^{r} w_{ist} x_{js} x_{kt} \quad .$$

Using observed values of $b_{jk}^{(i)}$ we will attempt to fit these by least squares using the model on the right.

4. Algorithm

Let $B_i = (b_{jk}^{(i)})$, $W_i = (w_{irs})$, $X_i = (x_{js})$. Compute

$$Y = (\textstyle\sum_i B_i X W_i)(\sum_i W_i X' X W_i)^{-1}$$

$$\tilde{X} = 1/2\ (X + Y)$$

$$\tilde{W}_i = (\tilde{X}'\tilde{X})^{-1}\tilde{X}' B_i \tilde{X}(\tilde{X}'\tilde{X})^{-1} \quad .$$

$\tilde{X}$ and $\tilde{W}$ are the new values of X and W. Iterate for convergence.

References

Carroll, J.D. and Chang, J.J. (1970), "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckhart-Young' decomposition," *Psychometrika* 35, 283-319.

Torgerson, W.S. (1958), *Theory and Methods of Scaling*, New York: John Wiley and Sons, Inc.

An auditory-motor theory of speech production

Peter Ladefoged, Joseph DeClerk, Mona Lindau,
and George Papçun

A great deal has been written about the perception of speech sounds, suggesting that it might be appropriate to have a "motor theory of speech perception". (Liberman et al. 1967) We will discuss a set of data on the shape of the tongue in vowels, and consider the converse possibility, namely, that it is equally plausible to talk about an auditory theory of speech production. Most of the evidence for the motor theory of speech perception comes from studies on consonants. It may be true that, as far as consonants are concerned, "speech sounds are perceived by reference to the articulatory movements that produced them" (Liberman, 1957). But in the case of vowels, speech sounds may be produced by reference to their acoustic images. Our principal reason for making this claim is that our data show that different speakers use different articulatory gestures when producing what is heard as the same vowel.

To see how it is possible for more than one combination of articulatory gestures to produce vocal tract shapes which have the same auditory effect we must consider the forces which determine the shape of the tongue. A basic description of the muscles involved has been given previously (Ladefoged, 1964). The slight divergencies between that account and the scheme to be presented here are due to insights provided by other investigators (Hirano and Smith, 1967; Lindblom and Sundberg, 1971; Smith 1971). It now seems as if the major forces acting on the tongue are as shown in Figure 1.

Because the tongue is located within the mandible, the first point to be considered is the position of the mandible itself. The muscles responsible for raising and lowering the mandible are not shown in Figure 1. We need only note that, since the effective hinge of the mandible is above and behind the tongue, the mandible moves downward and backward as the mouth is opened. Consequently lowering the mandible will lower the front part of the tongue more than the back part, and it will also cause the lower part of the tongue to become somewhat retracted, so that the pharynx is slightly narrowed. The effect of lowering the mandible is indicated in Figure 1 by the arrow marked 1.

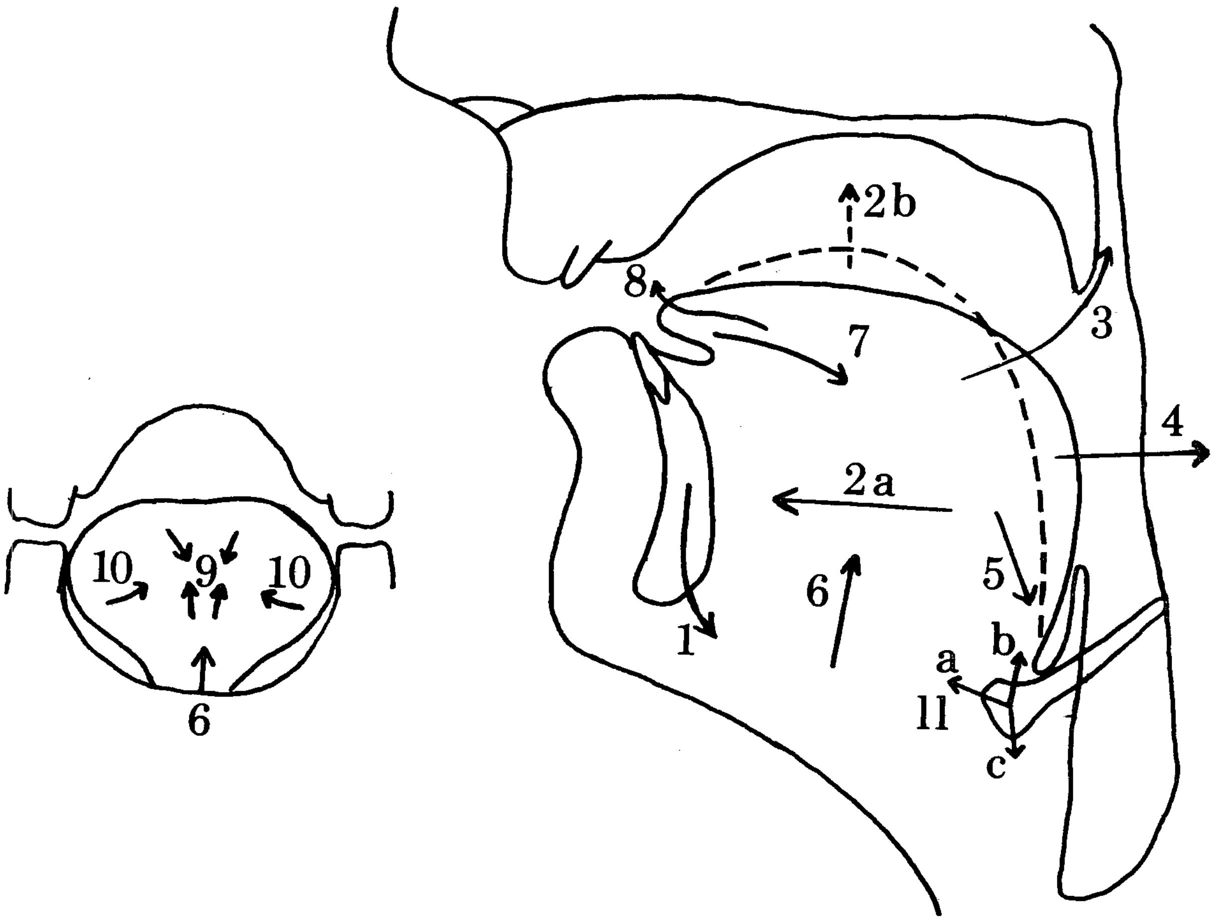


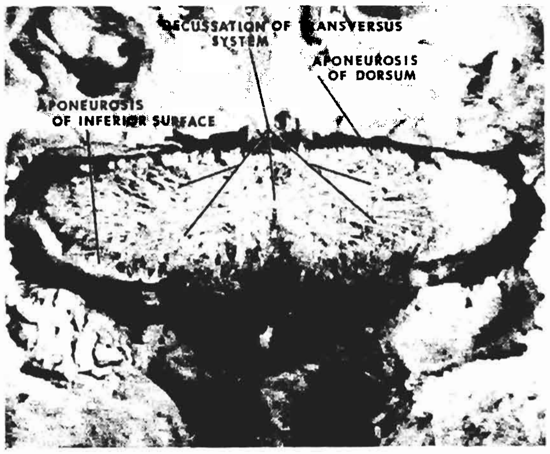
Figure 1. The major forces affecting the position of the tongue. (1) Jaw opening; (2a) M. Genioglossus; (2b) Affect of (2a); (3) M. Stylo-glossus; (4) M. Glossopharyngeus; (5) M. Hyoglossus; (6) M. Mylohyoideus; (7) M. Longitudinalis inferior; (8) M. Longitudinalis superior; (9) M. Verticalis; (10) M. Lateralis; (11) Forces affecting the hyoid: (a) M. Geniohyoideus; (b) M. Stylohyoideus; (c) M. Thyrohyoideus.

The principal muscle of the tongue is the genioglossus. Its most important function is to pull the root of the tongue forward, in the direction shown by the arrow marked 2a in Figure 1. A pull in this direction compresses the tongue within the mandible. But since the mass within the mandible must remain constant, the effect of simply pulling the tongue root forward is that the front of the tongue is displaced upward as indicated by the dashed position of the tongue, and the arrow marked 2b in Figure 1.

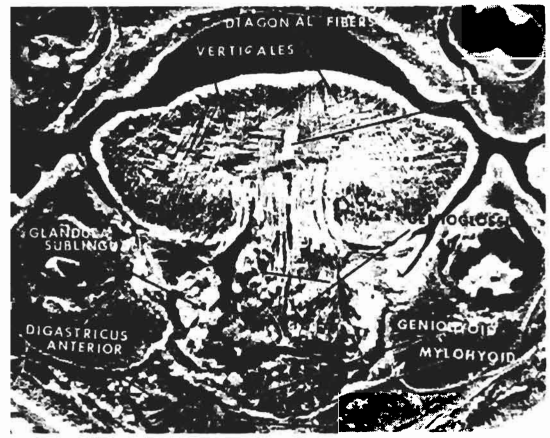
There are several muscles which can be used to oppose the effect of the genioglossus and pull the tongue backwards. The styloglossus will pull backwards and upwards as indicated by arrow 3. The glossopharyngeus (which some anatomists list as part of the superior pharyngeal constrictor) can pull the tongue directly backwards as indicated by arrow 4; it is the action of this muscle, possibly in conjunction with other parts of the pharyngeal constrictor muscles, which results in the lateral wall of the pharynx being drawn together in the vowel /a/ as observed by (Minifie et al. 1970). The tongue can also be slightly retracted by the action of the hyoglossus, which pulls the tongue down and back as indicated by arrow 5.

Figure 1 also shows the actions of the mylohyoid muscle (arrows 6). As can be seen from the cross section of the tongue in the plane of the second molars (on the left of the figure), this muscle is like a sling support beneath the tongue; when it contracts it will raise the whole body of the tongue. Since it is attached to the hyoid bone near the root of the tongue, its action will also cause the root of the tongue to be pulled slightly forward.

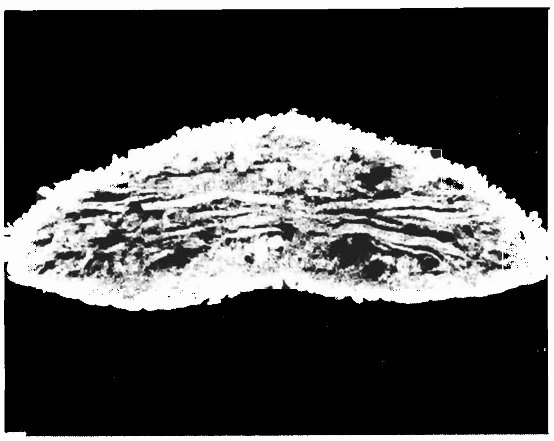
The actions of some of the intrinsic muscles of the tongue are also shown in Figure 1. Since these muscles do not have their origin at any point outside the tongue they cannot cause the mass of the body of the tongue to move in any direction. They can, however, cause the contraction of the tongue in certain directions, and hence its expansion in other directions. The inferior longitudinals (arrow 7) can cause a lengthwise contraction of the tongue, the tip being drawn back so that the center of the tongue is bunched up. The superior longitudinals (arrow 8) also act on the tip of the tongue. They are probably responsible for the raising of the tip of the tongue (an action which will not be considered in this paper). The verticalis (the arrows marked 9) and the lateralis (the arrows marked 10) have fibers running more or less orthogonally to the longitudinals, and to each other. These intrinsic muscles may play a role in the formation of the tongue shapes by supplying feedback information on the degree to which they have been stretched (cf. Bowman 1971). They may perhaps be regarded as position sensors rather than prime movers.



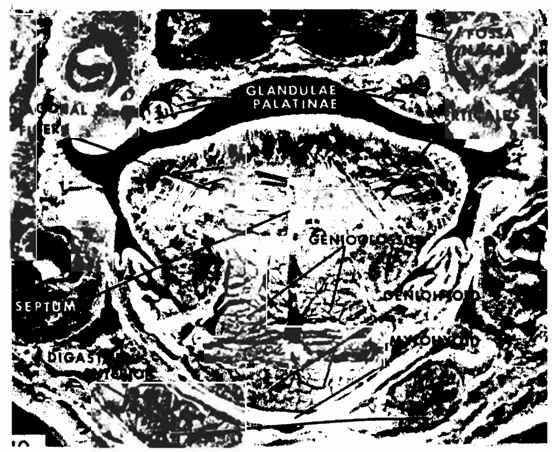
4 1/2 mo. fetus, Homo. Decussation through midline of bundles, just back of tongue-tip.



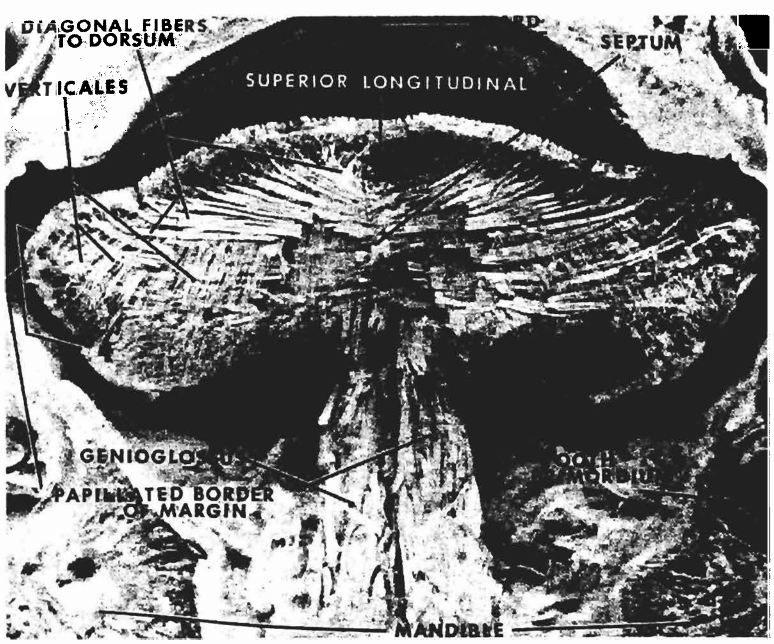
4 1/2 mo. fetus, Homo. Frontal section through middle of dorsum.



38 day Homo. Decussation as above.



4 1/2 mo. fetus, Homo. Frontal section, through first fetal molar.



4 1/2 mo. Homo. Frontal section dissected, just back of frenulum showing complete bundles of fibers from septum to dorsum of aponeurosis.

Figure 2. Illustrations from Strong, L. H. (1956) "Muscle Fibers of the Tongue" Anatom. Record 126, 61 - 79.

Finally we must consider the effect of movements of the hyoid bone. This bone forms the support for the root of the tongue. As is indicated by the arrows marked 11, it can be pulled (a) forward by the geniohyoid muscle, (b) upward and backward by the stylohyoid muscle, and (c) downward by the thyrohyoid muscle. Each of these movements will result in a corresponding displacement of the tongue root.

Several of the muscles which we have been considering can be seen in the excellent photographs published by Strong (1951) three of which are reproduced here (by permission) as Figure 2.

Our main conclusion from our study of the anatomy is that there are a number of muscular actions which have, at least in part, similar results. The present paper is most concerned with the mechanisms for varying the height of the tongue. There are three or four distinct ways in which tongue height can be varied. Firstly there is the movement of the jaw. Secondly there is the pull of the genioglossus on the root of the tongue, and the concomitant raising of the front of the tongue. Thirdly there is the lifting action of the mylohyoid and styloglossus muscles. Fourthly, and less importantly, the inferior longitudinals may bunch up the tongue and thus contribute to tongue height. The action of these muscles are difficult to distinguish from those of the genioglossus and the mylohyoid, and will not be considered further in this paper. But we can show from the data which we have collected that different individuals use different combinations of the first three articulatory gestures for controlling the height of the tongue.

Data Collection

We selected six male subjects between the ages of 21 and 40 who had been born, raised, and educated in the Midwest. Their speech was recorded and the recordings were evaluated by a panel of three linguists. The panel agreed that each subject was a speaker of the same variety of American English. Each subject was examined and found to have no hearing loss or other physical deficits which could affect his speech.

Subjects were given practice in saying (among other things) ten sentences of the form "say h_d again". The vowels in the frame were /i,ɩ,e,ɛ,æ,a,ɔ,o,ʊ,u / as in the words "heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd". In this paper we will be mainly concerned with the pronunciation of only the first five of these sentences, i.e. those containing front vowels in the test words.

Cinefluorograms were obtained by use of techniques described in full by DeClerk, Landa, Phyfe and Silverman (1965). The subject was placed between a Picker 150 KV X-Ray tube and a 9" Picker Image Intensifier, in a sound-proofed booth. The sound insulated booth was used to reduce the noise from the rotating anode of the X-Ray generator and the motor of the

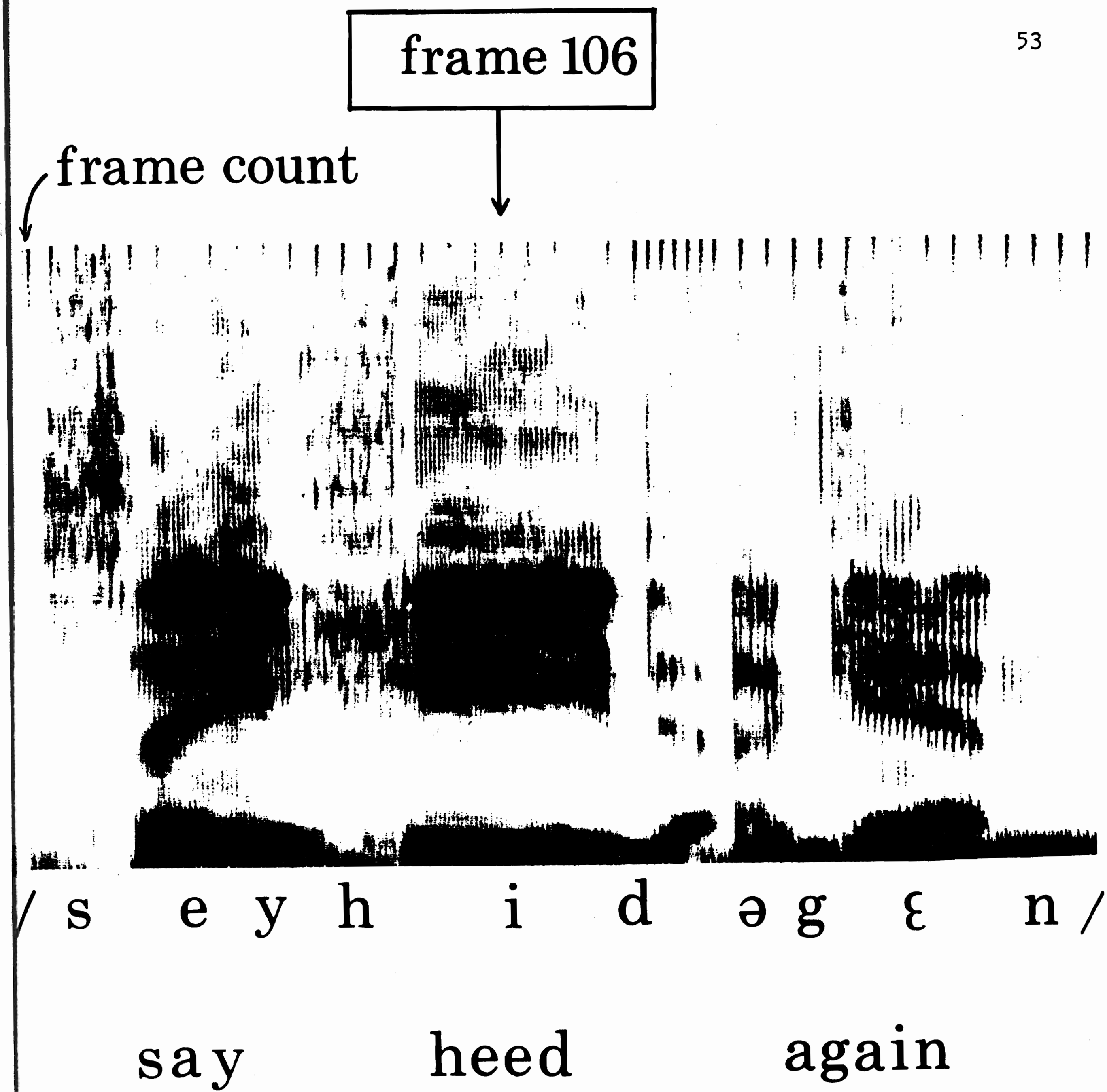


Figure 3. Spectrogram of the phrase "Say heed again" as said by subject 1, showing the digitally coded frame count, and the frame that was marked as being in the middle of the word *heed*.

high speed motion picture camera. A Brüel and Kjær condenser microphone was used to record the speech material onto one track of a two track tape recorder. A 9" Picker Image Intensifier was used to intensify the brightness of the X-Ray image and convert it to a visible image suitable for photographing.

The Image Intensifier is an evacuated glass tube containing a fluorescent screen on a thin aluminum base. In contact with the screen is a photoemissive surface. X-Radiation after passing through the subject strikes the fluorescent screen. The screen fluoresces, giving off light which impinges point by point on the photoemissive surface. Electrons are knocked off the photocathode of this tube by the impact of photons from the subject. The image leaving the surface is now an electron image which is accelerated through a 25 KV field and electrostatically focused onto an output phosphor,and reduced in size onto a phosphor screen where the electrical energy is reconverted to light. The luminous phosphor screen is then imaged on film using high speed conventional optics. The amplification arises from the high energies of the electrons striking the output phosphor and from the electro-optical minification of the image size. The amplification makes possible a reduction in radiation exposure to the subject.

The output phosphor of the image intensifier was photographed by means of a 35 mm Mitchell high-speech (120 frames per second) motion picture camera, using Ektachrome ER high speed (ASA 160) color film. The camera was placed in a double walled sound attenuating blimp so that its sound level during operation was reduced by approximately 45 decibels.

Before filming, the midline of the lips and the dorsal surface of the subject's tongue was striped with a radio-opaque substance (Microtrast). Rugar was sprayed in a thin concentration into the nasal cavity to show up the posterior surface of the soft palate. A 10mm steel pin was placed vertically on the chin surface to provide an appropriate scale for later measurements of vocal tract dimensions. Measurements showed radial distortion to be negligible within the area of the frame in which vocal tract measurements were made in the study.

The subject was cinefluorographed in a mid-sagittal view, while reading the sentences. A simultaneous audio record of the session was also obtained. By use of a specially constructed digital synchronizer a reference signal was placed on a second channel of the tape recorder and on the film, during every fourth frame, and a digital code derived from a 6 stage binary counter was recorded in both places every 128 frames. The digital code indicates which one of 64 code groups of 128 frames one is analyzing. This technique allows frame by frame synchronization between the articulatory event and the acoustic signal. A sound spectrograph representation of a typical utterance showing the synchronizing pulses on the upper part of the spectrogram is shown in Figure 3. Spectrograms were made of each of the sentences. Each spectrogram was examined, and

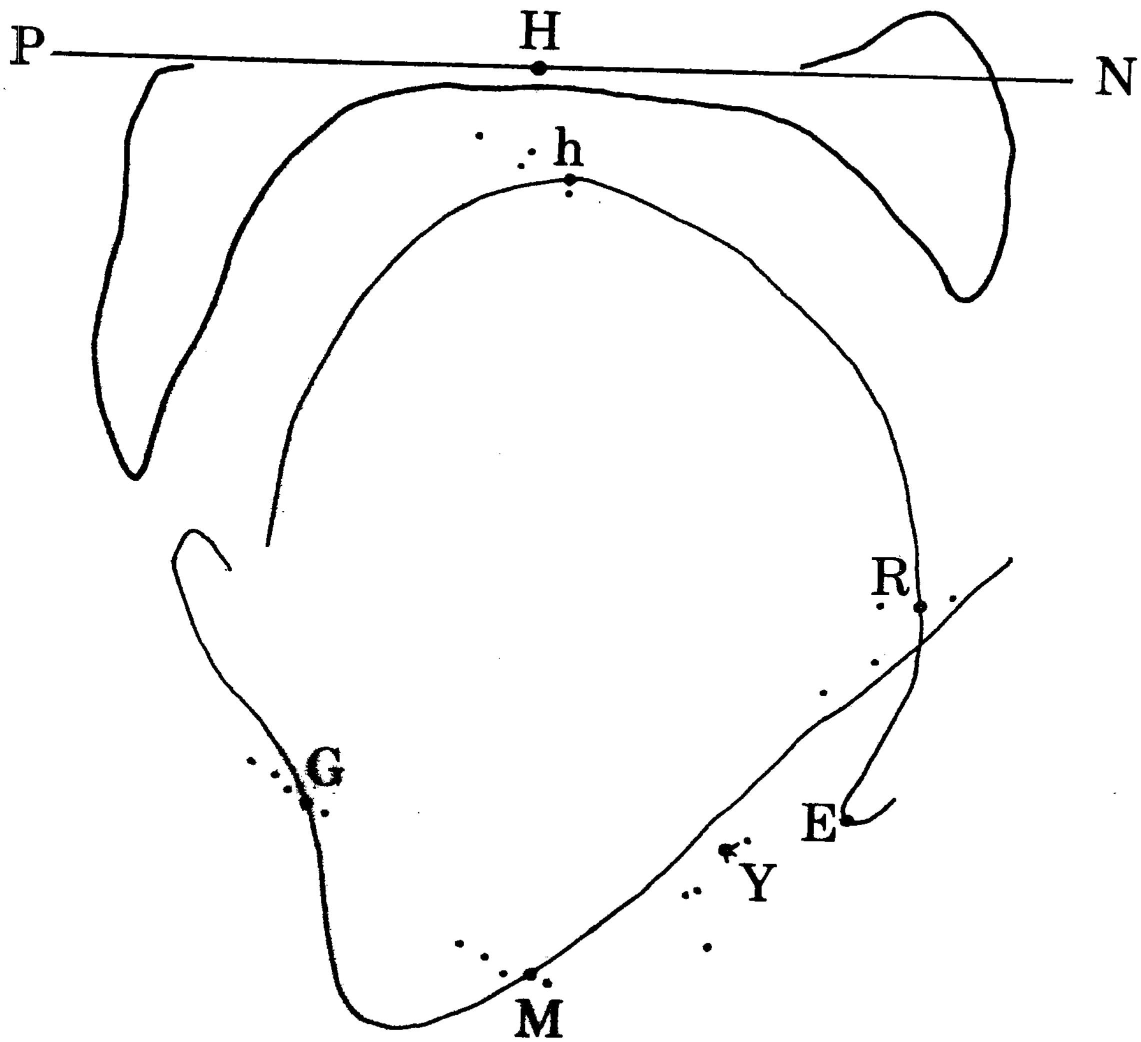


Figure 4. Points used in measurement. (PN) Plane of the base of the nasal cavity; (h) Highest point of the tongue for a given vowel; (H) Average horizontal position with respect to PN of the highest points of the tongue; (M) Fixed point on the mandible directly beneath H; (G) Fixed point on the mandible near the origin of the genioglossus; (E) Root of the epiglottis; (R) Tongue root 18 mm up from E; (Y) Most anterior and superior edge of the hyoid.

an appropriate point in the vowel in the test word was marked. In the case of the lax vowels in *hid*, *head*, and *had* a steady state part of the second formant was selected. For the tense vowels in *heed* and *hayed*, which for some speakers were diphthongal throughout, a point shortly after the first consonant was selected. The corresponding frame in the film was then located and traced.

Although image intensification was used to make the x-ray photographs as clear as possible, tracing them was still difficult, and some uncertainties remained in the tracings. The uncertainties were largely eradicated by making composite tracings. One composite was made in which the five separate tracings of the nasal cavity were superimposed on each other. In this manner it was possible to decide on the true outline of the nasal cavity, and correct errors made in the original tracings; i.e., where a particular tracing diverged from the others it was corrected. Occasionally an average needed to be taken from several representations, but usually it was possible to get two or three of the tracings to agree quite accurately and thus to dispel errors in the others.

Five tracings of the tongue in its position for the five vowels were added to the composite tracing of the nasal cavity in their correct positions relative to the nasal cavity as determined by their positions on the individual tracings. The position of the hyoid for the five vowels was also marked on this tracing.

A composite tracing of the mandible, including the tongue positions for the five vowels placed in correct position relative to the mandible, was made in a manner similar to that of the nasal cavity composite.

Then, when the representation of the tongue position for any given vowel from the nasal cavity tracing was overlaid upon the corresponding representation for the tongue on the mandible tracing, what resulted was the clearest and most correct representation we could develop of the outline and alignment of nasal cavity, mandible, and tongue for the given person saying the given vowel. It was from composites aligned in this manner that measurements were taken. Figure 5.1 through 5.6 are tracings from aligned composites. The unsteadiness of the jaw and tongue lines results from an attempt to trace them as minutely as possible. By contrast, Figure 5.2 is second-generation, steadier, but less accurate.

For explication of the method of taking the measurements, we use an abstraction from the composite tracings for subject 1. This figure is labeled here as Figure 4 and is the same as the /ɛ/ position in Figure 5.1 with the addition of labeled points indicating the landmarks used for measuring. The points which do not intersect the figure indicate the articulatory motions for the other vowels.

On the composite representation of the nasal cavity the plane of the base of the nasal cavity was established (line PN). The highest point of the tongue (h) for a given vowel is defined as the point of the tongue which most nearly approaches line PN. The average horizontal position of the highest point of the tongue (H) was determined by averaging the distances along line PN of the high points for the several vowels. From H a line perpendicular to PN was projected across the mandible with the mandible in the position for /ε/. Point M, the point on the mandible directly beneath H, was defined by projecting a perpendicular from PN at H across the mandible, with the mandible in the position for /ε/. The position for the vowel /ε/ was chosen because it was the vowel usually nearest average mandible height. (Efforts were made to determine a true average position for the mandible, but they were abandoned because it was found difficult or impossible to determine correct extrapolations for intermediate positions, let alone to specify what had been done explicitly enough to provide for replication, and also because of the very small changes in ultimate measurements such procedures would be likely to introduce.) Point G was marked on the innermost curve of the anterior of the mandible as representing the mandibular insertion of the genioglossus. Point R is a point on the root of the tongue 18 mm up from the root of the epiglottis (E).

The following measurements were taken to the nearest half millimeter:

1. Tongue height: The perpendicular distance between PN and h.

2. Mandible height: The perpendicular distance between PN and M, as the mandible was placed in position for each of the five vowels. (Notice that M is a fixed, physical point on the mandible.)

3. Tongue root advancement: The distance from G to R.

4. Hyoid height: The perpendicular distance between PN and the most anterior and superior edge of the hyoid.

5. Tongue lift: The distance along a perpendicular to PN between M and h. The figure for tongue lift was calculated by subtracting M from h, and therefore must be regarded as a derived rather than a direct measurement.

Individual subject behavior

The consolidated data for each subject is shown in Figures 5.1 through 5.6. The upper part of each figure shows the tongue position at a moment in each of the five front vowels. Below this are graphs showing the relative movements of the highest point of the tongue (always on the ordinate) as compared with (on the abscissa) the relative height of the mandible, the relative degree of lifting of the tongue within the mandible, the relative pharynx width resulting from variations in the

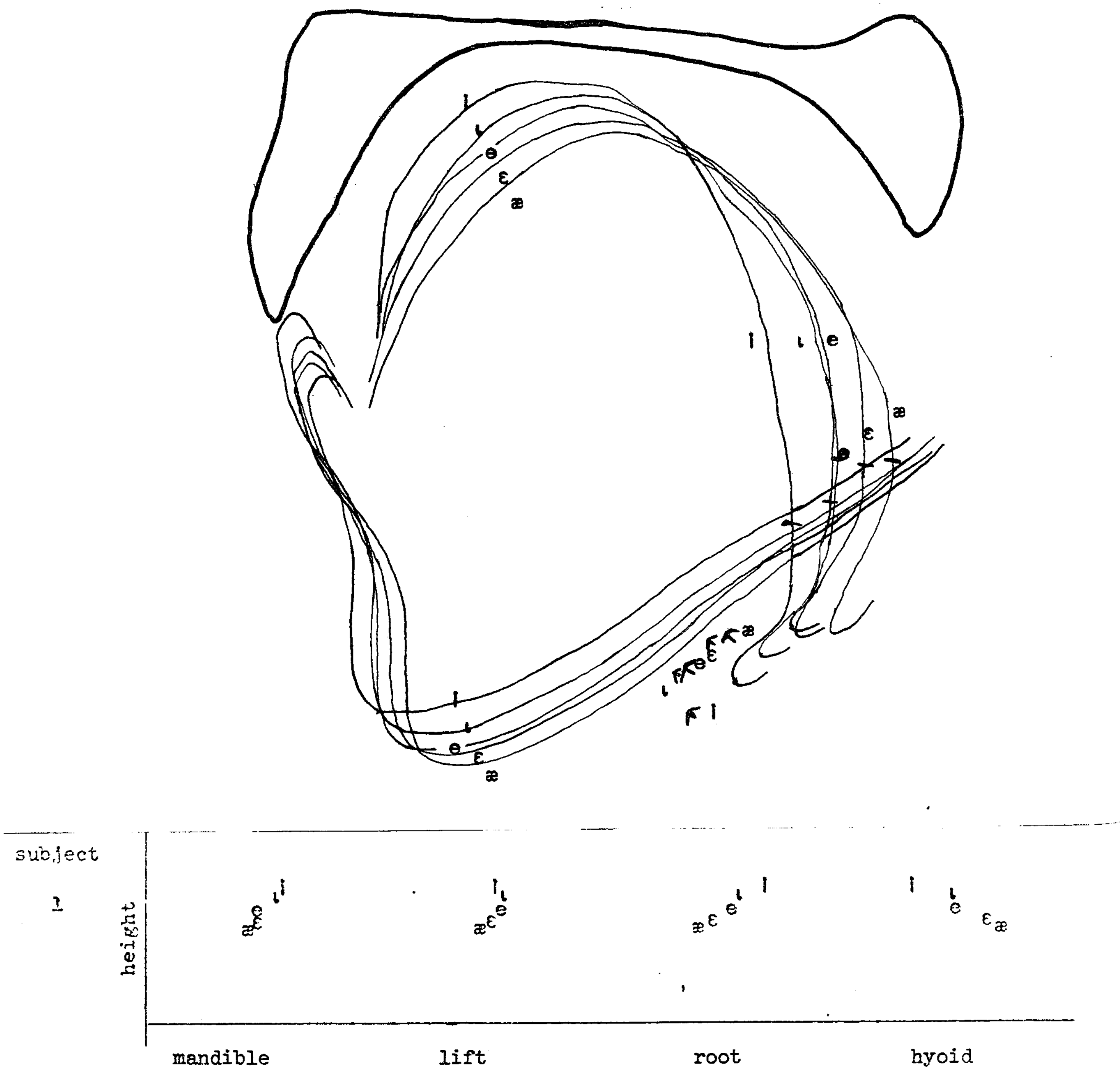


Figure 5.1 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 1.

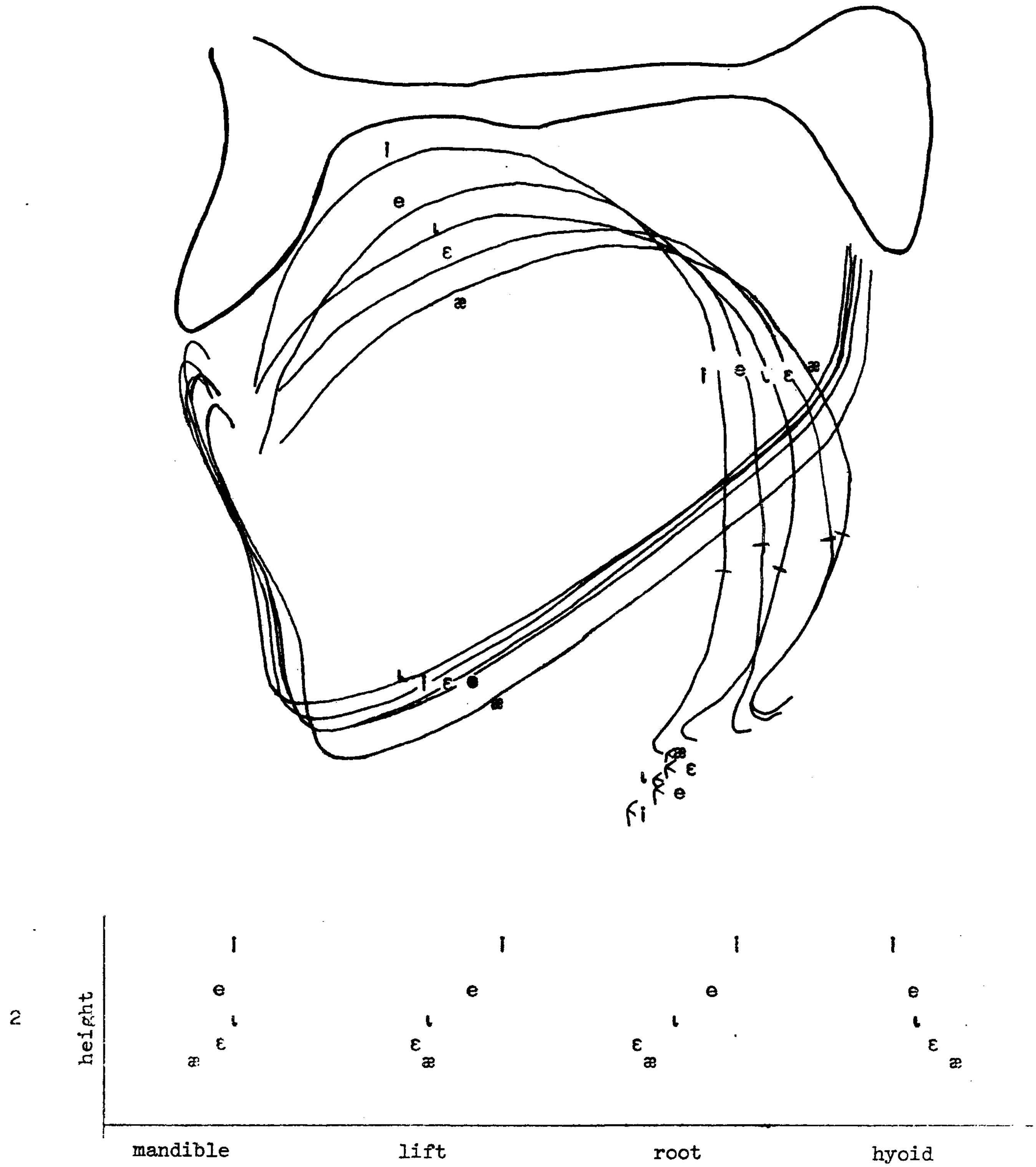


Figure 5.2 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 2.

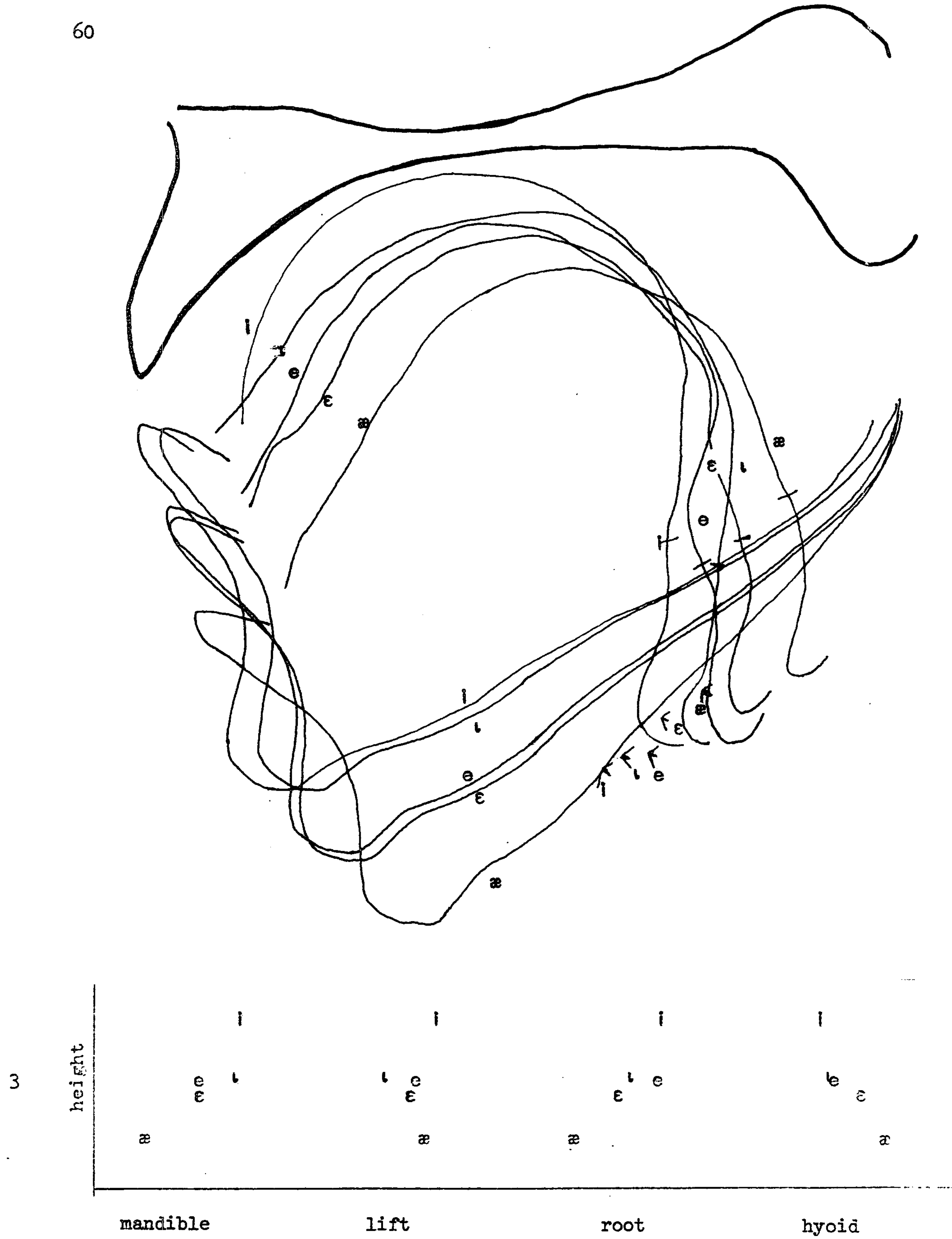


Figure 5.3 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 3.

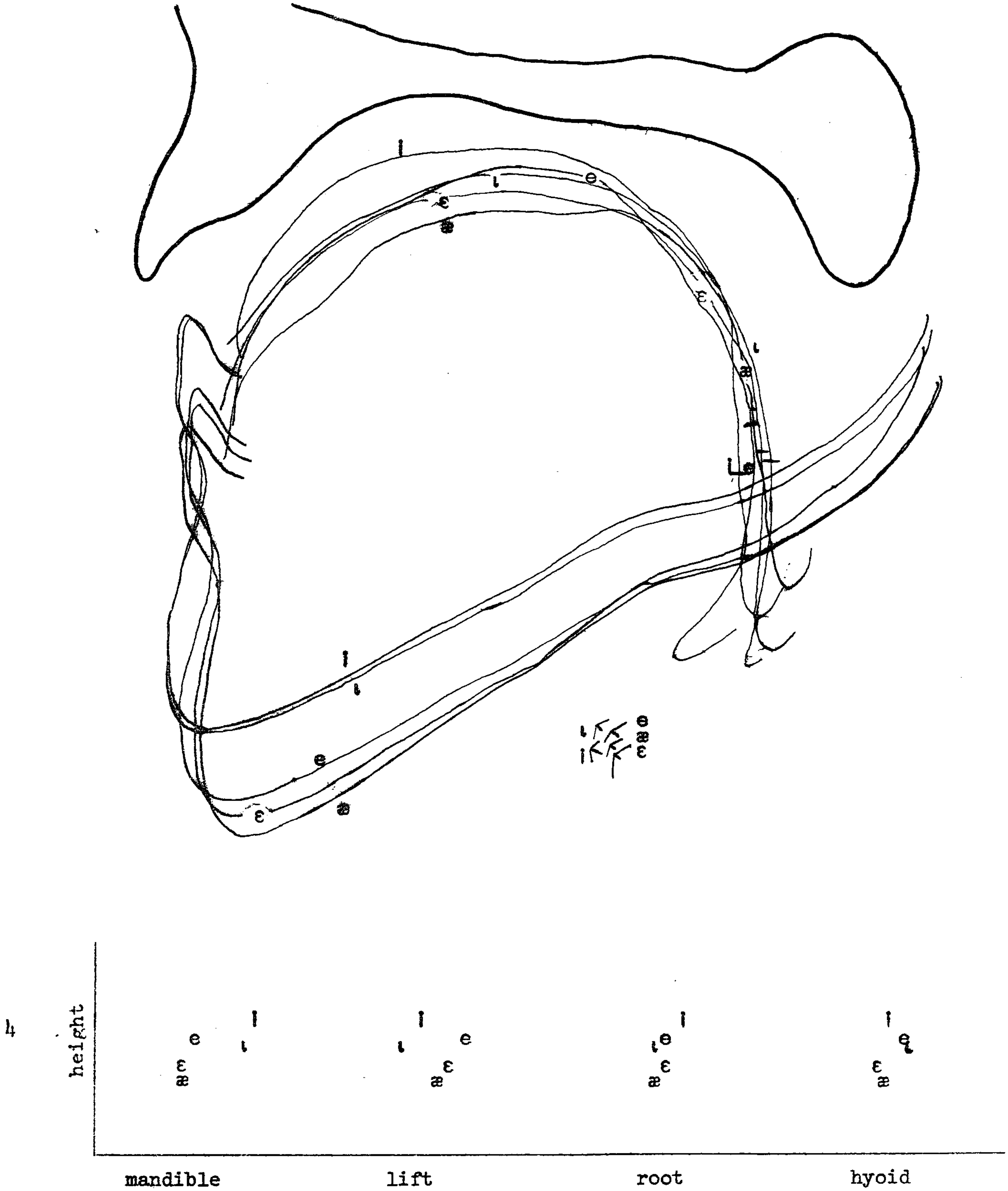
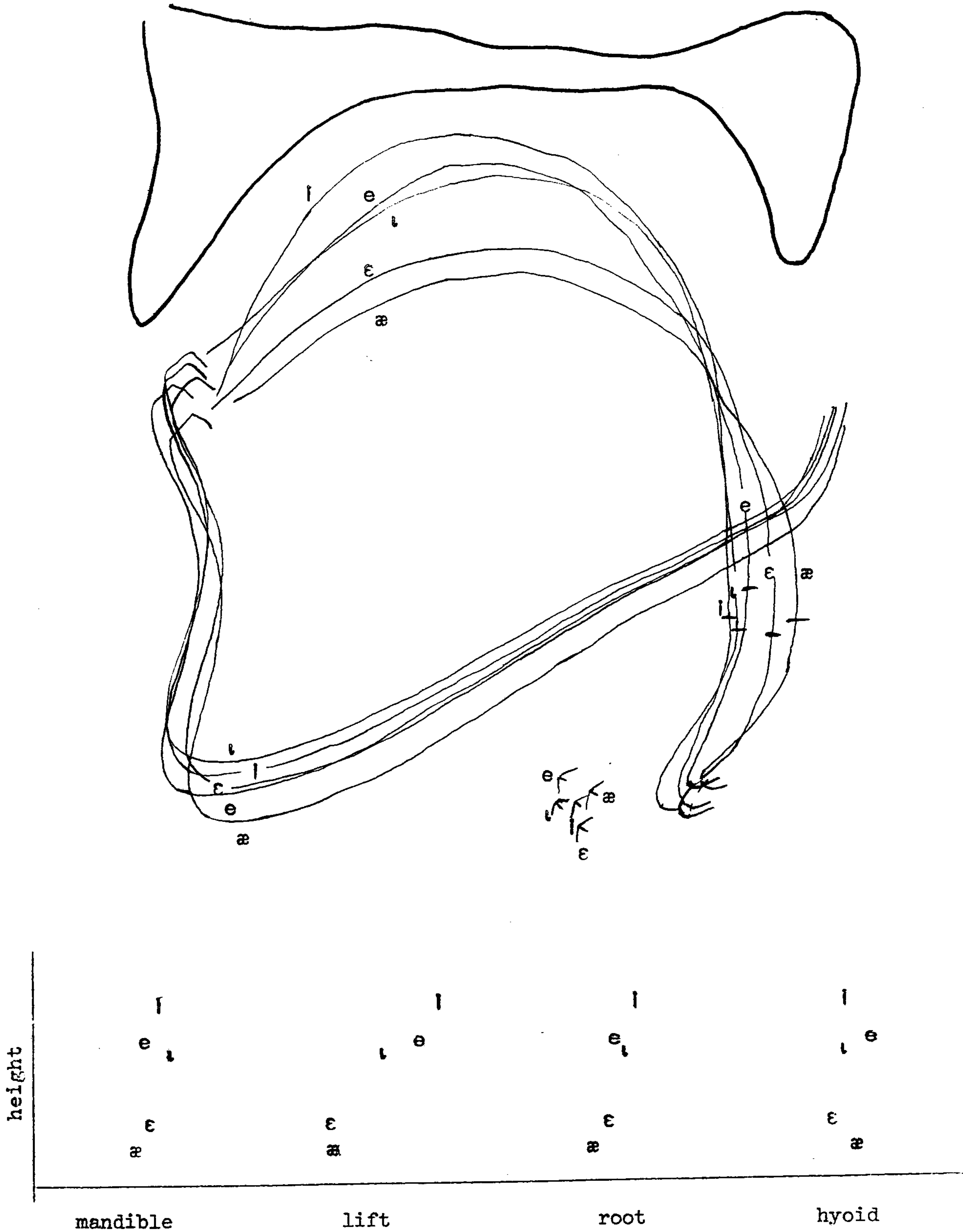


Figure 5.4 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 4.



5

Figure 5.5 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 5.

Figure 5.6 The tongue position in each of the five front vowels, and the relation between (on the ordinate) tongue height, and (on the abscissa) mandible height, tongue lifting relative to the mandible, tongue root advancement, and hyoid height for Subject 6.

degree of advancement of the tongue root, and the relative height of the hyoid. Each of the six subjects will be discussed separately, since (as may be seen from even a cursory glance at figures 5.1 through 5.6) there is a considerable degree of variation in the articulatory gestures used by the different subjects.

Subject 1, Figure 5.1. This subject steadily decreases the tongue height for each of the five front vowels. Accompanying these movements are almost equally steady increments of jaw opeing. This subject also controls the tongue position by increasing the pharynx width (advancing the tongue root) in accordance with the tongue height required except in the case of /ɩ/. But (as may be seen from the second graph) he makes little use of the possibility of lifting the tongue within the mandible. The hyoid bone does not seem to have been moved independently from the tongue root; it seems that this subject uses the genioglossus and geniohyoid muscles synergistically. But he also seems to at least maintain the downward pull of the thyrohyoid muscle so that as the hyoid bone is pulled forward it must also move downward. For this subject hyoid height and vowel height are inversely correlated.

Subject 2, Figure 5.2. This subject makes larger variations in tongue height than subject 1. But his jaw movements are about the same size. He makes considerable use of the tongue body lifting mechanism in the two tense vowels (those in *heed* and *hayed)*. He also has a greater pharynx width in these two vowels than in the other three. He has a very similar tongue shape in each of the lax vowels in *hid, head,* and *had,* and distinguishes between these three vowels largely by jaw position. As is the case for subject 1, the hyoid movement is similar to that of the root of the tongue, and is inversely correlated with tongue height. Note that this subject has a higher tongue position in /e/ than in /ɩ/.

Subject 3, Figure 5.3. This subject, like the previous one, distinguishes the tense vowels in *heed* and *hayed* from the corresponding lax vowels in *hid* and *head* by increasing the pharynx width and (to a lesser extent) by lifting the tongue within the mandible. But the graphs show that unlike the previous subject he distinguishes the two tense vowels from one another not so much by the lifting and advanced tongue root mechanisms, but by large variations in jaw position. The three lax vowels in *hid, head* and *had* are distinguished from one another both by large variations in jaw opening and by variations in the advancement of the tongue root. For the lax vowels these two mechanisms are more than enough to account for the variations in tongue height, and consequently, both for /ɩ/ and(to a lesser extent) for /ɛ/ there is less lifting of the tongue within the mandible than might be expected. The lifting feature is actually negatively correlated with tongue height for the lax vowels. Again, as with the first two subjects, the hyoid movements are in accordance with a synergistic action of the genioglossus and geniohyoid muscles, and hyoid height is negatively correlated with vowel height.

Subject 4, Figure 5.4. This subject has two distinct sets of jaw positions, one for the high vowels /i,ɩ / and the other for the non-high vowels /e,ɛ,æ/. From the graphs one may see that the members of each set are distinguished from one another principally by the lifting mechanism. Note that the differences in mandible position are so large that there have to be compensatory actions of the lifting mechanism; considering the high vowels /i,ɩ/ as a group and the low vowels /e,ɛ,æ/ as a group, this mechanism is negatively correlated with vowel height. The high vowels /i,ɩ/ are further distinguished from one another both by very small differences in jaw opening and by considerable differences in tongue root advancement; and among the non-high vowels /e,ɛ,æ/ there are also small differences in jaw opening distinguishing /e/ from /ɛ/, and in tongue root advancement distinguishing /ɛ/ from /æ/. The movements of the hyoid bone are more difficult to interpret; but, unlike the previous subjects, this subject does not have a systematic negative correlation between tongue height and hyoid height.

Subject 5, Figure 5.5. This subject is like subjects 2, 4, and 6 in that his vowel in *hayed* is higher than that in *hid*. But he is unlike any of the previous subjects in having a lower position of the jaw for /i/ than for /ɩ/ and a lower position for /e/ than for /ɛ/. In these pairs of vowels the differences in tongue height are in the same direction as those of other subjects; so in these pairs of vowels this subject has a *negative* correlation between tongue height and mandible height. There is compensation for the overly large jaw openings in /i/ and /e/ in that there is more use of the lifting mechanism in these vowels than in the others. In comparison with other subjects he does not have very large variations in pharynx width. The small pharynx width variations between the tense vowels /i,e/ and their lax counterparts /ɩ,ɛ/ may be due to the extreme use of the lifting action of the mylohyoid muscle, rather than the advancement of the tongue root by means of the action of the genioglossus muscle. As in the case of the previous subject, the movements of the hyoid are hard to interpret.

Subject 6, Figure 5.6. This subject makes the clearest articulatory distinction between the set of tense vowels and the set of lax vowels. He is like the previous subject in having a negative correlation between mandible height and tongue height in each of the tense-lax pairs /i-ɩ/ and /e-ɛ/. But he is unlike subject 5 in that he compensates for the comparatively lower jaw position in each of the tense vowels by large tongue root movements increasing the pharynx width, as well as by lifting the tongue. He distinguishes the lax vowels from one another by large variations in jaw opening, which are used in conjunction with some compensatory actions of the lifting mechanism. The tense vowels are also distinguished from one another by jaw position. The hyoid position is virtually the same for all vowels except for /e/, in which it is somewhat lower.

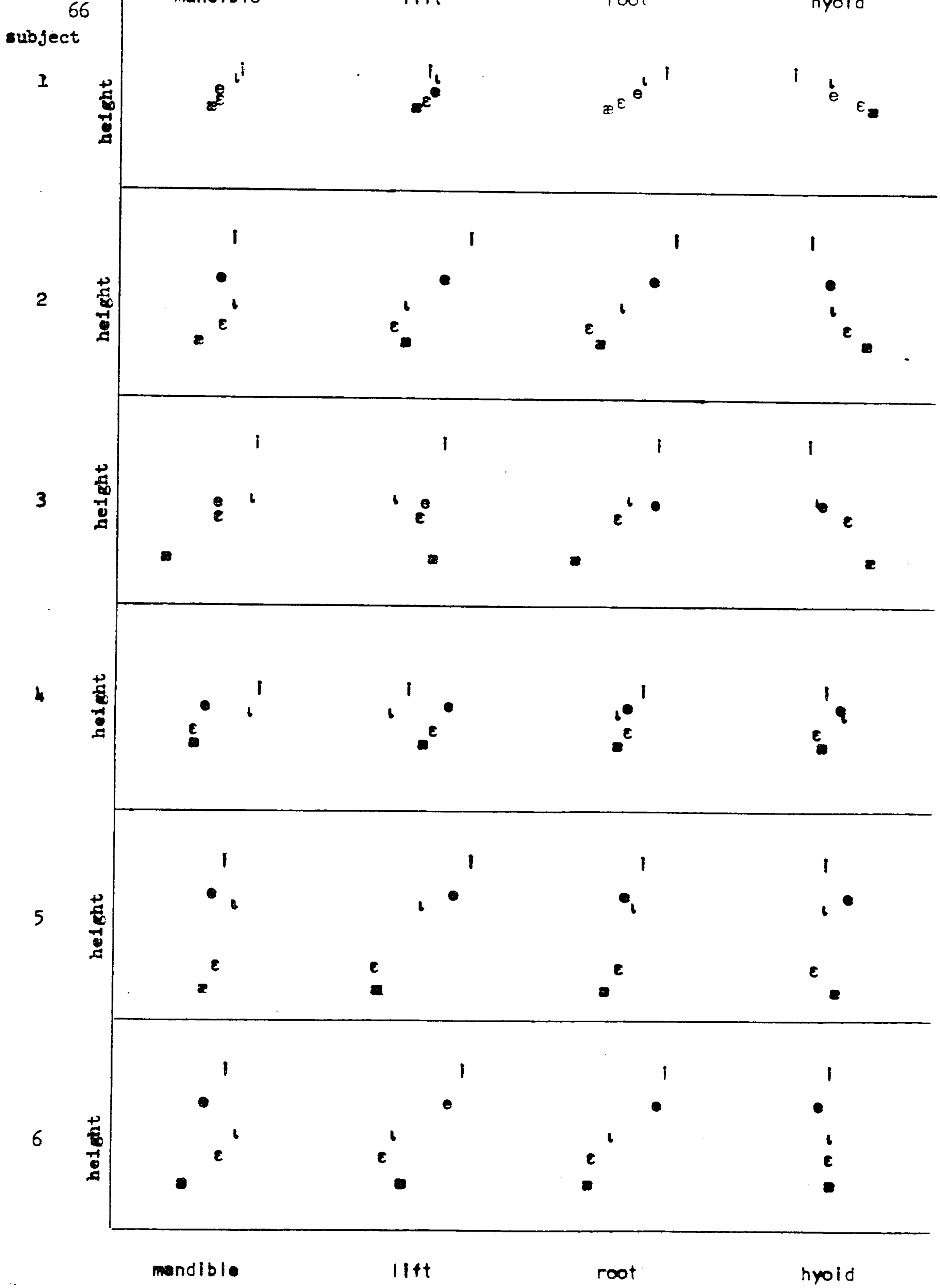


Figure 6. The relation between (on the ordinate) tongue height, and (on the abscissa) mandible height,tongue lifting relarive to the mandible, tongue root advancement, and hyoid height for all six subjects.

Table 1. Multiple correlation analyses of each subject, showing (a) regression coefficients, and (b) partial correlation coefficients for I Mandible height, advanced tongue root, and hyoid height, and II Tongue lifting within the mandible advanced tongue root and hyoid height, the dependent variable always being tongue height.

	S1		S2		S3		S4		S5		S6	
I	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Mandible height	.45	.73	-.43	-.29	.67	.59	.09	.70	-.35	-.87	.40	.96
Advanced tongue root	.46	.77	.33	.37	.68	.44	1.59	.99	3.74	.99	1.15	.99
Hyoid height	.09	.29	-1.89	-.59	.39	.15	1.36	.98	1.50	.99	1.14	.89
Mult. corr. coeff.	.9927		.9893		.9286		.9949		.9998		.9990	
II												
Tongue lifting	-.15	-.16	.65	.88	.65	.56	-.08	-.63	.28	.93	-.62	-.89
Advanced tongue root	.69	.80	-.14	-.31	-.21	-.14	1.71	.99	2.72	.99	1.88	.98
Hyoid height	.11	.26	-1.50	-.96	-1.6	-.61	1.50	.99	1.07	.98	1.94	.98
Mult. corr. coeff.	.9849		.9974		.9249		.9939		.9999		.9973	

Statistical analysis

The data we have been considering may be discussed in statistical terms. The first point to consider is the extent to which the tongue height can be predicted from the other four variables: jaw opening, pharynx width, tongue lifting, and hyoid height. As an aid in this discussion the data from Figures 5.1-5.6 is consolidated in Figure 6. We cannot determine a meaningful multiple correlation coefficient between the four variables and the tongue height. By our system of measurement the degree of jaw opening and the degree of tongue lifting within the mandible are bound to add up to exactly the recorded tongue height. If both of these variables are included in an analysis we will inevitably find that there is a perfect correlation between tongue height and those two measures. Accordingly in determining the extent to which our variables can be used to predict tongue height, we investigated first the relation between tongue height and the three variables jaw opening, pharynx width and hyoid height. We then made a similar comparison of the relation between tongue height and the three variables tongue lifting, pharynx width and hyoid height.

The top half of Table 1 shows the principal results of a multiple correlation analysis involving the three independent variables mandible height, advancement of the tongue root, and hyoid height; the lower half shows the results of a similar analysis involving tongue lifting within the mandible instead of mandible height. We will consider the individual partial correlation coefficients in a later version of this working paper. For the moment we merely wish to point out that for either set of three variables, the multiple correlation coefficient is always highly significant for each of the six subjects. This indicates that the four variables we have chosen are more than sufficient to account for essentially all the possible variations in tongue height for front vowels.

Articulatory mechanisms.

Another way of summarizing our results is by discussing them in relation to a number of points which have been raised in the literature. We will consider: (1) the effect of jaw opening; (2) the distinction between tense and lax vowels; (3) movements of the hyoid bone; and (4) the notion of a neutral position of the tongue.

Jaw opening. Lindblom and Sundberg (1971) have emphasized that the position of the jaw may be an important determiner of the shape of the vocal tract. Previous authors (eg Jones 1956) have described vowels in terms of the position of the highest point of the tongue, but have drawn diagrams which indicate that the jaw position is the same for all vowels. This view is clearly untenable. But so, in our opinion, is Lindblom and Sundberg's suggestion that the primary difference among the front vowels is simply one of relative jaw opening.

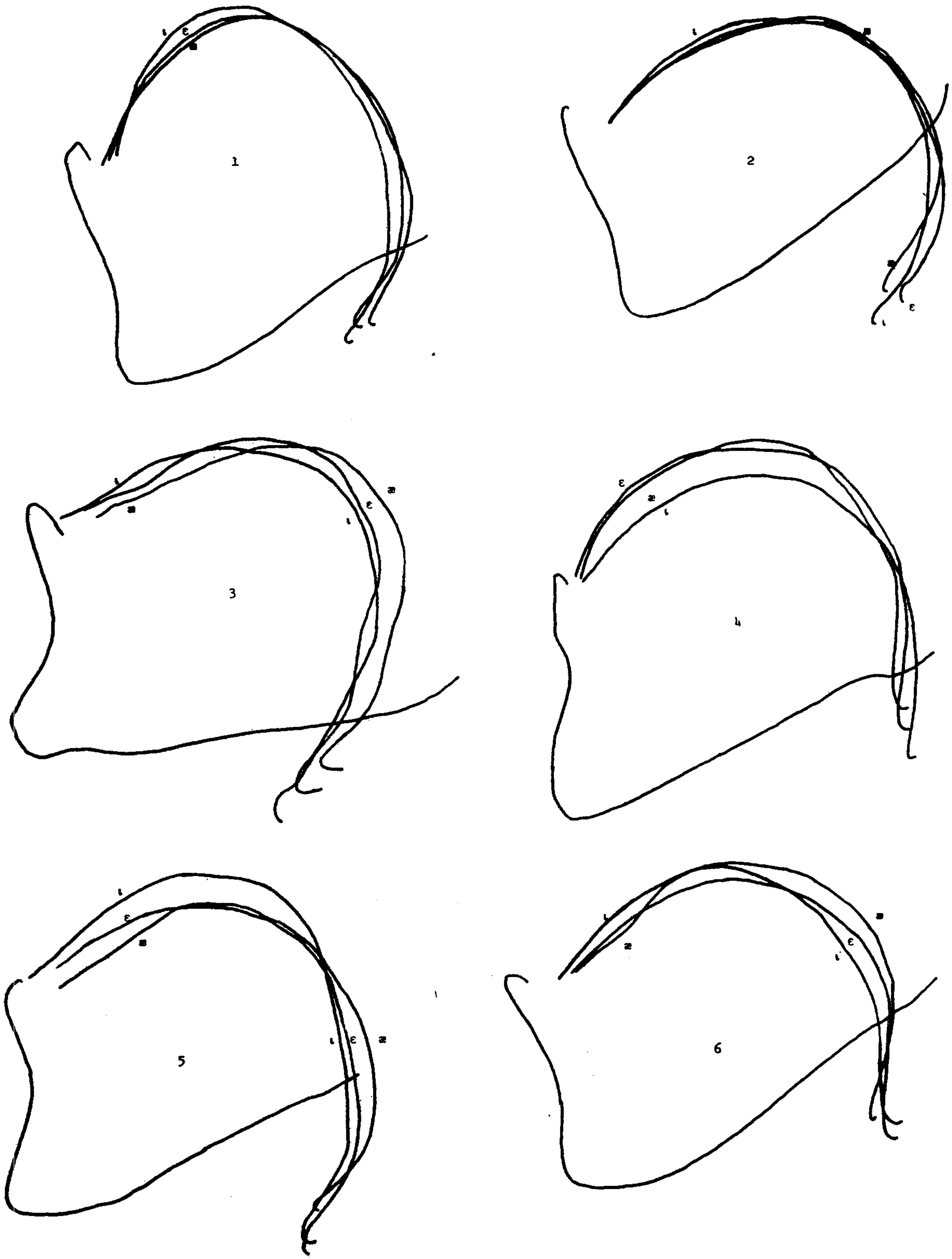


Figure 7. The lax vowels /ι, ε, æ/ superimposed onto a fixed mandible for each of the six subjects so as to show only the movement of the tongue (if any) with respect to the mandible.

To estimate the extent to which jaw opening is used by a subject Lindblom and Sundberg superimposed tracings of the tongue positions in different vowels onto a fixed mandible. If the mandible position were the only difference between members of a particular set of vowels, the tongue shapes of a given subject when superimposed in this way should lie on top of one another. They showed that, for their subject, the tongue shapes did remain constant with respect to the mandible throughout a set of front vowels. As a result they concluded that the degree of jaw opening was the most important difference between these vowels.

Examination of our data shows that none of our subjects behaved in exactly this way. Even if we restrict the discussion to differences among the lax vowels /ɩ,ɛ,æ/, then we find that only subject 2 (and perhaps subject 1) behave in the same way as Lindblom and Sundberg's subject. Figure 7 shows that the shapes of subjects 1 and 2 are fairly constant with respect to the mandible for all three lax vowels, and the differences between them can be mainly attributed to the degree of jaw opening. Subject 4 has similar shapes for /ɛ/ and /æ/ but a very different shape for /ɩ/; so he also could be said to be using only variations in jaw opening to distinguish between two of these three vowels. None of the other subjects has similar tongue shapes in any of these vowels. They cannot, therefore, be using the degree of jaw opening as the main method of distinguishing between front lax vowels.

Tense-lax distinctions. One possible correlate of this distinction which has been suggested (Perkell, 1971; quoting a personal communication from Halle and Stevens) is that tense vowels have an advanced tongue root. As we have seen, at least for some speakers, advancement of the tongue root correlates with vowel height, and may be used to differentiate all the front vowels from one another. So this suggestion concerning tense and lax vowels might have to be interpreted as meaning that a given tense vowel has a greater degree of advancement of the tongue root than is commensurate with its height. The previous discussion has indicated that it is in fact true that subjects 2, 3 and 6 separate the tense vowels /i,e/ from the lax vowels /ɩ,ɛ,æ/ by advancing the tongue root and thus increasing the width of the pharynx. But the other subjects do not distinguish between these two groups of vowels in this way. Subject 4 separates /i/ from /ɩ/ by advancing the tongue root but not /e/ from /ɛ/; subject 5 makes very little use of the tongue root mechanism; and subject 1 uses it as part of the mechanism for varying tongue height for all vowels, both tense and lax. The feature Advanced Tongue Root has been discussed by some of us at greater length elsewhere (Lindau et al. 1972).

Advancement of the tongue root is not the only way of bunching up the tongue. Accordingly we may usefully compare the tongue shapes in different vowels by superimposing the tongue shapes of the front vowels of each subject in such a way as to maximize the degree of agreement in the curvature of the tongue. The results of this procedure are shown in

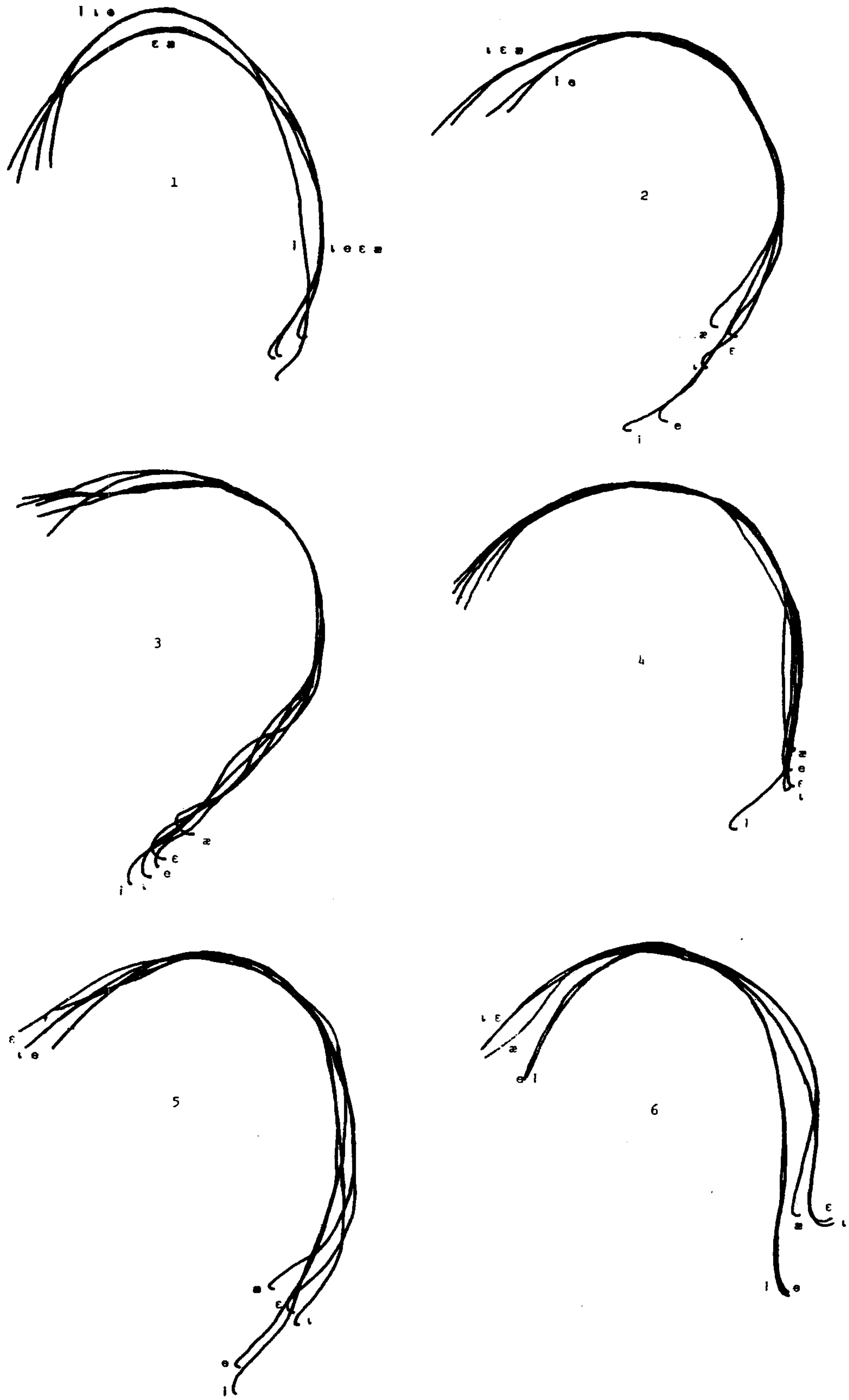


Figure 8. The tongue shapes in the front vowels of each subject, superimposed in such a way as to maximize the degree of agreement in the curvature of the tongue.

Figure 8. Subject 1 clearly has a different curvature in /i/, and also has a slightly greater curvature in /ɩ/ and /e/ (which are very similar to one another) than in /ɛ/ and /æ/. Subject 2 has a slightly more bunched tongue in /i,e/. Note also in this subject and several of the others that we achieve the degree of agreement in curvature by rotating the diagram of the tongue shape, so that the root is lower for high vowels. When the tongue is rotated in this way subjects 3 and 4 each have essentially only one tongue shape in all the front vowels. Subject 5 has slight differences between the tense and lax vowels, and subject 6 has a very clear difference in curvature for the two sets of vowels.

Hyoid-height. The hyoid bone is closely joined to the thyroid cartilage. Consequently raising the hyoid bone might well lead to an upward pull on the thyroid cartilage; and this, in turn, might lead to the vocal cords being stretched so that there is an increase in pitch. It has been suggested by Ladefoged (1964) that a mechanism of this kind would account for the increase in pitch which has often been noted in high vowels (Lehiste 1970). We have not yet measured the fundamental frequency of our subject's vowels, but if they have the normal increase in pitch then Ladefoged's explanation is erroneous in that for the majority of our subjects the hyoid position is inversely correlated with vowel height. We must admit that we find the movements of the hyoid bone (and, incidentally, the increases in pitch commonly observed in high vowels) somewhat mysterious. The hyoid obviously can be pulled downwards so that the whole body of the tongue is lowered. Movements of this kind can be used to compensate for an advancement of the tongue root and thus allow an increase in pharynx width without a concomitant increase in tongue height. This mechanism is used to differentiate classes of vowels in some languages (Lindau et al. 1972). Thus in Twi (an African language of the Akan family) there are two sets of vowels, the one being distinguished from the other by advancement of the tongue root. Within each of the sets of Twi vowels, the tongue height of each vowel is characterized by some other articulatory gesture, such as jaw opening. The Twi data indicate that in some languages variations in pharynx width and variation in jaw opening *can* be used independently to produce systematic linguistic distinctions. In these languages they cannot be used to compensate for one another.

Neutral position of the tongue. It has been suggested by Chomsky and Halle (1968) that the neutral position of the tongue is as in the vowel /ɛ/ as in *bed*. Our data give no support to the notion that the tongue should be in the neutral position in *bed*.

Auditory control of speech

The data we have been discussing clearly show that perceptually similar vowels may be produced by articulatorily different gestures. The vowels were all pronounced in normal utterances in which the subject was not constrained to behave in any particular way. Until we have examined further utterances by the subjects we cannot tell whether each subject consistently distinguishes between these vowels in the way that he did on the occasion when these articulations were photographed. It may well be that subjects sometimes use one articulation and sometimes another when making a given vowel.

We know from an experiment by Lindblom and Sundberg (1971) that a speaker can produce a given vowel with different articulatory gestures. These authors constrained the degree of jaw opening by putting a wedge between the subject's teeth. They showed that a subject could produce a given vowel with many different jaw openings: moreover the tongue must have been in the correct position for a given vowel before any sound was made, since acoustic measurements showed that the requisite vowel quality was present during the first period of the sound; there can have been no question of the subject modifying the vowel quality as a result of auditory feedback.

In considering how we can account for all the facts about the production of vowels, it is useful to recall what the Haskins Laboratory group have said about the perception of consonants. According to Liberman et al (1962) "articulatory movements and sensory feedback (or, more likely, the corresponding neurological processes) become part of the perceiving process, mediating between the acoustic stimulus and its ultimate perception". We would feel that it is equally true that acoustic properties and their sensory counterparts (or, more likely, the corresponding neurological processes) become part of the producing process, mediating between the possible articulatory gestures and their ultimate production. Articulatory gestures are organized within the sensory-motor cortex. The information available at this level includes not only sensory feedback information revealing dynamic and static information about the state of the muscles, but also neural projections from the auditory cortex. A speaker may be able to use an auditory image to arrive at a suitable tongue position without reference to any stored pattern of articulatory gestures.

This hypothesis is supported by the facts of language acquisition. When a child learns to talk he has no direct access to knowledge of the positions of the tongue which may be involved. He is constrained by physiological limitations so that often it is possible to make a given sound only by a specific articulatory gesture; and in these circumstances he may build up a specific articulatory image for a given phonetic unit. But in the case of vowels it is quite possible that he may learn to produce a given auditory image without using a specific articulatory image.

Phonological studies (Ladefoged 1971) also indicate that on some occasions the sound patterns which can be observed in languages are more easily explicable by reference to acoustic rather than to articulatory data. This is particularly true for phonetic and phonological phenomena involving vowels. Thus Liljecrants and Lindblom (in press) has shown how it is possible to account for the vocalic patterns which occur in different languages by reference to their acoustic properties.

Phoneticians usually describe vowels in terms of the position of the highest point of the tongue, and in terms of the degree of lip rounding. These traditional terms are best regarded as labels for auditory phenomena, rather than precise descriptions of vocal tract shapes. They are reasonably appropriate in discussions of the front vowels we have been considering. But they cannot be so easily applied to all vowels. In fact, if we consider the total range of possible vowel sounds, then two of the traditional terms are more clearly correlated with acoustic rather than articulatory measures. Thus formant one is correlated with so-called vowel height; and the distance between formant one and formant two is correlated with the traditional front-back dimension. But there is no uniform acoustic correlate for the third traditional dimension, the degree of lip rounding. Nevertheless the degree of lip rounding is undoubtedly an important attribute of vowels, playing a major role in many phonological rules. Since it has no uniform auditory or acoustic correlates, it is at least possible that it is simply an articulatory property. In summary, vowels cannot be adequately described using only acoustic terms; nor can they be described entirely in articulatory terms. Listeners and speakers (and languages) organize two aspects of vowel quality primarily in auditory/acoustic terms and have a third feature which is organized primarily in articulatory terms.

Considerations of this kind lead us to suggest that it is inappropriate to speak of a dichotomy between auditory and articulatory theories. At the sensory-motor level of the cortex, which is where phonetic units are encoded for the production of speech, both articulatory and auditory images are available; and similar images are probably used in the process of decoding incoming speech signals. Human beings are not maximally efficient encoders and decoders, using only the minimum amount of information. Speech is a highly redundant system at all levels, and is best described in terms of an auditory-motor theory of speech production and perception.

References

Bowman, J.P. (1971) *The Muscle Spindle and Neural Control of the Tongue. Implications for Speech.* Illinois; Charles C.Thomas.

Chomsky, N. and M. Halle. (1968) *The Sound Pattern of English.* New York Harper and Row.

DeClerk, J.L., L.S. Landa, D.L. Phyfe, and S.I. Silverman. (1965) "Cinefluorography of the vocal tract" *Proc. 5th Int. Cong. Acoust.* Liege.

Hirano, M. and T. Smith (1967) "Electromyographic study of tongue function in speech" A preliminary report. *Working Papers in Phonetics, 7.* 45-56. UCLA.

Jones, D. (1956) *An Outline of English Phonetics.* (8th ed.), Cambridge, Heffer & Sons.

Ladefoged, P. (1964) "Some possibilities in speech synthesis" *Language & Speech* 7. 205-214.

_____. (1971) "Phonological Features and their Phonetic Correlates." *Working Papers in Phonetics* 21. 3-12. UCLA. (forthcoming in *Journ. of the Intern. Phonetic Assoc.)*

Lehiste, I. (1970) *Suprasegmentals.* Cambridge: MIT Press.

Liberman, A.M. (1957) "Some results of research on speech production" *Journ. Acoust. Soc. of Amer.* 29, 117-123.

_____, F.S. Cooper, K.S. Harris, P.F. MacNeilage. (1962) "A Motor Theory of Speech Perception" *Proc. Speech Comm. Seminar.* Stockholm: Royal Institute of Tech. 1963, D3.

_____, F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy (1967) "Perception of the speech code" *Psychological Review* 74:6. 431-461.

Liljecrants, J. and B. Lindblom (in press) "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast" *Language.*

Lindau, M., L. Jacobson, and P. Ladefoged. (1972) "The Feature Advanced Tongue Root" *Working Papers in Phonetics 22,* UCLA.

Lindblom, B. and J. Sundberg. (1971) "Acoustic Consequences of Lip, Tongue, Jaw, and Larynx Movement." *Journ. Acoust. Soc. of Amer.* 50:4, 1166-1179.

Minifie, F.D. T.J.Hixon, C.A.Kelsey, and R.J.Woodhouse. (1970) "Lateral Pharyngeal Wall Movement during Speech Production" *Journ. Sp. & Hear. Re.* 13, 584-594.

Perkell, J. (1971) "Physiology of Speech Production: A Preliminary Study of Two Suggested Revisions of the Features Specifying Vowels" *Quart. Prog.Rep. No.*102, *Res. Lab. Electron.* M.I.T., 123-139

Smith, T. (1971) "A Phonetic Study of the Function of the Extrinsic Tongue Muscles" *Working Papers in Phonetics* 18. UCLA.

Strong, L.H. (1956) "Muscle Fibers of the Tongue" *Anatom. Record 126,* 61-79.

The Feature Advanced Tongue Root

Mona Lindau, Leon Jacobson, and Peter Ladefoged

[Expanded version of a paper given at the 82nd meeting of the Acoustical Society of America, October 1971]

Introduction. This paper discusses some phonetic mechanisms that are used to distinguish between sets of vowels. In some African languages two sets of vowels are used in what is called vowel harmony: only vowels belonging to the same set will occur in the same word. In a typical vowel harmony language there are eight or nine distinct vowels, four or five occuring in one set of words, and the other four or five occuring in the other.

The feature which distinguishes between the two sets of vowels has been described in terms of a number of phonetic mechanisms. Christaller (1875) distinguishes the two sets by 'tongue height' for Twi; while Welmers (1946) introduces the suprasegmental feature 'raising' for this distinction in Fante. Berry (1955) claims that the vowel sets exhibit different phonation types--the tense vowels being accompanied by a "dull and breathy voice" and the lax vowels by a "tight or creaky quality of voice." The terms 'tense' and 'lax' were used by Schachter and Fromkin (1968) as phonological classifiers with no claims regarding their phonetic content. Chomsky and Halle (1968) introduce the feature 'Covered' and define 'covered' sounds as "...produced with a pharynx in which the walls are narrowed and tensed and the larynx raised;" 'uncovered' sounds are described as "...produced without a special narrowing and tensing in the pharynx" (p. 315). No examples are given so it is ambiguous which set 'covered' is supposed to represent, particularly since Chomsky and Halle obfuscate the issue by associating the dull quality with covered vowels. Where the dull or breathy phonation is mentioned elsewhere it refers to the set that also is tense, or raised, or has a wider pharynx.

Ladefoged (1964) and Stewart (1967) put forward the hypothesis that vowel harmony is conditioned by differences in pharynx width controlled by advancing the tongue root. They point out that the differences between the two sets of vowels in some languages cannot be satisfactorily described in terms of tongue height. Halle and Stevens (1969) support this notion and suggest a revision of the feature system by replacing the ill-defined feature Covered with the feature Advanced Tongue Root.

Halle and Stevens (according to Perkell, 1971) recently suggested that the feature Advanced Tongue Root also substitute for the feature Tense, at least for nonlow vowels. The feature Tense is commonly used in the description of many languages, such as English, which do not have vowel harmony systems of the kind described above. Evidently Halle and Stevens argue that the same phonetic mechanism, advancing the tongue root, is used in both English and West African languages. We presume this claim can be generalized to other Germanic languages with vowels similar to those of English, and to those East African languages that have a vowel harmony similar to that of West African languages.

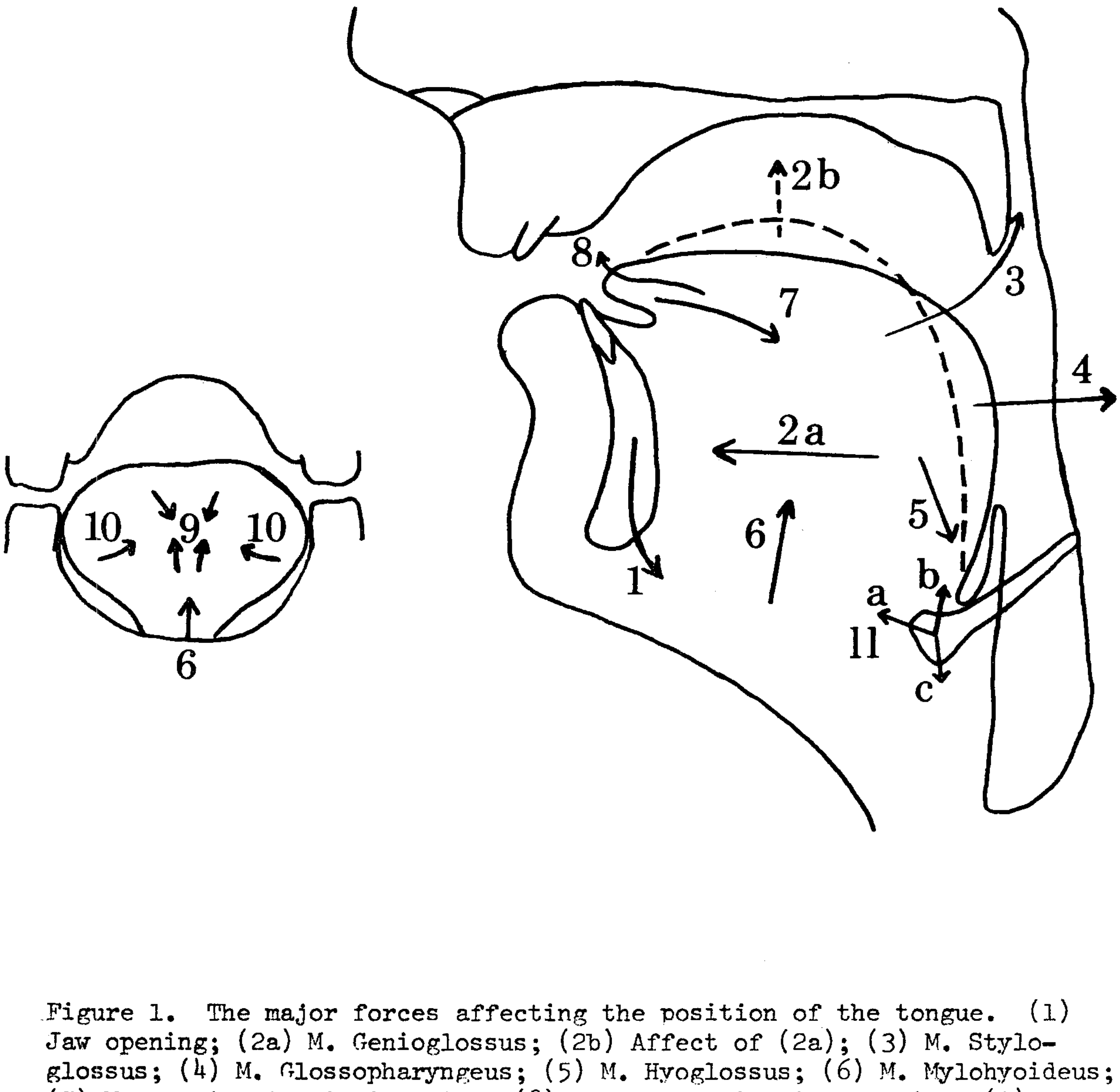


Figure 1. The major forces affecting the position of the tongue. (1) Jaw opening; (2a) M. Genioglossus; (2b) Affect of (2a); (3) M. Styloglossus; (4) M. Glossopharyngeus; (5) M. Hyoglossus; (6) M. Mylohyoideus; (7) M. Longitudinalis inferior; (8) M. Longitudinalis superior; (9) M. Verticalis; (10) M. Lateralis; (11) Forces affecting the hyoid: (a) M. Geniohyoideus; (b) M. Stylohyoideus; (c) M. Thyrohyoideus.

In summary, therefore, we may note that there are two claims which can be made about the feature Advanced Tongue Root. 1) It is a feature which distinguishes between vowels in some way other than the ways in which other features such as High and Low can be used to distinguish between vowels. 2) The so-called tense vowels in English and other Germanic languages differ from the so-called lax vowels in the same way as the vowels differ in the vowel harmony sets in African languages.

Before we can evaluate the hypothesis that Advanced Tongue Root is a feature with phonetic correlates which are distinguishable from those of other features such as High and Low, we must consider the ways in which the height of the tongue can be varied. Figure 1 is a simplified diagram of the major forces affecting tongue shape.

Tongue height can be varied by movements of the jaw: raising and lowering the mandible will differentiate tongue heights. Since the effective hinge of the mandible is above and behind the tongue, lowering the mandible will necessarily lower the front part of the tongue more than the back part of the tongue. It will also cause the back part of the tongue to retract, so there is less space for variation of the sagittal cross-dimension of the pharynx for low vowels than for nonlow vowels. We have so far excluded low vowels in this investigation.

Tongue height can be varied by the pulling forward of the tongue root by the genioglossus muscle. A pull in this direction compresses the tongue within the mandible. But since the mass within the mandible must remain constant, the effect of simply pulling the tongue root forward is that the front of the tongue is displaced upward as indicated by the dashed position of the tongue, and the arrow marked 2b in figure 1.

Tongue height can be varied by the action of the mylohyoid, which lifts the whole body of the tongue up from the mandible, and the bunching up and resultant lifting of the tongue by the styloglossus muscle, which varies tongue height for back vowels by an upward and backward pull.

In theory then, a speaker can control tongue height by the use of three forces, either singly or in combination: opening the jaw, advancing the tongue root, lifting the tongue.

Now, returning to the first part of the hypothesis, we can see that in order to demonstrate that movement of the tongue root serves some function other than varying tongue height, one would have to show that the variations in tongue height in a given set of vowels can be accounted for in some other way. To demonstrate the second part of the hypothesis, the part that was suggested neither by Ladefoged nor Stewart, but only by Halle and Stevens, the independence of tongue root advancement from height of the tongue would have to be shown for Germanic languages as well as for African languages.

<u>Analysis</u>. In order to investigate the feature Advanced Tongue Root, we analyzed radiographic data from two West African languages, two East African languages, and two European languages. For the West African languages we used two languages of the Akan group, the Akyem dialect of

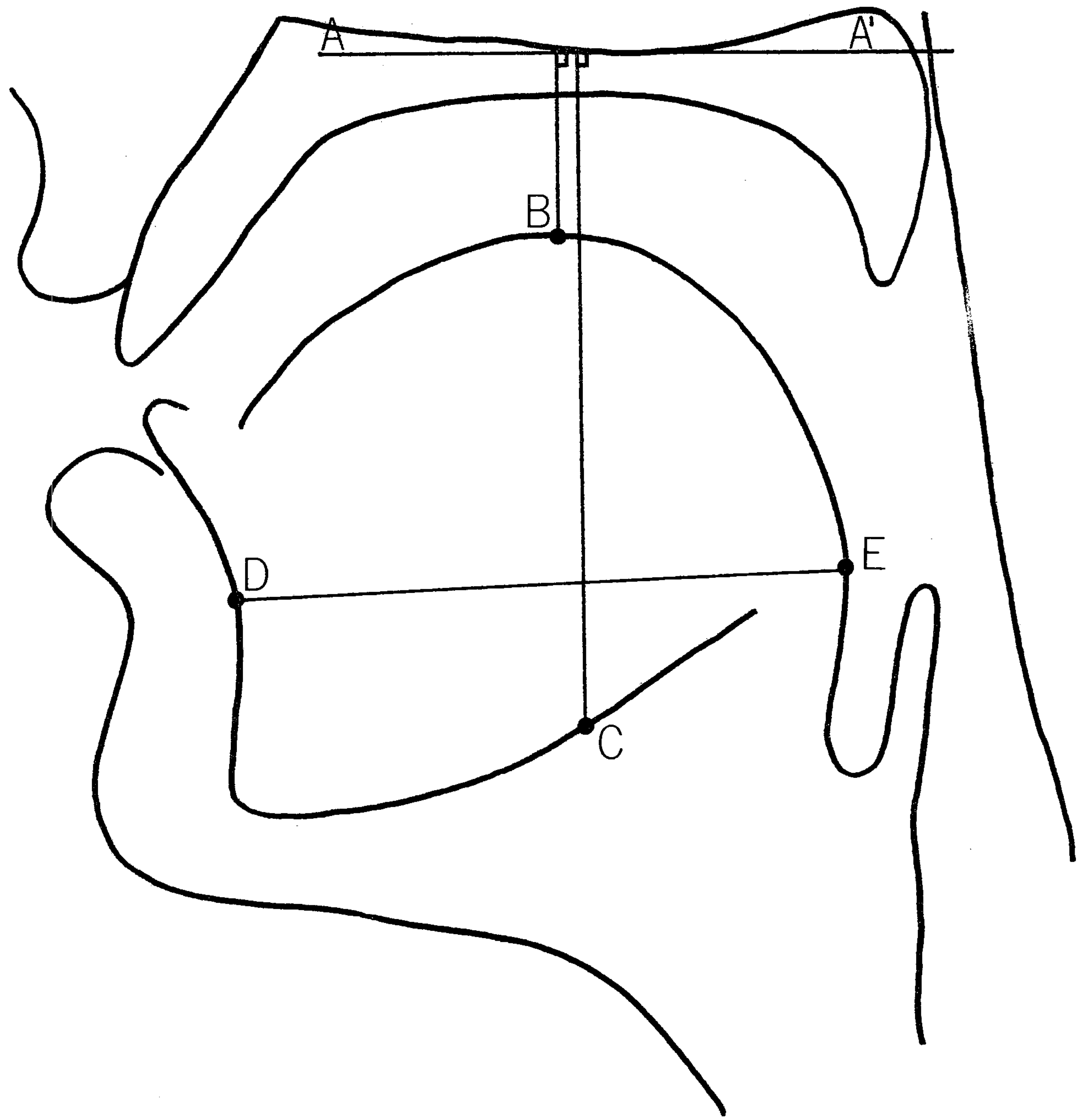


Figure 2. Points and measurements used in quantifying articulatory gestures.

Asante Twi and Igbo. The East African languages were Dho-Luo (Western Nilotic) and Ateso (Eastern Nilotic); these two languages are not related to the West African languages of this study, and are only distantly related to each other. The European languages were English and German.

For Twi and Ateso the data were obtained by radiography of sustained vowels. Both subjects had had some phonetic training, and we ensured that the sustained vowel was in fact of the intended vowel quality.

For Dho-Luo, the data were obtained by cineradiography. The text consisted of a list of words illustrating the harmonizing vowels. Simultaneous oscillomink recordings were made of the audio signal and of the frame counter attached to the camera. A steady state portion of each vowel was located on the film and a tracing was made of that frame. For Igbo we used the cineradiography data described in Ladefoged (1964). All these African languages exhibit vowel harmony so that a word contains vowels from one or the other of two sets. In this paper we will refer to the two sets as 'tense' and 'lax'. We emphasize that we use these terms only as labels for separating the sets.

For German our data consists of the X-rays published by Wängler (1961), while our data for American English consist of tracings from a cineradiographic film of six speakers saying sentences illustrating ten English vowels (Ladefoged, et. al., 1971).

Figure 2 shows the points used for measuring articulatory gestures. The starting point for two of the measurements is an arbitrary plan A-A', parallel to the floor of the nasal cavity. The highest point of the tongue (marked B) is defined in accordance with traditional phonetic practice as the highest point with reference to an arbitrary plane parallel to the floor of the nasal cavity (A-A'). Point B is also usually close to the point of maximum constriction of the vocal tract. Our measure therefore defines a critical point in the determination of the acoustic structure of the sound, as well as a point with articulatory significance.

The degree of jaw opening is determined by taking an arbitrary point on the mandible (marked C) and measuring the distance A-C. Point C is chosen so that it is immediately below the median position of B in the vowels under consideration for each speaker. In this way we can regard the distance A-C as a true index of the influence of jaw opening on tongue height. We also measure B-C to get an indication of how much the tongue is lifted within the mandible. The degree of tongue root advancement is determined by measuring the distance between an arbitrary point E on the surface of the tongue 18mm above the root of the epiglottis, and a fixed point D on the anterior aspect of the mandible near the origin of the genioglossus muscle on the posterior aspect of the mandible.

The measurements are on the same scale for each subject, but only relative comparisons can be made across subjects. The results of the measurements were plotted with tongue height on the abscissa against jaw opening, lifting, and advanced tongue root on the ordinate (figures 4, 7, and 8). The vowel symbols are paired in terms of tense and lax, the lax ones being marked with a subscripted comma. The correlation between tongue height and each of the other features shows to what extent that feature contributes to tongue height. For example, a nonsignificant

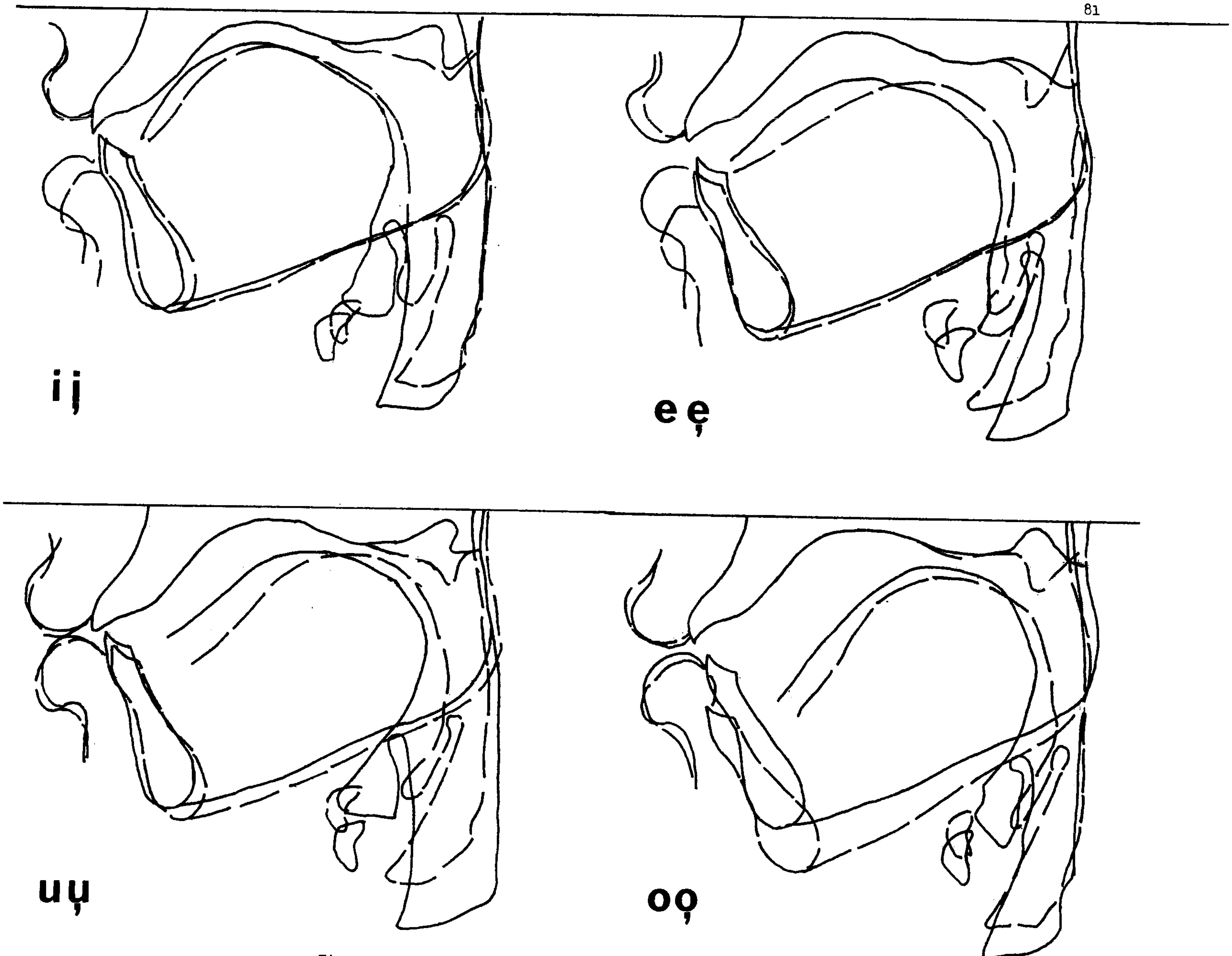


Figure 3. Tracings from radiology data showing the articulatory positions of the two sets of vowels in Twi. The broken lines are the lax vowels. Subscripted commas denote the lax vowels of a pair.

correlation between tongue height and advanced tongue root demonstrates that advancing the root of the tongue does not contribute to the tongue height, in which case the advanced tongue root may constitute a separate articulatory mechanism from that of tongue height. On the other hand, a significant correlation between tongue height and advanced tongue root means that the advancing of the tongue root contributes to raising the height of the tongue and therefore it is not a separate feature.

Results. Figure 3 shows the articulatory positions of the two sets of vowels in Twi.* The tracings are made with reference to the maxilla. In each pair the pharynx is considerably widened by advancing the root of the tongue in the tense vowel. A slight lifting of the tongue occurs with the advanced tongue root, only in the case of /o/. The other pairs of vowels have similar tongue heights, so it seems very likely that the advanced tongue root is not employed to control tongue height but can be used to control tenseness.

Figure 4 shows the functional relationships between tongue height and jaw opening, lifting, and advanced tongue root of the Twi vowels. The only significant correlation is between tongue height and jaw opening (correlation coefficient 0.75, $p < 0.05$). Jaw opening is the main control of tongue height. The lifting factor does not contribute to the height of the tongue, but it is of interest to point out that, except in the case of /i/ and /ɨ/, the speaker tends to use lifting of the tongue body antagonistically to jaw opening to control tongue height. This is illustrated by the negative correlation between lifting and jaw opening in Figure 5. When /i/ and /ɨ/ are excluded from consideration, the negative correlation is significant (correlation coefficient 0.94, $p < 0.002$). The speaker trades off an increase in jaw opening with an increase in tongue lifting; that is, the lower the jaw is, the higher the body of the tongue tends to be lifted above the mandible. In controlling tongue height with jaw opening this speaker apparently overdoes the gesture to some extent, so the lifting mechanism is employed in compensation. (The notion of compensatory articulations between jaw opening and tongue shape was recently discussed by Lindblom and Sundberg (1971).

Although there is a considerable range of variation of the tongue root, it does not contribute significantly to the tongue height. Advancing the tongue root is clearly used to separate tense and lax vowels in Twi. It can be seen from the graph in Figure 4 that all the tense vowels have a more advanced tongue root than all the lax vowels. A matched t-test shows that the difference in tongue root advancement between tense and lax vowels is significant ($p < 0.025$). Associated with the advanced tongue root of the tense vowel set there is also a consistent lowering of the position of the hyoid and the larynx. The gesture of advancing the tongue root seems thus to involve at least the genioglossus, the sternohyoid, and the sternothyroid muscles.

* This is not the set of all vowels in Twi. We have excluded the low vowels.

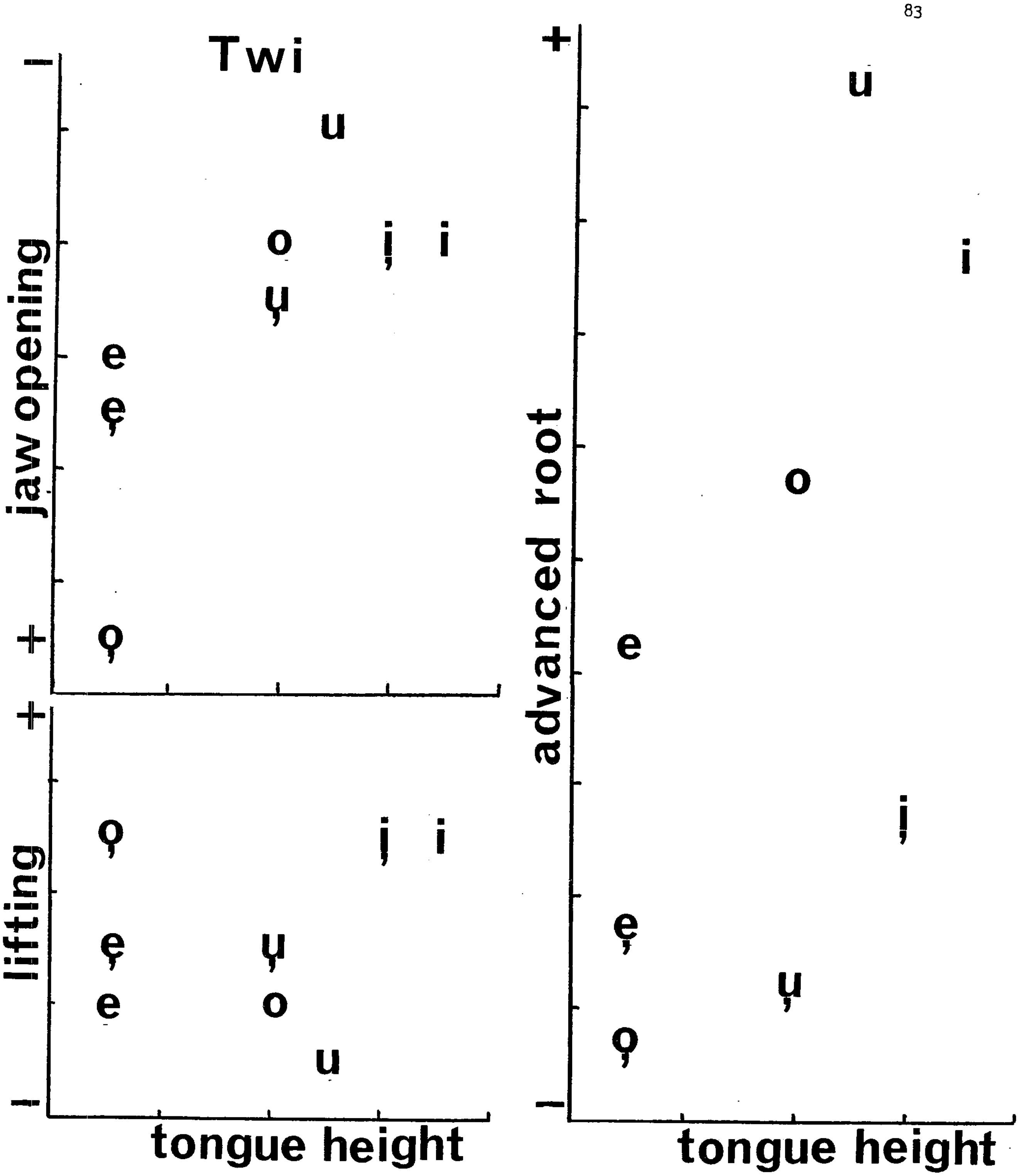


Figure 4. Tongue height related to jaw opening, lifting, and advanced tongue root in Twi. The minus sign (-) denotes the minimum value on the scale, and the plus sign (+) the maximum value.

The original data published by Ladefoged (1964); here reproduced in Figure 6 of Igbo vowels indicate that the harmonizing set of the high vowels are partly differentiated by the mechanism of advanced tongue root. The nonhigh vowels do not all use this feature, however. Instead, /ɛ/ and /a/ are differentiated by jaw opening.

Figure 7 shows tongue height related to jaw opening, lifting, and the advancement of the tongue root, for Dho-Luo. Here the situation for control of tongue height is more complex. The correlations between tongue height and jaw opening, and between tongue height and lifting are both very small and nonsignificant (correlation coefficients 0.3, $p < 0.2$ and 0.6, $p < 0.2$, respectively); and there is no significant correlation between tongue height and advanced tongue root. How does this speaker control tongue height? In the plot of tongue height against lifting, the tense [u] is the vowel that deviates most from the regression line. If tense [u] is disregarded the correlation between tongue height and lifting improves and is highly significant (correlation coefficient 0.9, $p < 0.002$). Except for tense [u] then, tongue height is controlled by lifting.

Excluding any one of the most deviant vowels in the plot of tongue height against jaw opening does not improve the correlation between these two variables. However, if all the back vowels are considered in isolation, they are found to be almost on a straight line. So for back vowels, jaw opening also contributes to attaining the required height.

The difference in advanced tongue root between tense [u] and lax [u̜] is considerably more than between any of the other pairs. By advancing the tongue root more for the tense [u] the speaker compensates for the low lift above the mandible. In summary, tongue height for front vowels in Dho-Luo is attained by lifting; tongue height of back vowels except [u] by lifting and jaw opening; and tongue height of [u] by jaw opening and advanced tongue root.

Again, advancement of the tongue root separates the two classes of vowels in this language. The difference between the tense and lax vowels is significant ($p < 0.05$). Together with the lack of significant correlation between tongue height and tongue root it could support a claim that the distinctive feature for the two harmonizing sets of vowels in Dho-Luo is Advanced Tongue Root.

The results from Ateso are shown in Figure 8. Both the correlations between tongue height and lift, and between tongue height and advanced tongue root are high (correlation coefficient are 0.9 and 0.8, respectively) and at significant levels ($p < 0.01$ for both). Jaw opening is not used systematically for varying tongue height (correlation coefficient 0.5, ns). Note also that this speaker makes use of compensatory articulation. When the jaw opening is too small in [e̜] the tongue is lifted more than expected. Tongue height in Ateso is a result of the combined effects of

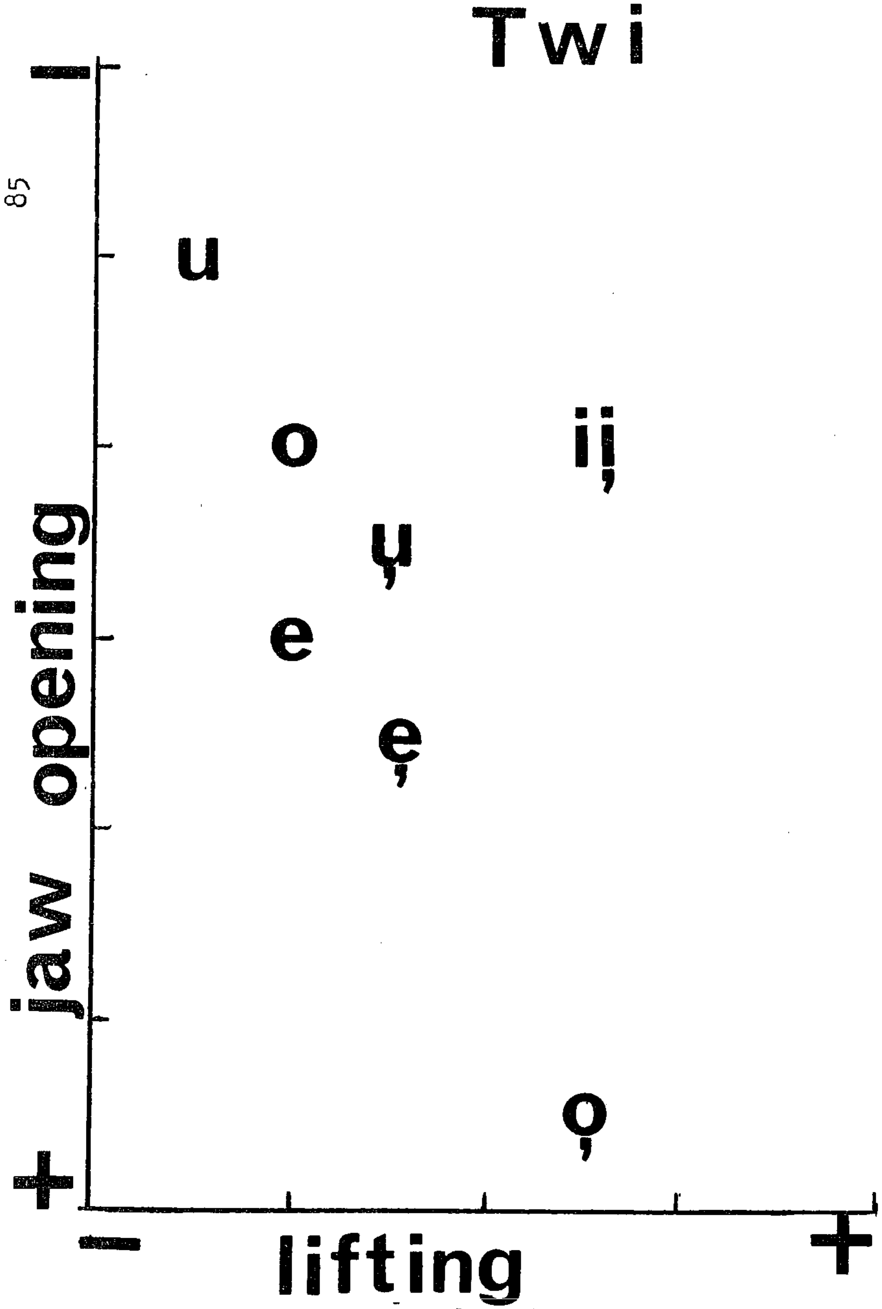


Figure 5. Relation between lifting and jaw opening in Twi.

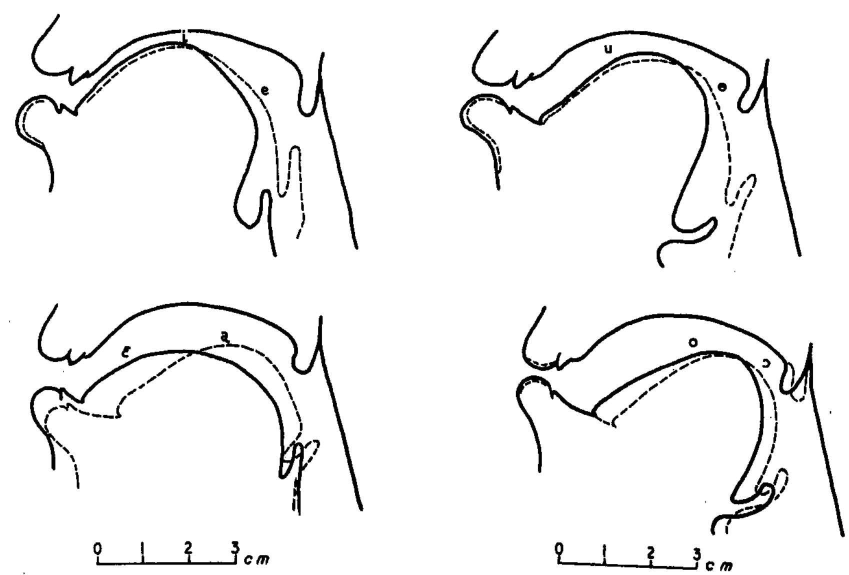


Figure 6. Tracings from single frames in a cine-radiology film showing the tongue positions in the two sets of Igbo vowels. (= Fig. 12 in Ladefoged 1964).

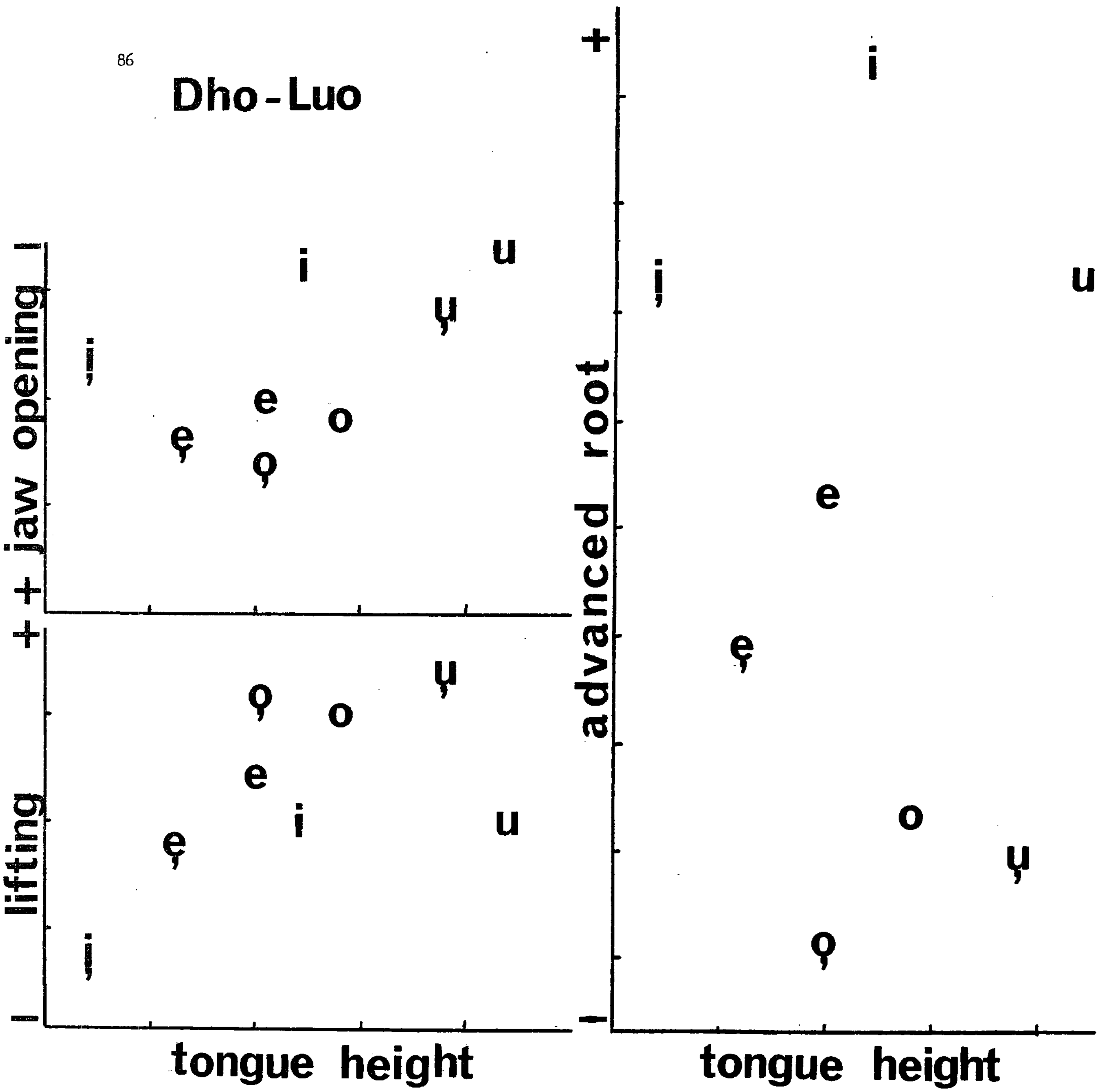


Figure 7. Tongue height related to jaw opening, lifting, and advanced tongue root in Dho-Luo.

lifting and advanced tongue root. This is an example of a language where the vowel harmony sets are not distinguished by any separate mechanism of the tongue root, but all vowels differ by the height mechanism only.

In Twi an advanced tongue root is accompanied by a lowered hyoid bone and larynx. This does not happen in Ateso. The heights of the hyoid bone and larynx remain the same for tense and lax vowels of the same pair. Our data from Igbo and Dho-Luo do not include the hyoid and the larynx.

Figure 9 shows the results for German. In this language the tense vowels differ from the lax vowels in quantity as well as quality: the tense vowels are noticeably longer than the lax vowels, as well as having a slightly higher tongue position. Before we can say that the lax vowels and the tense vowels differ in terms of the feature Advanced Tongue Root,and not by the feature of height, we have to see whether advancing of the tongue root is the general mechanism for making all height distinctions in German.

The results are similar to those of Ateso. The subject does not use jaw opening in any consistent way in achieving different tongue heights. Tongue height is attained by lifting and by advanced tongue root, both of which are strongly correlated with tongue height (both correlation coefficients are 0.9, $p< 0.002$). The difference in advanced tongue root between the tense and lax vowels is nonsignificant. Note too, that the range of variation of advanced tongue root in German is only fractionally larger than that of tongue height, while for Twi and Dho-Luo the variation of advanced tongue root was considerably larger than that of tongue height. These two facts indicate that advanced tongue root is not a separate feature in German. The mechanism is used (at least by this one speaker), but all it does is contribute to tongue height. The difference in vowel quality between long and short vowels in German is simply one of vowel height, which is what some investigators of phonetics have suggested (e.g. Wängler).

An account of our analysis of the vowels of six speakers of American English is presented in more detail elsewhere (Ladefoged, et al., 1972). The main result relevant to this paper is that there is a substantial variability in the mechanisms used to distinguish between vowels. Tongue height is attained by different combinations of jaw opening, lifting, and advanced tongue root for different speakers. Three of our six speakers do indeed separate out the vowels in <u>heed</u> and <u>hid</u>, and the vowels in <u>hayed</u> and <u>head</u>, by making the vowel in the first word of each pair with a more advanced tongue root. But even these three speakers do not all distinguish between a pair of back vowels such as those in <u>who'd</u> and <u>hood</u> in the same way. And the other three speakers do not make consistent differences between any pair of tense and lax vowels solely by advancement of the tongue root.

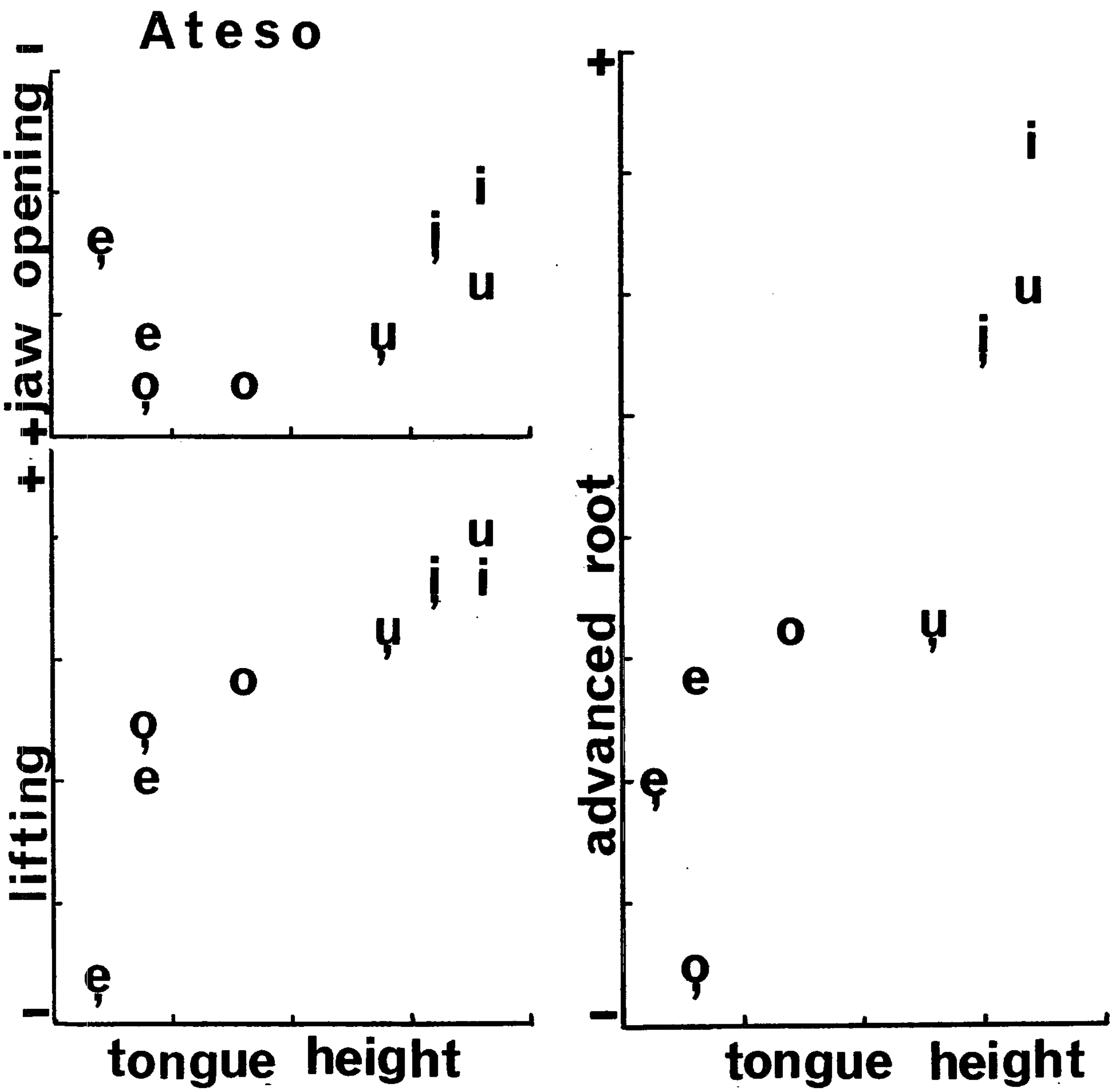


Figure 8. Tongue height related to jaw opening, lifting, and advanced tongue root in Ateso.

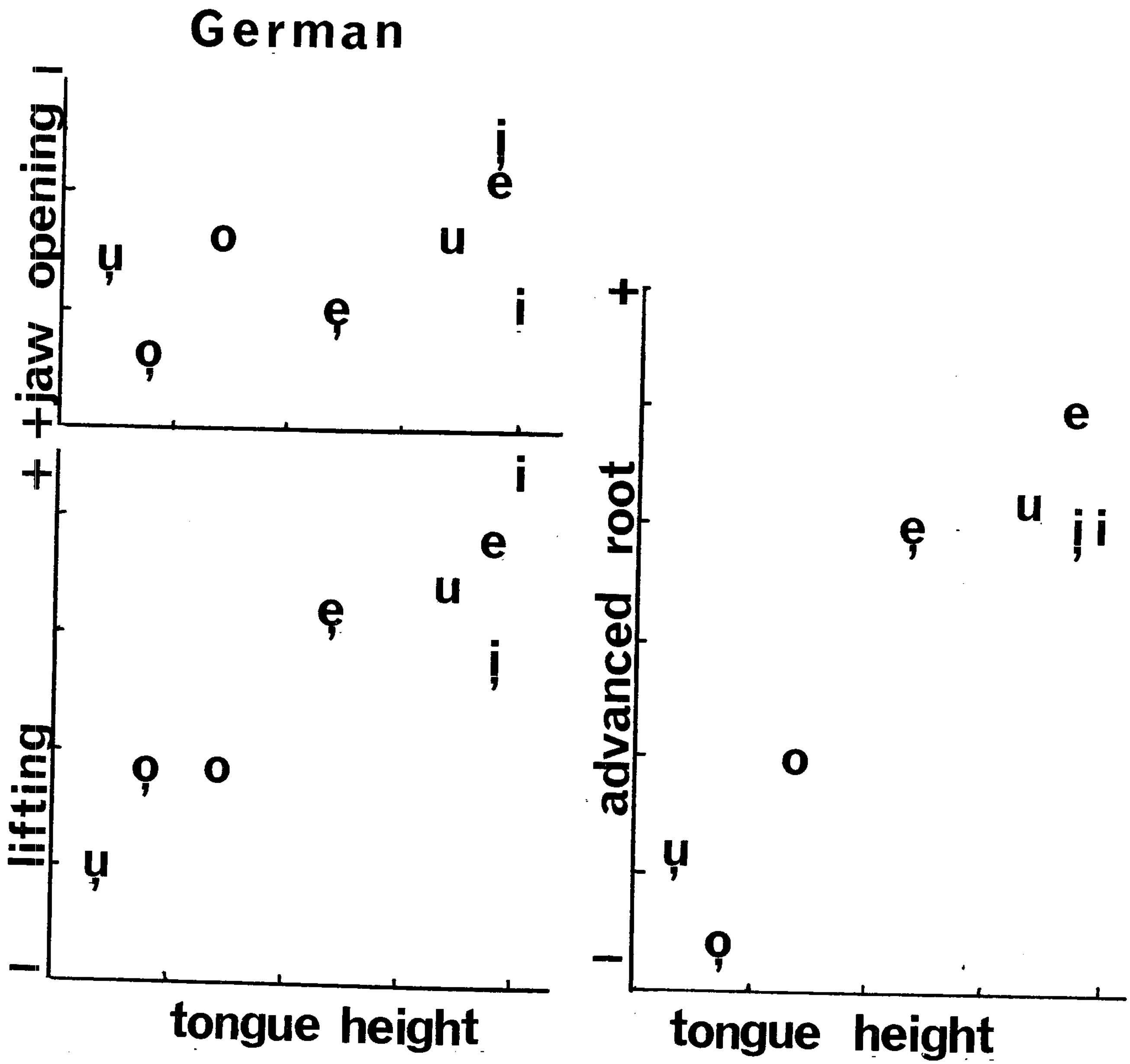
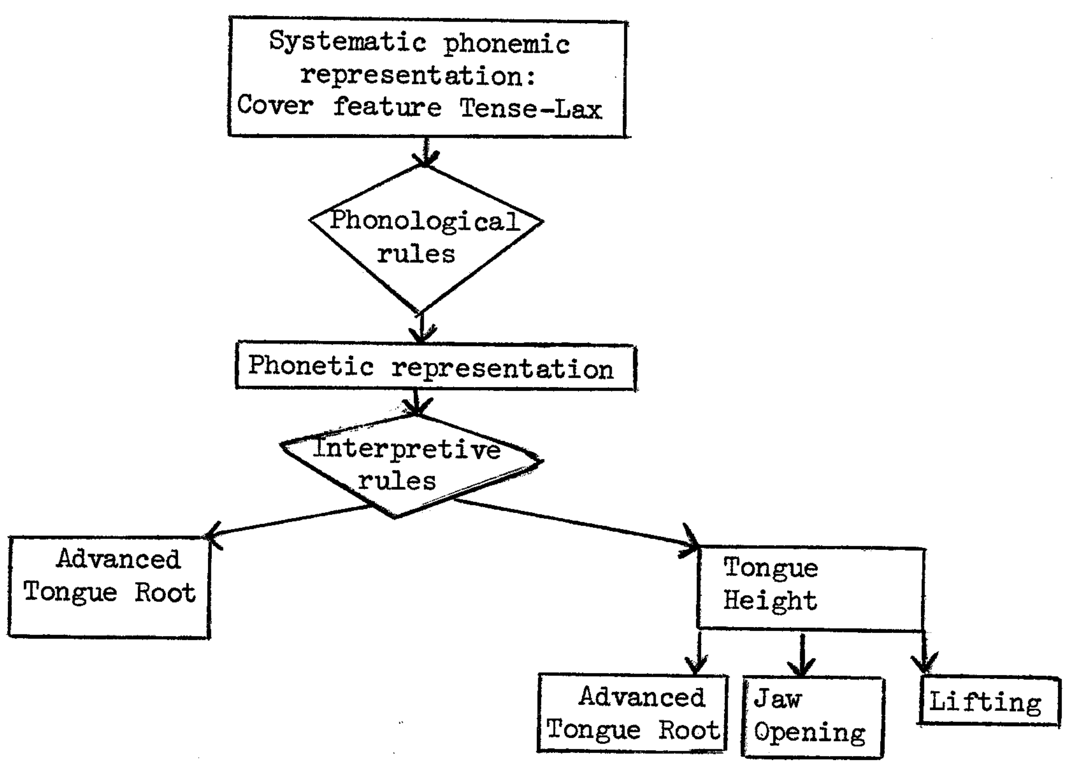


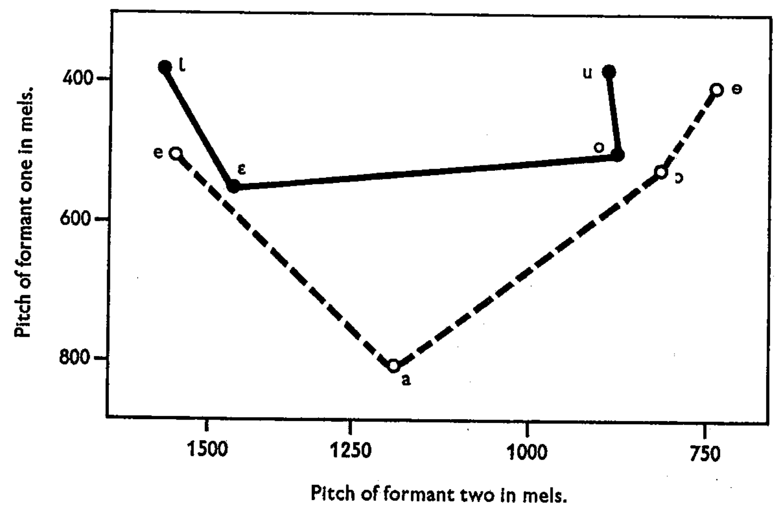
Figure 9. Tongue height related to jaw opening, lifting and advanced tongue root in German.

To recapitulate, tongue height in Twi is controlled by jaw opening interacting with lifting in Dho-Luo by combinations of lifting, jaw opening and advanced tongue root, in German by lifting and advanced tongue root. When more than one speaker was used from English it was found that all three mechanisms may contribute to tongue height.

Discussion. The African languages under consideration exhibit vowel harmonies that are structurally similar to each other. Part of the task of a systematic phonemic representation is to make possible comparisons of the operation of the "same phonological rule" across languages. Therefore it becomes desirable to represent the conditioning factor of vowel harmony with a uniform feature, so that this structural similarity can be represented in the grammar. The phonetic level includes a specification of what phonetic mechanisms are involved, but the phonetic interpretation of the vowel harmony feature will not be the same in the different languages. In Twi and Dho-Luo the articulatory mechanism is Advanced Tongue Root, in Ateso it is Tongue Height, and in Igbo it is partly Advanced Tongue Root, partly Tongue Height. This leads to positing different representations of vowel harmony at different levels in the phonology. At the lexical and systematic phonemic levels it is represented by an abstract cover feature (Cf. Ladefoged, 1972). At a later stage interpretive rules assign the cover feature to some possible phonetic mechanism(s). The diagram below illustrates this model.



a.



b.

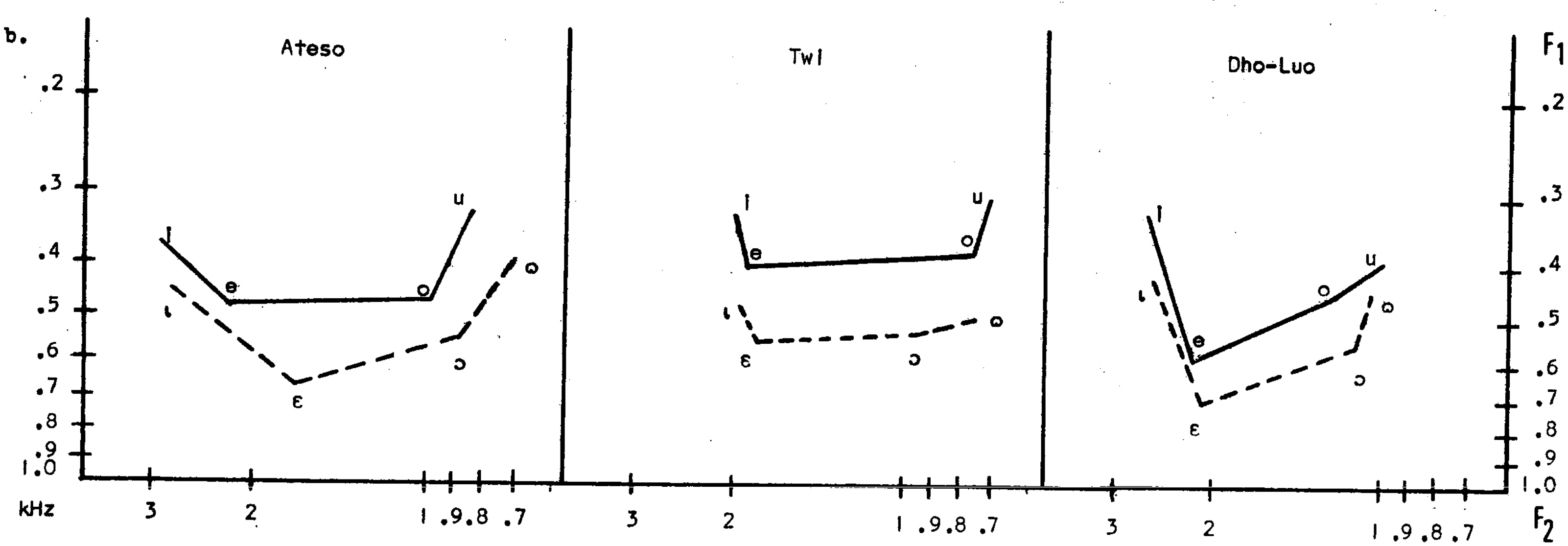


Figure 10. a. Formant chart of Igbo vowels (= Figure 13 in Ladefoged 1964).
b. Formant charts of Ateso, Twi, and Dho-Luo.

The phonological feature of vowel harmony can in the case of Twi be interpreted as a phonetic feature Advanced Tongue Root with a fairly simple relation to an articulatory gesture that pulls the root of the tongue forward and simultaneously lowers the hyoid and the larynx. In Ateso the articulatory mechanism of vowel harmony cannot be distinguished from that of Tongue Height. This feature bears a much more complex relation to articulatory gestures than the feature Advanced Tongue Root. It may involve the same articulatory mechanism as for Advanced Tongue Root, and in addition gestures of jaw opening and lifting.

The previous discussion has assumed that the "phonetic mechanism" is articulatory in nature, and that what the speaker tries to accomplish, his phonetic target, is an articulatory gesture. As a result we find rather complex relationships between a phonetic feature and vowel targets in different languages, and even between different vowel targets of one speaker. The obvious next question to ask is whether these relationships are simpler if we instead assume that vowel targets are acoustic.

Figure 10a shows the formant chart of Igbo (Fig. 13 in Ladefoged, 1964, and Fig. 10b), the charts for Ateso, Twi, and Dho-Luo. The formants are measured from the midpoint of the vowels in words. The set of 'tense' vowels are connected with a straight line, and the set of 'lax' with a broken line.

The first formants of the back vowels for Ateso differ somewhat more than the first formants of the back vowels for Igbo, but the overall pictures of the acoustic spaces are very similar. Igbo and Ateso both have the 'tense' vowels inside the space of the 'lax' vowels. The F_1 - F_2 space of Dho-Luo resembles that of Igbo and Ateso. The only tense vowel that is not inside the space of the lax ones is [u], where the second formant is too low. This could however be a result of the fact that [u] is the only vowel in Dho-Luo where an advanced tongue root participates in attaining tongue height.

The formant chart of Twi displays a different picture of the relation between the two vowel sets. The main difference is due to the first formant. The first formant of <u>all</u> the tense vowels is lower than the first formant of <u>all</u> their lax counterparts. The same result was obtained after an acoustic analysis of the vowels of three other Twi speakers. The Twi chart exaggerates, as it were, the acoustic distinction between tense and lax vowels. Twi is also the only language where the pharynx width in the vicinity of the tongue root is greater for all tense vowels than for all lax vowels. Igbo and Dho-Luo have a more advanced tongue root for tense vowels than for their lax counterparts, but the tongue root mechanism does not separate out all tense vowels from all lax vowels.

In summary, the articulatory data of Twi demonstrate that an advanced tongue root separates out all tense vowels from all lax vowels. In Twi the advanced tongue root is also accompanied by larynx lowering which further enlarges the pharyngeal cavity. This mechanism contributes nothing to tongue height. When the acoustic data are considered we also find one factor, the first formant, that separates out tense and lax vowels. Thus the phonological feature of vowel harmony in Twi could be attributed in a fairly simple manner to either an articulatory parameter of advanced tongue root and lowering of the larynx, or to an acoustic parameter of a downward shift of F_1.

The results of the articulatory data of the other African languages show that in Ateso the two sets differ by the height mechanism, in Igbo some vowels differ by height, others by the tongue root mechanism, and in Dho-Luo the advanced tongue root separates each tense vowel from its lax counterpart, but it does not separate all tense vowels and all lax vowels as in Twi. In spite of all the separate articulatory means involved the acoustic formant spaces look very similar in all three languages. The tense vowels are "inside" the space of the lax vowels. The first formant does not separate out the two classes of vowels unambiguously, and tenseness becomes confused with the dimension of acoustic vowel height.

Conclusion. When the phonological feature Vowel Harmony is interpreted in articulatory terms the four African languages result in four different interpretations. If the phonological feature is interpreted with reference to acoustic parameters the relation between the abstract cover feature and its interpretation becomes much simpler. Two patterns can be distinguished: one where all tense vowels have a lower first formant than lax vowels, and another where the tense and lax vowels seem to be on a dimension that is very similar to vowel height. Looked at from the point of view of a model of speech production this provides support for the claim that what a speaker aims at in vowel production, his target, is a particular configuration in an acoustic space where the relations between formants play a crucial role. The nature of some vowel targets is much more likely to be auditory than articulatory. The particular articulatory mechanism a speaker makes use of to attain a vowel target is of secondary importance only.

The variation among speakers of English should lead us to view all our data on other languages with caution. We are currently recording additional speakers of these languages. But at the moment all our remarks on other languages should be regarded as indicating simply what is possible. Thus we can say that the first part of the original hypothesis is supported and there may be a feature of advanced tongue root at work in some languages. Three of the four speakers of the African languages considered use it to distinguish their vowel harmony sets. But it clearly cannot be used in a feature system to replace all cases of what have previously been called tense-lax distinctions among high vowels, which is the second part of the hypothesis suggested by Perkell (1971). Some tense-lax differences between vowels, as in Ateso and German, and probably as in English, are better interpreted in terms of tongue height and length differences.

Acknowledgements: We wish to thank Peter Cockshott, Bill Greenberg, John Okeju, Achola Pala, Edward Twum-Akwaboah for their invaluable assistance and discussions.

References

Berry, J. (1955) "Some notes on the phonology of the Nzema and Ahanta dialects." *Bull. School Oriental and Afr. Studies*, 17.1, 160-5.

_____ (1957) "Vowel Harmony in Twi." *Bull. School of Oriental and African Studies*, 19, 124-130.

Chomsky, N. and M. Halle (1968), *The Sound Pattern of English*, New York: Harper and Row.

Christaller, J.G. (1875) *A Grammar of the Asante and Fante Language Called Tshi (Chwee, Twi)*. Basel.

Greenberg, W. (1970) "Harmonically neutral vowels --- phonologically non-existant, or?" UCLA: unpublished MS.

Halle, M. and K.N. Stevens (1969), "On the Feature 'Advanced Tongue Root' ", *Quart. Prog. Rep. No. 94, Res. Lab. Electron.*, M.I.T., 209-215.

Ladefoged, P. (1964), *A Phonetic Study of West African Languages*. London: Cambridge University Press.

_____ (1972) "Phonological features and their phonetic correlates". *J. Internat. Phon. Assoc. 2*.

Ladefoged, P., J. DeClerk, M. Lindau, and G. Papçun (1972) "An Auditory Theory of Vowel Production." *Working Papers in Phonetics 22*. UCLA.

Lindblom, B. and J. Sundberg (1971) "Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement." *J. of the Acoust. Soc. of Amer.* 50,4, 1166-1179.

Perkell, J. (1971) "Physiology of speech production: a preliminary study of two suggested revisions of the features specifying vowels." *Quart. Prog. Rep. No.* 102, *Res. Lab. Electron., M.I.T.*, 123-139.

Schachter, P. and V. Fromkin (1968) "A Phonology of Akan: Akuapem, Asante, and Fante." *Working Papers in Phonetics 9*. UCLA.

Stewart, J.M. (1967) "Tongue Root Position in Akan Vowel Harmony." *Phonetica 16*. 185-204.

Wängler, H.-H. (1961,2nd ed.) *Atlas Deutscher Sprachlaute*. Berlin: Akademie-Verlag.

Welmers, W.E. (1946) *A Descriptive Grammar of Fanti*. Supplement to *Language* 22.3.

The Three Glottal Features

Peter Ladefoged

[Expanded version of a paper presented at the 82nd Meeting of the Acoustical Society, October 1971.]

In this paper I will consider the linguistic aspects of speech that are controlled by the actions of the larynx. In doing so I am to a great extent restating a position which I have documented from a linguistic point of view in an earlier monograph (Ladefoged, 1967) and in its latest revision (Ladefoged, 1971). In the present paper I will contrast this position with some recent suggestions by Halle and Stevens (1971).

The glottis viewed from above is like a triangular slit. The two principal forces acting on it are shown very schematically in Figure 1. One force causes it to be stretched in the anterior-posterior dimension (in the direction of the line a-a') mainly by varying the position of the thyroid cartilage. The other force results in the movement of the arytenoid cartilages, along the lines shown as b-b', thus decreasing or increasing the length of the base of the triangle. There are, of course, numerous other forces acting on the glottis. For example, there are movements in the vertical dimension, which cannot be shown in Figure 1. To the best of my knowledge none of the other possible movements are used linguistically independently of the two forces we have discussed. In fact, even these two forces interact, in that movements of the arytenoids obviously also affect the length of the glottis.

There are three ways in which languages use laryngeal features. In the first place, all languages make syntactic differences between utterances by varying the PITCH. Some languages also use pitch differences to make lexical contrasts. The mechanisms for varying the pitch of the voice are well known and need not be considered in detail here. The most important is the stretching of the vocal cords principally in the anterior-posterior dimension, through the action of the forces along a-a'; one of the most dramatic visual changes in the appearance of the glottis is the increase in length which occurs with an increase in pitch. Along with the stretching of the glottis in the anterior-posterior dimension shown in the figure, there may be some stretching in the superior-anterior dimension (not shown) due to the raising of the larynx. Pitch changes are also the result of tensing the vocal cords by means of the intrinsic muscles, which can thus vary the *mass* which is able to vibrate. In addition the rate of vibration depends on the size of the Bernoulli effect which causes the vocal cords to be sucked together. The size of the Bernoulli suction is proportional to the rate of flow of air through the glottis. The rate of flow of air depends in part on the subglottal pressure -- a parameter which will not be considered further in this paper where we are exclusively concerned with laryngeal features. But it is also partially dependent on the position of the vocal cords themselves. When the glottis is spread open there is obviously a potentially higher rate of flow than when it is narrowed. Hence, even from the point of view of laryngeal features, we must remember that the physiological correlate of pitch variations is not just the muscular adjustments which can be summarised in terms of a label such as glottal tension; pitch variations may also occur as a result of variations in the degree of constriction of the glottis.

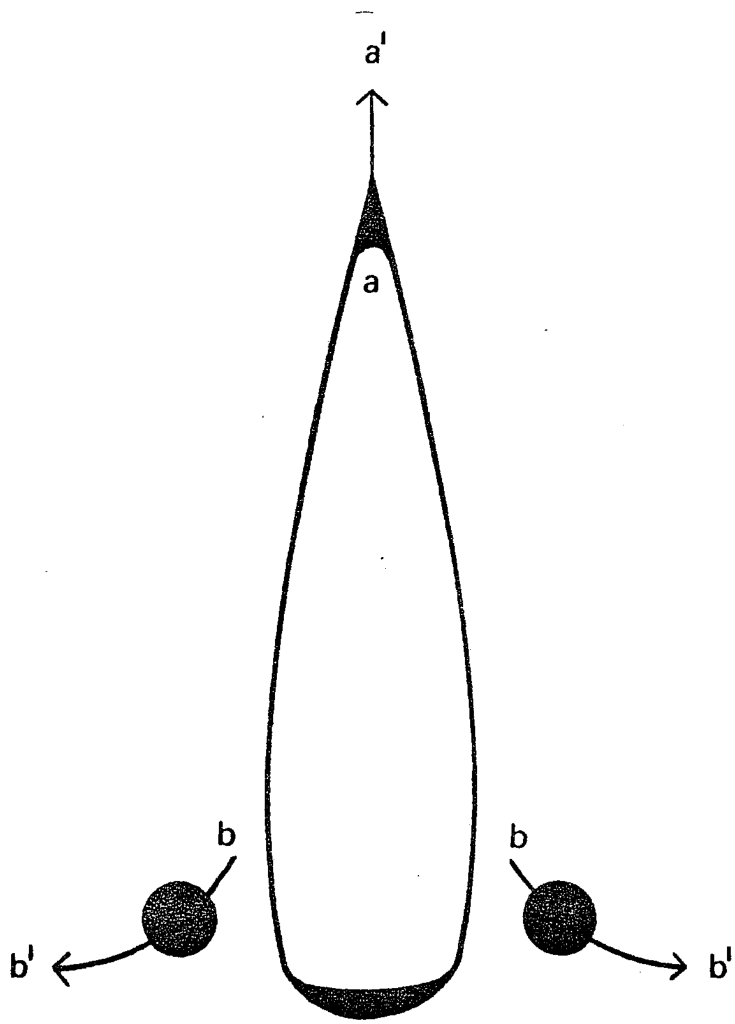


Figure 1. A schematic view of the glottis showing the two principal forces involved in linguistic contrasts. Movements along the line a a' represent a stretching of the glottis and are associated with PITCH features. The solid circles represent the arytenoid cartilages; movements along the lines b b' are due to the actions of these cartilages and represent the linguistic feature GLOTTAL STRICTURE.

The second way in which languages use laryngeal gestures is to vary the phonation type. There is more speculation on this topic than on most others in phonetic theory. What little evidence there is seems to suggest that variations in phonation type are principally dependent on movements along the lines b-b' in figure 1, just as variations in pitch are principally dependent on movements in the direction of the line a-a'.

In support of this contention, consider the photograph shown in figure 2 (on the facing page). These photographs, which were taken by John Ohala and Ralph Vanderslice, show the state of the glottis in (a) voiceless sounds, such as the voiceless vowel symbolized by /h/ at the beginning of the word *hot;* (b) murmured or breathy voiced sounds of the kind that occur in Hindi and other Indo-Aryan languages; (c) voiced sounds such as the vowel /a/ in *hot;* and (d) laryngealized or creaky voiced sounds of the kind that occur in Indo-Chadic languages. Immediately below each photograph is a schematic representation of the glottis, with the vibrating parts of the vocal cords indicated by dashed lines and the centers of the arytenoid cartilages shown by solid circles. The distance between the arytenoids is arbitrarily taken to be 100% in voiceless sounds as in (a). In sounds of this kind the arytenoids are far apart, and the vocal cords cannot be set into vibration. In breathy voiced sounds as in (b) the distance between the arytenoids is only 90% of that in (a). Because of the basically triangular shape of the glottis, the anterior portions of the vocal cords are now close enough together to be set into vibration. In ordinary voiced sounds as in (c) the arytenoids are still closer together (about 75% of the distance in voiceless sounds) so that the whole glottis is able to vibrate. Finally, in creaky voiced sounds as in (d) the arytenoids have come even closer (67% of the distance in voiceless sounds) so that the posterior portion of the glottis is held together and unable to move, leaving only a small part of the ligamental vocal cords vibrating. Note that, because the arytenoid cartilages move forward as they come together, the vocal cords tend to be less stretched in creaky voiced sounds; they are therefore likely to vibrate at a lower frequency. But the coming together of the arytenoids and the movements of the thyroid cartilage that stretch the vocal cords are independent laryngeal gestures, so that it is quite possible for creaky voiced sounds to occur on any pitch. They are, in fact, used with contrastive pitches in tone languages such as Kuman.

The four states of the glottis shown in figure 2 represent points in a continuum of possible arytenoid positions. Other photographs of a similar kind have shown quite clearly that the arytenoids are further apart than in figure 2(a) when breathing in, and perhaps also in strongly aspirated sounds. They are also closer together than in figure 2(d) when making a glottal stop, which is a laryngeal gesture used by many languages (e.g. Tagalog and other Malayo-Polynesian languages; or for most Americans as the form of /t/ which occurs in a word such as *button*). Furthermore several stages intermediate between those shown in figure 2 are possible. Thus there is a noticeable difference in the degree of breathiness that occurs in the murmured stops in the languages spoken in different parts of India; and the murmured sounds that occur in Southern Bantu languages are different from any that occur in India in that they are much more like sounds with regular voicing. Differences in the degree of laryngealization (creaky voicing) also occur. For example a very slight degree of laryngealization may be heard in the vowel after so-called fortis unaspirated stops in Korean. Kagaya (1971) has shown by a motion picture film of the larynx taken through a fiberscope that the fortis Korean stops have a much more constricted glottis than the corresponding lenis stops. But the slightly laryngealized Korean stops are very different from the strongly laryngealized stops that occur in Hausa. Furthermore, the difference between the two sets of vowels that occur in a Nilotic language such as Luo is that one set is slightly laryngealized, whereas the other set has regular voicing or a tendency towards murmur. But the degree of

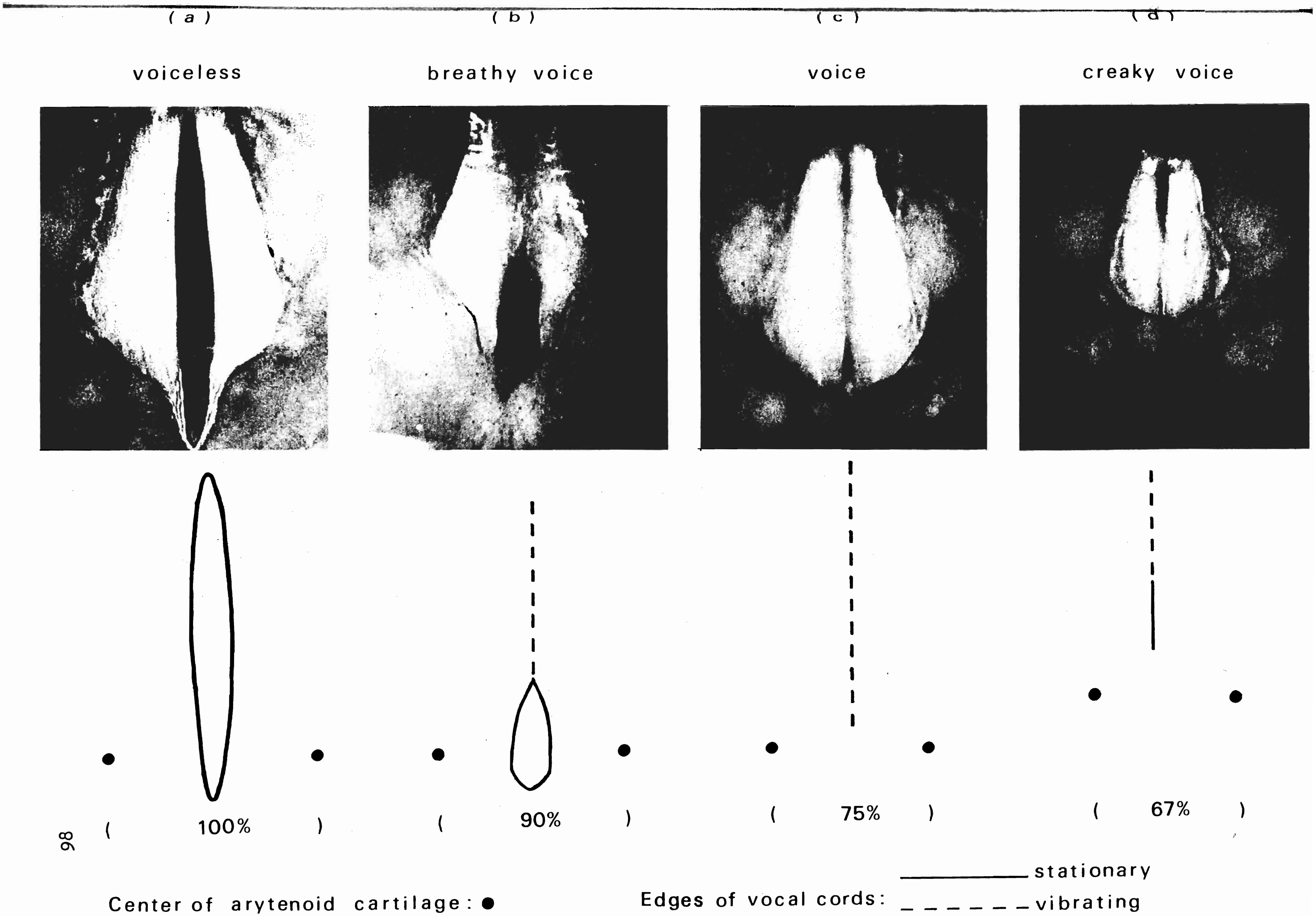


Figure 2. Photographs (taken by John Ohala & Ralph Vanderslice) and schematic representations of four states of the glottis. The distance between the centers of the arytenoid cartilages in the voiceless position is arbitrarily taken to be 100%.

laryngealization which occurs in Luo is nothing like so extensive as that found in the creaky voiced semivowels in Chadic languages such as Margi.

The multi-valued feature underlying all these differences in phonation type may be referred to as GLOTTAL STRICTURE. A set of terms for possibilities within this feature might include: (1) glottal stop; (2) creak; (3) creaky voice; (4) stiff voice; (5) voice; (6) slack voice; (7) murmur; (8) breathy voice; (9) voiceless. The terms stiff voice (for a moderate degree of laryngealization) and slack voice (for a tendency towards murmur) are taken from Halle and Stevens (1971), where they are defined slightly differently. It should be emphasized that each member of the set of terms corresponds to an arbitrary area in the range of possibilities. It would have been quite possible to have added other terms. Alternatively we could have collapsed some adjacent pairs of terms. As far as I am aware no language contrasts more than three possibilities within this feature. There is, for example, no language that differentiates between a group of sounds solely by making one with creaky voice, another with regular voice, a third with breathy voice, and a fourth which is voiceless.

The third way in which languages use laryngeal features is by varying the timing of laryngeal events, with respect to the articulatory gestures. Thus in stops and fricatives the vocal cords may be vibrating throughout the articulation (as in English *zoo* or Burmese /zan/ 'levitation'), or they may start vibrating just after the articulation (as in English *Sue*, or Burmese /san/ 'example'), or there may be a short period of aspiration before regular voicing starts (as in Burmese /s^{h}an/ 'rice'). No language uses this feature to distinguish more than three possibilities. Linguistic differences in voice onset time have been well described by Lisker and Abramson (1971, and references therein) and need not be elaborated further here.

Finally we must note that there is no reason to believe that these three glottal features are independent, or even that they allow for the specification of speech sounds in a Euclidean space. Terbeek and Harshman (1971) have discussed some of the spatial distortions that may occur in the perception of vowels, and it is possible that similar effects apply in the feature specification of the laryngeal aspects of sounds. It is clear, for example that there is an interaction between the value of the voice onset feature and the glottal stricture feature, in that only a limited range of phonation types can be potentially aspirated or unaspirated. Voiceless sounds and sounds accompanied by a glottal closure may, or may not, be followed by a period of voicelessness (aspiration); but laryngealized sounds, and sounds with regular voicing are, to the best of my knowledge, never followed by a period of voicelessness before regular voicing starts. However the voice onset and the glottal stricture features cannot be collapsed into one feature because sounds with several different phonation types (e.g. glottal stops, murmured stops, and voiceless stops) can all be aspirated.

There is also an interaction between the value on the pitch feature and that on the glottal stricture feature. We have already noted that laryngealized sounds tend to have a lower pitch, because they are the result of a glottal state in which the arytenoid cartilages are not only held together but also moved forward so that the vocal cords become less stretched. Conversely, after a voiceless sound in which there is a high rate of flow, the pitch tends to be higher because of the increased Bernoulli effect. We must emphasize however, that there is only a tendency towards a correlation between the rate of vibration of the vocal cords and the degree of glottal stricture. The glottal stricture and the pitch of a sound can be varied independently in many Nilotic languages, so at least in these languages (and therefore within the theory of general phonetics) these two features cannot be collapsed into one.

The foregoing discussion seems to indicate quite conclusively that we need at least three laryngeal features in linguistic description; and I would not be at all surprised to find that we need even more. There are some states of the glottis (such as whisper) which cannot be characterized in terms of a single variable such as distance between the arytenoids; and many authors (cf. Fromkin 1971) have shown that a single feature, pitch, is not sufficient to account for the contrasts that occur in tone languages.

Halle and Stevens (1971) have suggested that all the phenomena we have been discussing can be described in terms of only two variables: the degree of stiffness of the vocal cords; and the degree of constriction of the glottis. (They in fact characterize each of these variables in terms of two binary oppositions: [± stiff], [± slack]; and [± spread], [± constricted], with a convention that neither [+stiff, + slack] nor [+ spread, + constricted] can co-occur. They are thus expressing two ternary oppositions in terms of four binary oppositions. But this idiosyncracy is irrelevant to the present discussion.) Halle and Stevens use their features to capture a number of interesting linguistic generalizations. For example, they claim that they are able to formalize the reason for the occurrence of alternations in which a vowel occurs with a high tone after voiceless stops but with a lower tone after voiced stops. In their view, this is just an assimilation in which the vowel is taking on the stiffness value of the consonant. High tone vowels have stiff vocal cords, and so (they claim) do voiceless consonants; but low tone vowels have slack vocal cords, and (in their view) so do voiced consonants.

Halle and Stevens do not produce any evidence for their claim that the pitch tends to rise after voiceless consonants because voiceless consonants have stiff vocal cords. Personally I would rather assume there is no special stiffening in voiceless consonants; it seems to be an unmotivated and, indeed, unlikely additional complication. There are two other reasons why there may be an increase in pitch after voiceless consonants. Firstly: voiceless consonants may have a higher larynx position than the corresponding voiced consonant. As we noted earlier, this may result in a stretching of the vocal cords; but it seems curious to call this stretching a stiffening of the vocal cords. Secondly the increase in pitch may be due to the increase in the flow rate which is associated with voiceless sounds. Probably the best way of dealing with this phenomenon in phonological theory is by means of marking conventions, rather than by combining the feature specifying the degree of glottal stricture and pitch.

Furthermore, as many of the examples discussed above have shown, a correct feature system must make it possible to seperate out differences in phonation type, which are due to differences in glottal stricture, from differences in pitch which depend on another mechanism. It must be possible to show that voiced sounds, breathy voiced sounds, and creaky voiced sounds can all occur over a range of different pitches. In addition, I cannot see any reason for dispensing with any simple way of specifying voiced and voiceless sounds (which, under the Halle-Stevens proposal, are specified in terms of differences in stiffness-slackness, and are inextricably confused with pitch variations). As has been shown by Fromkin (1971), the use of a feature system of this kind results in some ridiculously unnatural rules. It seems clear that there must be at least three multi-valued glottal features, not just two ternary oppositions.

References

Fromkin, Victoria. 1971. Tonal features and rules and generative phonology. Working Papers in Phonetics 21:

Halle, Morris and Kenneth N. Stevens. 1971. A note on laryngeal features. Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T. 101:198-213

Kagaya, R. 1971. Laryngeal gestures in Korean stop consonants. Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 5: 15-24

Ladefoged, Peter. 1967. Linguistic phonetics. Working Papers in Phonetics 6.

Ladefoged, Peter, 1971. Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Lisker, Leigh and Arthur Abramson. 1971. Distinctive features and laryngeal control. Language 47 (forthcoming).

Terbeek, Dale and Richard Harshman. 1971. Is vowel perception non-Euclidean? Working Papers in Phonetics 22

Respiration, Laryngeal Activity and Linguistics

Peter Ladefoged

[Abstract of a paper to be presented at the International Symposium on Ventilatory and Phonatory Control Systems, 24-29 January 1972, London, England. The full text of this paper is not given here, as it is extremely similar to the previous paper, supplemented by additional material from Preliminaries to Linguistic Phonetics]

This paper will discuss the ways in which the languages of the world use the respiratory and phonatory mechanisms to produce linguistically significant differences in sounds. Each of these mechanisms will be discussed first in relation to the production of segments (phonemes) then in relation to suprasegmental systems (accent and intonation).

Few languages (if any) use the respiratory system as the sole means of distinguishing between segments. Chomsky and Halle have proposed that "heightened subglottal pressure" may be a linguistically significant feature. The little evidence that there is for this feature will be reviewed. There are numerous occasions when there are increases in respiratory activity (e.g. in /h/ sounds), but it is by no means clear whether these are programmed for individual segments or whether they are the result of feedback mechanisms.

All known languages make extensive use of different phonatory mechanisms. There seem to be two parameters involved, one being the glottal stricture (roughly, the degree to which the arytenoid cartilages are pulled together or apart) and the other the timing of the onset of phonation. Languages distinguish up to three possibilities on each parameter. Thus, on the glottal stricture parameter, Hausa has /ɓ/ (laryngealized or "creaky" voice), /b/ (regular phonation), and /p/ (voiceless); and Gujerati has /b/ (regular phonation), /bɦ/ (murmur or "breathy" voice), and /p/. Three way variations in voice onset time occur in Thai, Hindi, and many other languages.

The phonatory and respiratory mechanisms function together to produce suprasegmental phenomena. Accented syllables have an increase of respiratory energy, accompanied by a variation in the tension of the vocal cords. Variations in intonation contours can be characterised mainly in terms of glottal tensions. In English most intonation contours can be characterised in terms of two binary features, Cadence (a falling pitch) and Endglide (a terminal rise in pitch).

On the Reality of Linguistic Constructs: Evidence from Speech Errors

Victoria A. Fromkin

[Revised version of a paper given at the VIIth International Congress of Phonetic Sciences, Montreal, Canada]

During the period when anti-mentalism dominated American linguistics the 'reality' of abstract constructs was denied by those linguists who agreed with Twaddell (1935) that 'the linguistic processes of the "mind" are..simply unobservable (and)...we have no right to guess about the linguistic workings of an inaccessible "mind".' But just as there is evidence for such 'unobservable' physical entities as atoms and electrons so too is there evidence for linguistic processes of the mind. Speech error data provide such evidence.

In fact, linguistic behavior cannot be explained without assuming the reality of highly abstract units and processes. Arguments have been put forth previously (Fromkin, 1971) for the reality of independent distinctive features and discrete phonemic segments, despite the fact that the emitted speech signal is a continuous one. The substitution and transposition, deletion or addition of segments and individual features which occur in deviant utterances (deviant in the sense that the spoken utterance differs from the intended target) justify their 'reality'. The examples under (1) are errors in which only feature values are switched.

(1) *a. metaphor → menaphor (t → n ; [-nasal] → [+nasal]/[+nasal] x __)

b. is Pat a girl → is bat a curl (p → b,g → k; [-voice] ← → [+voice])

---

* In examples, the arrow is to be interpreted as 'spoken as'. The pronunciation of the utterance will be given in phonetic symbols, within square brackets, only when the orthography may create an ambiguous interpretation or obscure the actual speech errors. A dash represents a pause by the speaker; a series of dots (...) indicates that no errors occurred in the intended words.

It should be noted in these examples that if we view segments as being indissoluble units one can not explain in any coherent fashion the 'cause' of the errors. However, in (1a) the change of the [t] to [n] shows that (at least in English) nasality implies voicing, and the switch from [g] to [k] in this phonetic context causes a delay in the onset of voicing, or aspiration of the voiceless stop. This suggests that the 'feature-switch' must occur prior to the phonetic constraints of the language which in turn must occur prior to the stage where articulatory commands are determined.

But to suggest that features are 'real' does not deny the reality of segmental units as can be seen by the examples under (2).

(2) a. dinner at eight → inner at date (d ← → ∅)

b. pinch hit → pitch hint ([pĩntʃ hɩ∅t] → [pɩ∅tʃ hĩnt]
n ← → ∅)

c. a whole box of flowers → a whole blocks of flowers b → bl

These errors show the addition of a segment where the intended utterance had no such unit, or the deletion of a segment present in the target, implying the existence of segment-type units.

The fact that consonant clusters can be separated is further evidence for the reality of discrete segmental units, and, in addition, reveals that at the point in the production process where the error occurs the clusters consist of *sequences* of such discrete segments. This is illustrated in (3)

(3) a. slumber party → lumber sparty (sl ... p → l ... sp)

b. there's a pest in every class → there's a pet in every class (st → t)

c. but what is deep structure → but what is steep ducture (d ... str → st ... dr)

The fact that affricates do *not* behave as other clusters, that is, the stop and fricative are never split in errors, reveals that such affricates are underlying single segments in English.

The positing of semi-independent features and discrete segments does not deny the possibility that at some later stage in the production of an utterance, clusters or syllables or even larger units may constitute one articulatory program. In fact, it is also necessary, if we wish to account for the data, to posit syllable units composed of ordered segments, as shown previously by Fromkin (1971). Any attempt to explain the complexities of language and speech by positing only a single 'basic unit' whether this be the 'feature' the 'phoneme' or the 'syllable' is bound to fail.

What appears to be even more interesting than the 'behavioral' evidence for such abstract units (abstract, in the sense that one cannot find them as discrete units in the physical signal) is the evidence provided by speech error data supporting the reality of even more highly abstract constructs.

The examples under (4) are fairly straightforward. They all illustrate the split of nasal-stop clusters.

(4) a. The bank [bæ̃ŋk] will pay [pej] 5.6% interest →
The ban [bæ̃n] will [pejk] 5.6% interest.
([ŋk] → [n], ∅ → [k])

b. The rank [ræ̃ŋk] order [ɔrdər] of the subjects →
The rand [ræ̃nd] orker [ɔrkər] of the subjects
([ŋk]...[rd] → [nd]...[rk])

c. The red tide will stink [stĩŋk] up the sea [sɩj] →
...will [stĩn] up the [sijk]

Example (4a) is interesting in that it shows that when the final k, is deleted (or moved), the remaining nasal is an alveolar [n] rather than the'expected' velar [ŋ]. This suggests that underlying the [ŋk] in *bank* is /nk/. It also attests to the reality of the homorganic nasal rule in English phonology, which must operate in the production process *after* the misordering of units has occurred.

The second example supports this hypothesis in a more ambiguous way, since if all examples were like (4b) one could conclude that the homorganic rule was operative but that *rank* was stored as /ræŋk/; the switch of the /k/ and /d/ would change the velar nasal to an alveolar. But (4c) further attests to the underlying /n/ in phonetic [ŋk] sequences.

The examples under (5) also involve velar nasals, but ones in which the intended utterances do not include phonetic clusters.

(5) a. swing and sway [swĩŋ] and [swej] → [swĩn] and [swejg]

b. sing for the man [sĩŋ... mæ̃n] → [sɩg ... mæ̃ŋ]

c. Chuck Young [tʃʌk jʌ̃ŋ] → chunk yug [tʃʌ̃ŋk jʌg]

d. cut the string [kʌt] the [strĩŋ] → [kʌ̃nt] ...[strɩg]

e. I'll wring your neck → I'll rink your neg.
([rĩŋ] [nɛk] → [rĩŋk] [nɛg])

f. morphemes in the verb string → morphemes in the verg strin
[vʌrb strɩŋ] → [vʌrg strɩn]

g. springtime for Hitler → sprig time for hintler
([sprĩŋ] [hɩtlər] → [sprɩg] [hɩntlər])

Errors that produce a phonetic [g] which does not occur in the intended *phonetic* utterances support the proposal made by Sapir (1925), Chomsky and Halle (1968) and others, for deriving the phonetic [ŋ] from an underlying sequence of /ng/ in English.

Some of the examples above may be alternatively explained by suggesting that these errors involve feature anticipations and switches. (5b), for example, may result from a denasalization of the [ŋ] in *sing* following the non-nasal [s], and the perseverance of the velarity feature i.e. [ŋ] → [g] / s x__; [n] → [ŋ] / [g] x __ . Similarly, (5e) may result merely from the addition of an anticipated [k] i.e. [rɩŋ] → [rɩŋk]/ __ x [nɛk].

The examples (5a) and (5g), however, are not so 'easily' explained. To suggest that what occurs in the process of producing '[swɩn] and [swejg]' is that the velar nasal becomes alveolar in *swing* and in addition a non-nasal velar stop is added to *sway* may describe the result but does not explain it. (5g) is equally complex without positing an underlying /ng/. If on the other hand we assume that prior to the articulatory stage the phonological representation of *swing* is /swing/, and the phonological representation of *spring* is /spring/, then the errors are similar to others in which a cluster is broken up, or a single segment is misordered. This is illustrated in (6).

(6) a. /swɩng/ and /swej/ → /swɩnØ/ and /swejg/

b. /sprɩng/time for /hɩtlər/ → /sprɩØg/ .... /hɩntlər/

All the errors cited under (5) are more plausibly explained by assuming the existence of the /ng/. Such errors also provide 'behavioral' support for the English phonological rule (7).

(7) $g \rightarrow \emptyset$ / N --#

Since the [g] emerges if the nasal is deleted or transposed during speech production, the /g/ must be present at some stage in the production process. Should this hypothesis be accepted it shows that the homorganic nasal constraint must occur *after* the error in serial ordering since the nasal in *swing* is alveolar rather than velar. This evidence for ordering is also given in the examples under (4).

These errors reveal, in addition, the extent to which abstract generalizations are constructed in acquiring a language. While *swing* is never heard as *[swɪng] (nor even as *swinger* *[swɪŋgər]) by 'analogy' with *long* [lɔŋg] and *longer* [lɔŋgər] and by recognizing the distributional constraints of [ŋ], general rules are devised which account for all the data. Some might suggest that these rules (and the underlying forms) are only posited after the speaker has learned to spell. Even if this is the case, the result of the spelling system, together with the phonological facts of the language, can combine to permit the storage of abstract segments never heard phonetically.

It is not surprising to find evidence for such underlying segments. Even more abstract structures are shown to be 'real' as illustrated in (8)

(8). a. I regard this as imprecise → I disregard this as precise*

b. It's not possible that he's going → it's possible that he's not going -- it's not possible that he's not going -- I mean, it's impossible that he's going.

c. People agree that it is not well formed → people don't agree that it's well formed.

d. The vowels are not....are unrestricted.

e. I know he's not going to Valley College →
I don't know he's going to Valley...I mean...
I know he's NOT going to Valley College.

f. He had no idea → he had an idea...uh...no idea.

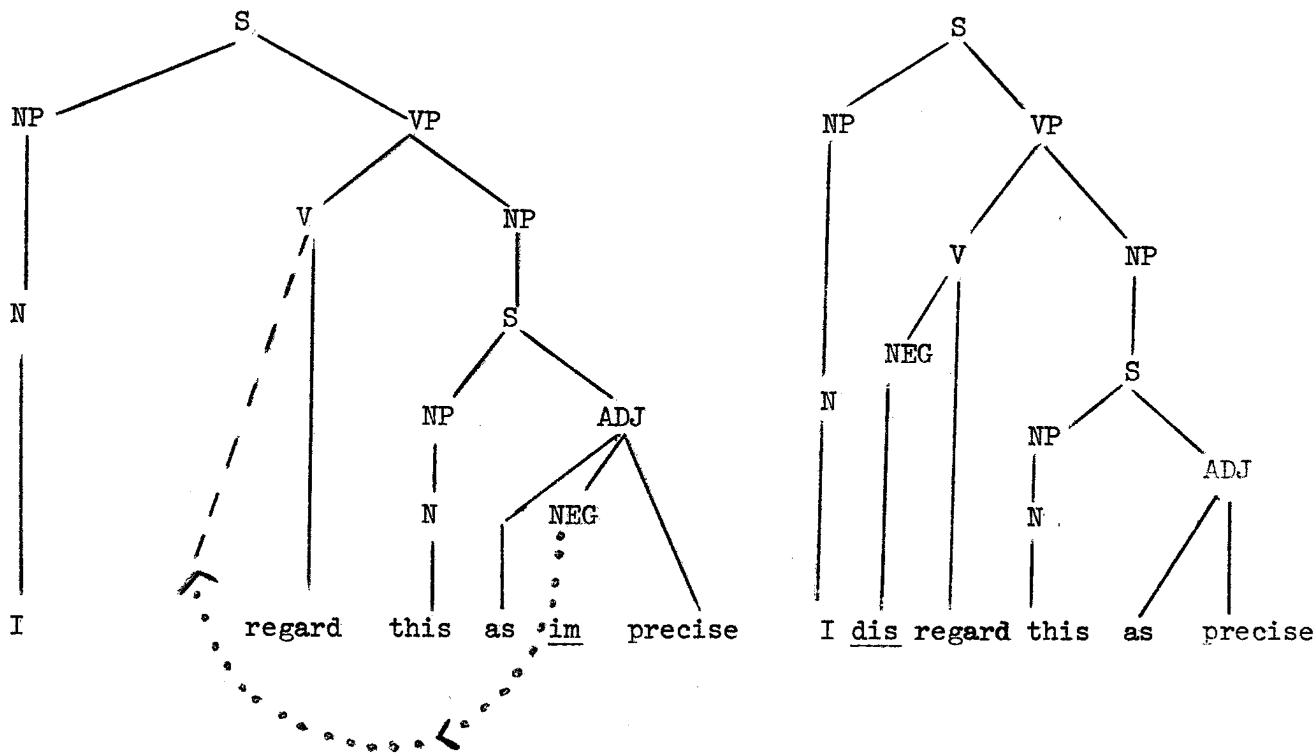
Examples under (8) demonstrate that in producing a negative sentence a speaker must first generate an abstract NEGATION element which is independent of any particular word or place in the string. In non-deviant sentences this element is placed in a specified place in the string. But just as segments or features may be disordered, so may such syntactic elements.

The intended surface structure and the surface structure of the actual utterance of (8a) are given under (9).

---

* This error was provided by Roland Hausser. I would like to thank him, in addition, for his stimulating insights regarding syntactic errors of this type.

(9) Surface Structure of Intended Utterance   Surface Structure of Spoken Utterance



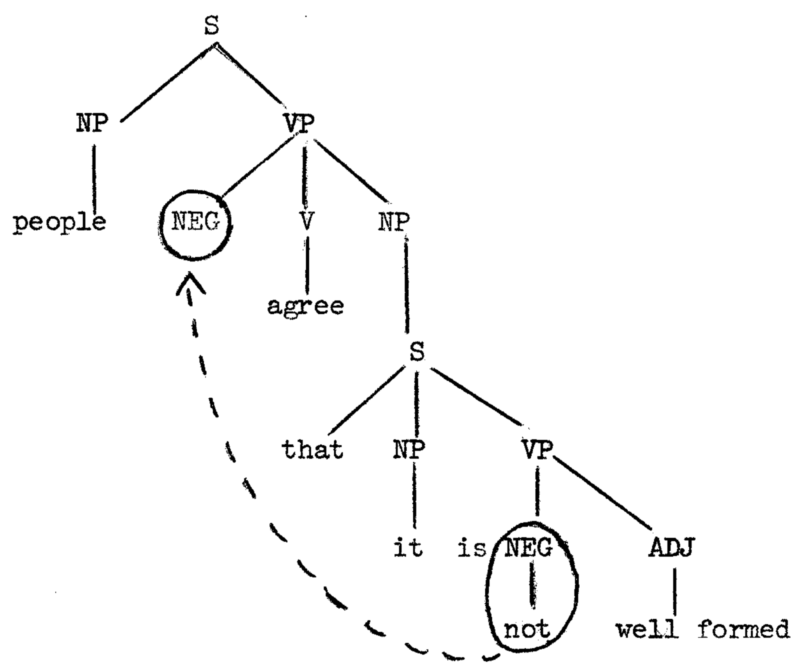
In the intended utterance, the negation occurs in the embedded sentence, while in the deviant spoken utterance the negation occurs in the higher sentence. One cannot explain this error by a performance model which posits a Markovian process of word selection. Such errors provide further evidence to Lashley's hypothesis concerning a hierarchy of structures in the serial order of behavior. (Lashley, 1961).

These examples also show that the morphophonemic rules which produce the phonological and final phonetic realizations of negation must occur after the Negative element has been moved, that is, NEGATIVE + *precise* → *imprecise* while NEG + *regard* → *disregard*.

As is shown by (8b) NEG + *possible* may alternatively become *not possible* or *impossible* under certain conditions, just as NEG + *restricted* may occur as *unrestricted*.

Examples (8c) and (8e) once more illustrate the existence of underlying abstract elements which may not occur on the surface. In c, the intended utterance has no *do* form, i.e. the present tense of the verb *agree* in the affirmative has no phonetic expression. The *do* insertion must occur after the NEG is disordered. (10) shows the intended surface structure and the disordered surface structure (before the *do* is inserted).

(10)



What was actually said was not 'People not agree that it is well formed' but rather 'People *don't* agree that it is well formed.'

Since *do* does not occur in the intended utterance, it must be inserted after the NEG element is disordered, in keeping with the syntactic rules of the language.

I am not claiming that syntactic rules can not be broken. Speech errors also reveal those rules which either fail to apply in the production of an utterance, or where misordering results in ungrammatical sentences. Such errors are given under (11)

(11) a. The last I <u>knew</u> about that → The last I <u>know</u> about that.

b. It seems like all the dogs in the neighborhood bury <u>their</u> bones in our back yard → ...<u>its</u> bones....

c. Now you can relax → Now you can relaxed

d. That's what we're going to be doing today → ...going to be do today.

e. Laurie's boyfriend has longer hair than he does...than she does.

f. I did it yesterday → I'll do it yesterday.

While such syntactic errors occur (and these are but a small sample of the kinds of syntactic errors) the examples under (9) show that there is constant monitoring which occurs during the production process. Such monitoring is not foolproof or no errors would occur. But the 'overlooking' of one error does not preclude the 'correction' of other aspects of syntax or phonology. Thus, when NEG is moved incorrectly, the *do* insertion rule may 'correct' another possible syntactic error.

Evidence for many other such performance rules (which are identical with grammatical rules), for semantic features and classes, for the generation of syntactic structures prior to lexical insertion, for major category nodes, etc. all of which must be posited in a model of linguistic performance, as well as in a grammar, is similarly provided by speech error data. In this paper just a few illustrations were presented supporting Sapir's statement that 'the notion of speech sounds is (not) explicable in simple sensory motor terms (since)...a complex...psychology...is implicit in the utterances of the simplest consonant or vowel.' (1925) Speech errors provide a way to investigate this complex psychology.

References

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English.* N. Y. Harper & Row.

Fromkin, V.A. (1971) 'The Non-anomalous Nature of Anomalous Utterances' *Language* 47.1, 27-52.

Lashley, K.S. (1961) 'The problem of serial order in behavior' in L.A. Jeffress (ed) *Cerebral Mechanisms in Behavior,* Wiley.

Sapir, E. (1925) 'Sound patterns in language' *Language* 1. 37-51.

Twaddell, W.F. (1935) 'On Defining the Phoneme' *Lang. Monograph No.16.*

Determination and proof of minimum uniqueness conditions for PARAFAC1

Richard A. Harshman

Introduction

The most important single fact about PARAFAC1 is that it provides a unique solution when the data fulfills certain necessary conditions. It is obviously important, then, to determine just what these conditions are. This problem has been approached in two different ways: empirically and mathematically.

Empirically, the necessary and sufficient conditions for uniqueness are explored by analysing "synthetic" data. This synthetic data is generated by computer so that the underlying structure and mathematical properties of each data set are known exactly. The computer starts with factors and their relationships and then generates the data that those factors would produce. In testing each data set, the PARAFAC1 program is used to repeatedly analyse the data from different starting positions. When, for a given data set the PARAFAC1 program converges on the same solution from all different starting positions, it is concluded that the PARAFAC1 model provides a unique description of that set of data. From a number of such experiments on different types of data sets (i.e. different sized data sets, different types of factors underlying the data sets, etc.) it is possible to observe the common properties of those data sets which have unique solutions, and thus generalize to those conditions which any data set must fulfill in order to have a unique solution. Further, it is possible to observe any regular patterns in the types of data that yield non-unique solutions and generalize to typical ways in which data sets can be "inadequate" to uniquely determine a solution. When non-unique solutions occur, one can observe the ways in which uniqueness can partially or completely "break down", and relate each type of non-uniqueness to a corresponding type of data "inadequacy". The results of such empirical studies of PARAFAC1 are described in Working Papers 16 (Harshman, 1970).

Mathematically, the conditions for uniqueness are explored by attempting to derive them from the algebraic properties of the PARAFAC1 model. If all our questions could be answered algebraically, there would be no need for the empirical studies (except to study the quirks of a particular algorithm, or the effects of different amounts and kinds of "noise" added to the data.) Unfortunately, the algebraic properties of PARAFAC1 are relatively subtle, and most of all the mathematical insights are yet to be discovered. The proof which is given below should be viewed as one more step in an ongoing effort to gain mathematical understanding of the subtler properties of PARAFAC1.

The first uniqueness proof for PARAFAC1 was discovered by Robert Jennerich of the UCLA Department of Biomathematics. It was published, along with the results of a number of empirical studies on PARAFAC1, in Harshman 1970. Jennrich's Uniqueness Theorem showed that the potential for a unique solution exists when one has as many parallel two-way data sets as there are factors underlying the data. For example, if there are three factors underlying a set of measurements, then one would need to have at least three measurements, on three objects, repeated on three different occasions, to fit the uniqueness conditions required by Jennrich's Theorem. (Certain other conditions were also required which are described in Harshman, 1970).

Empirical results, however, suggested that the conditions required by Jennrich's theorem were stronger than were really necessary in order to determine uniqueness. It was empirically possible to uniquely extract at least 6 factors from data sets having only three repetitions of the two-way tables (three different "occasions"). The *minimal* conditions of uniqueness could not be determined, however, since convergence became too slow if the number of occasions was too small compared to the number of factors. (See Harshman 1970, chapter 4).

The proof which follows shows that <u>two</u> occasions can be sufficient to uniquely determine any number of factors, provided that the factors change size from the first to the second occasions and that the percent change of each factor is different than that of the other factors. The proof reveals some of the necessary and sufficient conditions for uniqueness. These findings shed light, for example on patterns of "partial breakdown" of uniqueness by clarifying how and why uniqueness fails, as well as when it fails.

<u>Theorem</u>

If

(1) $X_1 = A\, D_1\, B'$

(2) $X_2 = A\, D_2\, B'$

where A and B are n by l matrices which are non-horizontal ($n \geq l$) and "basic" (of rank l) and D_1 and D_2 are non-singular diagonal matrices such that

(2b) $(D_1)(D_2^{-1}) = D_p$ where D_p has distinct diagonal elements.

And if

there exists some alternate representation of X_1 and X_2, such as

(3) $X_1 = \overset{*}{A}\, \overset{*}{D}_1\, \overset{*}{B}'$

(4) $X_2 = \overset{*}{A}\, \overset{*}{D}_2\, \overset{*}{B}'$

then

(5) $\overset{*}{A} = A\, \Delta_x$

(6) $\overset{*}{B} = B\, \Delta_y$

where Δ_x is some combination of a diagonal and a permutation matrix, and the same is true of Δ_y i.e. Δ_x, Δ_y are nonsingular with one nonzero element in each column.

<u>Proof</u>

We start with our expressions for X_1, X_2

(1) $X_1 = A\ D_1\ B'$

(2) $X_2 = A\ D_2\ B'$

now let us seek the restrictions imposed on any alternate A matrix. We can describe any alternate A matrix as our original A matrix subjected to some transformation T_A.

(7) $\overset{*}{A} = AT_A$

This is possible since the columns of A and $\overset{*}{A}$ must span the same space (e.g. X_1). In fact we can write an expression for T using equations 1, 2, 3, and 4. Since

(8) $\overset{*}{A}\ \overset{*}{D}_1\ \overset{*}{B}' = A\ D_1\ B'$

it then follows that by post-multiplying both sides by $(\overset{*}{D},\ \overset{*}{B}')^{-1}$ the right pseudo-inverse of $\overset{*}{D}$, $\overset{*}{B}'$, i.e. $(\overset{*}{B}(\overset{*}{B}'\overset{*}{B})^{-1})\overset{*}{D}^{-1}$

(9) $\overset{*}{A}\ \overset{*}{D}_1\ \overset{*}{B}'(\overset{*}{D}_1\overset{*}{B}')^{-1} = A\ D_1\ B'\ (\overset{*}{D}_1\ \overset{*}{B}')^{-1}$

(10) $\overset{*}{A} = A(D_1\ B'\ \overset{*}{B}'^{-1}\ \overset{*}{D}_1^{-1})$

let $T = D_1\ B'\ \overset{*}{B}'^{-1}\ \overset{*}{D}_1^{-1}$ and we obtain, by substitution, equation (7).

In a similar fashion, we obtain

(11) $\overset{*}{B} = B\ T_B$

now since $T_A\ T_A^{-1} = I$, $T_B\ T_B^{-1} = I$, we write

(12) $X_1 = A(T_A\ T_A^{-1})\ D_1\ (T_B'^{-1}\ T_B')\ B'$

(13) $X_2 = A(T_A\ T_A^{-1})\ D_2\ (T_B'^{-1}\ T_B')\ B'$

or, simply regrouping parentheses,

(14) $X_1 = (A\ T_A)(T_A^{-1}\ D_1\ T_B'^{-1})(T_B'\ B')$

(15) $X_2 = (A\ T_A)(T_A^{-1}\ D_2\ T_B'^{-1})(T_B'\ B')$

which can be rewritten, substituting $\overset{*}{A} = A\ T_A$, $\overset{*}{B}' = T_B'\ B'$ as

(16) $X_1 = \overset{*}{A}\ (T_A^{-1}\ D_1\ T_B'^{-1})\ \overset{*}{B}'$

(17) $X_2 = \overset{*}{A}\ (T_A^{-1}\ D_2\ T_B'^{-1})\ \overset{*}{B}'$

now since equations (3) and (16) both give expressions for X_1

(18) $\overset{*}{A}\ \overset{*}{D}_1\ \overset{*}{B}' = \overset{*}{A}\ (T_A^{-1}\ D_1\ T_B'^{-1})\ \overset{*}{B}'$

using the fact that $\overset{*}{A}$ and $\overset{*}{B}$ are basic, and thus have left pseudo-inverses we can pre and post multiply by appropriate inverses obtaining

$$(19) \quad (\overset{*}{A}^{-1}\,\overset{*}{A})\,\overset{*}{D}_1\,(\overset{*}{B}'\,\overset{*}{B}'^{-1}) = (\overset{*}{A}^{-1}\,\overset{*}{A})(T_A^{-1}\,D_1\,T_B'^{-1})(\overset{*}{B}'\,\overset{*}{B}'^{-1})$$

or

$$(20) \quad \overset{*}{D}_1 = (T_A^{-1}\,D_1\,T_B'^{-1})$$

In precisely the same fashion, by operating on equations for X_2 (i.e. equations (4) and (17)) we can obtain

$$(21) \quad \overset{*}{D}_2 = (T_A^{-1}\,D_2\,T_B'^{-1}).$$

Now equations (20) and (21) can be premultiplied on both sides as follows

$$(22) \quad \overset{*}{D}_1^{-1}\,\overset{*}{D}_1 = \overset{*}{D}_1^{-1}\,(T_A^{-1}\,D_1\,T_B'^{-1})$$

$$(23) \quad \overset{*}{D}_2^{-1}\,\overset{*}{D}_2 = \overset{*}{D}_2^{-1}\,(T_A^{-1}\,D_2\,T_B'^{-1})$$

or

$$(24) \quad I = (\overset{*}{D}_1^{-1}\,T_A^{-1}\,D_1\,T_B'^{-1})$$

$$(25) \quad I = (\overset{*}{D}_2^{-1}\,T_A\,D_2\,T_B'^{-1})$$

setting these expressions equal to one another, we get

$$(26) \quad \overset{*}{D}_1^{-1}\,T_A^{-1}\,D_1\,T_B'^{-1} = \overset{*}{D}_2^{-1}\,T_A\,D_2\,T_B'^{-1}$$

postmultiplying both sides by T_B' we get

$$(27) \quad \overset{*}{D}_1^{-1}\,T_A^{-1}\,D_1 = \overset{*}{D}_2^{-1}\,T_A\,D_2$$

Premultiplying both sides by $\overset{*}{D}_1$ and postmultiplying by D_1^{-1} we obtain

$$(28) \quad T_A = \overset{*}{D}_1\,\overset{*}{D}_2^{-1}\,T_A\,D_2\,D_1^{-1}$$

now

$$D_2\,D_1^{-1} = D_p \text{ from (2b)}$$

we can let

$$(29) \quad D_q = \overset{*}{D}_1\,\overset{*}{D}_2^{-1}$$

we know that D_q is nonsingular since $\overset{*}{D}_1$, $\overset{*}{D}_2$ are nonsingular (if they were singular, the rank of $\overset{*}{A}\,\overset{*}{D}_1\,\overset{*}{B}'$ would not be the rank of X_1 and similarly for X_2)

So from (29) and (2b), we can rewrite (28) as

$$(30) \quad T_A = D_q\,T_A\,D_p$$

now this implies that T_A is diagonal and/or permutation. To see this, consider two elements in a given row of T_A. Let us take t_{ab} and t_{ac} elements from row a and in column b and c. From (30) we can write the scalar equations

$$(31) \quad t_{ab} = (d^q{}_{aa})(t_{ab})(d^p{}_{bb})$$

$$(32) \quad t_{ac} = (d^q{}_{aa})(t_{ac})(d^p{}_{cc})$$

Now if <u>both</u> t_{ab} and t_{ac} are non-zero we could simplify by dividing both sides of equation (31) by t_{ab}, and both sides of (32) by t_{ac}.

$$(33) \quad 1 = (d^q{}_{aa})(d^p{}_{bb})$$

$$(34) \quad 1 = (d^q{}_{aa})(d^p{}_{cc})$$

from which we would obtain

$$(35) \quad (d^q{}_{aa})^{-1} = d^p{}_{bb}$$

$$(36) \quad (d^q{}_{aa})^{-1} = d^p{}_{cc}$$

or

$$(37) \quad d^p{}_{bb} = d^p{}_{cc}$$

But this contradicts our hypothesis that all diagonal elements of D_p are distinct. Evidently, we cannot take two nonzero elements from any row of T_A.

But since T_A is non-singular, each row has at least one nonzero element. In the same fashion (but by taking two elements from the same <u>column</u>) we can show that each column has at most one nonzero element.

But any non-singular matrix which has only one nonzero element in each row and column is either a diagonal or a permutation matrix, or some combination thereof.

But Δ_x is defined as some matrix which has just this form, so we can let

$$(38) \quad \Delta_x = T_A$$

$$(39) \quad \overset{*}{A} = AT_A = A\Delta_x$$

In the same fashion using symmetry we can obtain

$$(40) \quad \overset{*}{B} = BT_B = B\Delta_y$$

We can see that when two elements of D_p are not distinct, the corresponding factors are not uniquely determined, while the remaining factors (with elements of D_p distinct from all other elements of D_p) will <u>still</u> be uniquely determined. This explains how the uniqueness can break down "in stages" as discovered empirically in Harshman (1970, p39-44).

Application to INDSCAL and other models

By taking the case where matrix B=A, the minimal conditions theorem applies to analysis of scalar product and cross-product matrices, providing that one assumes orthogonal factors. Thus it applies to Carroll and Chang's INDSCAL model (Carroll and Chang, 1970), and provides minimal conditions of uniqueness for INDSCAL solutions. By the same type of interpretation, it provides a uniqueness proof for the orthogonal factor case of PARAFAC2 (for a description of PARAFAC2 see Harshman 1972, this issue). The general oblique proof for PARAFAC2 has not yet been discovered (although progress has been made using reasoning which is along the same lines as the proof reported here).

Further work to be done

Of course, not all questions about the uniqueness of PARAFAC1 have been answered by this minimal conditions theorem. For example, it does not deal with the circumstance where some or all of the D_i matrices are of a lower rank than the full number of factors--i.e. when some factors have a zero influence on some occasions. Empirical results show that this circumstance need not interfere with uniqueness when there are a sufficient number of occasions. Just how many are necessary and sufficient has not been mathematically determined. Of course, if any two of the D_i matrices satisfy the conditions of the preceding theorem, then they will suffice to provide the unique solution. But when no two such D_i exist, uniqueness must be determined by a more complicated set of interdependencies. Jennrich's Uniqueness Theorem applies to this more complicated circumstance whenever there are as many "occasions" as factors. It thus establishes some possible sufficient conditions for uniqueness in such a case. In this respect Jennrich's theorem is "stronger" than the minimal conditions theorem presented here, although it does not handle the situations where there are more factors than occasions, or provide minimal conditions for uniqueness when two D_i are of full rank. In these respects the theorem presented here is "stronger".

Another interesting empirical finding which is not covered by either proof is the discovery that 10 factors can apparently be uniquely determined by an 8 by 8 by 8 data set. This would correspond to the case where the A and B matrices in the minimal conditions proof were horizontal rather than the current stipulation that they are square or vertical. Mathematical treatment of this number of factors in such a sized data set poses the problem of carrying through the proof without the pseudo-inverses of A and B. This, too, is a problem for further work in the future.

References

Carroll, J.D. and Chang, J.J. (1970), "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika* 35, 283-319.

Harshman, Richard (1970), "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," *Working Papers in Phonetics* No. 16, 84 pp.

Harshman, Richard (1972), "PARAFAC2: Mathematical and technical notes," *Working Papers in Phonetics* No. 22

Acknowledgements

Without the constructive criticism of Robert Jennrich and David Cantor (of the UCLA Department of Mathematics), the proof would have been far less clear and rigorous. Their help and interest is greatly appreciated.

The Left Hemisphere is Specialized for Speech, Language, and Something Else

George Papçun, Steve Krashen, and Dale Terbeek

In WPP 19, we reported that

a) Morse operators were clearly left-hemisphere lateralized for dichotically presented Morse code letters ($p < .001$).

b) Naive subjects, taken as a group, were not lateralized; but

(1) The four subjects who were run on a list of all the letters in the Morse code alphabet (hereinafter ALLETTER) showed a slight but insignificant left ear, hence right-hemisphere advantage.

(2) The five subjects run on a restricted list (only the shorter Morse code letters, hereinafter RLIST) tended to show a right ear, left-hemisphere advantage ($p < .03$).

We also noted that the RLIST subjects were tested at a softer intensity than the ALLETTER group.

Since our previous report, we have run six Morse code naive subjects on the RLIST stimuli at approximately the ALLETTER intensity (the louder level). In Table 1 below we summarize our results.

Table 1

Stimuli	Intensity	Subjects	N	Right ear % correct	Left ear % correct	Average accuracy	Percent of total errors contributed by right ear
ALLETTER	loud	Morse operators - instructors	6	50.3	28.0	39.2	40.9
ALLETTER	loud	Morse operators - recent graduates	6	65.8	22.3	44.0	30.6
ALLETTER	loud	naive	4	34.8	40.0	37.4	52.1
RLIST	loud	naive	6	70.8	61.5	66.1	43.1
RLIST	soft	naive	5	52.3	43.4	47.9	45.7

The partial credit scoring procedure explained in our previous report was used in scoring the results for naive subjects.

Under dichotic listening conditions, a certain set of long and short tones (RLIST) is lateralized to the right ear, left hemisphere, for subjects who do not know them as code. (Combining all naive RLIST subjects, $p < .02$) This is the first report of any non-language stimulus to be lateralized to the left hemisphere under dichotic listening conditions. Precisely what the crucial characteristics of these stimuli are that cause them to be lateralized to the left hemisphere is yet uncertain. In general, the ability to make timing or order judgements would seem to be involved. This interpretation is in accord with the findings of Efron that judgements about the order of presentation of brief flashes of light and tactile shocks are made in the left hemisphere (Efron, 1963).

In the last column of Table 1, lateralization is reported in terms of the percentage of the total errors contributed by the right ear. In these terms, a percentage lower than 50% indicates right ear superiority, hence left hemisphere lateralization: a percentage above 50% indicates left ear superiority, hence right hemisphere lateralization. In Harshman and Krashen (1972), this method is explicated and justified as "An 'unbiased' procedure for comparing degree of lateralization of dichotically presented stimuli". This method is claimed to measure degree of lateralization without interference from accuracy variations, as opposed to other measures which give smaller ear difference scores when accuracy is higher. It also corrects for the influence of different amounts of guessing.

For operators and naive subjects on RLIST, the correlation between the percent of errors scores and accuracy is only +.0806, while within groups the percent of errors scores remain consistent. This is further confirmation of the validity of this method of calculating degree of lateralization.

We find that the RLIST soft group and the RLIST loud group differ in their total accuracy; as expected, the loud group has a greater accuracy (66.1%) than the soft group (47.9%). However, the percentage of errors for the right ear remains essentially the same (43.1% and 45.7%, respectively). When we compare the Morse operators and the naive subjects, we find that the Morse operators are more lateralized than the naive subjects.

From our results so far, we conclude that the left hemisphere is specialized for speech, language, *and* something else -- the ability to make timing or order judgements. Next we are led to wonder how these functions are interrelated within the left hemisphere, and how they function together in the production and perception of language.

References

Efron, R. (1963) "The effect of handedness on the perception of simultaneity and temporal order." *Brain* 86, 261-284.

Harshman, R., and S. Krashen (1972) "An 'unbiased' procedure for comparing degree of lateralization" *Working Papers in Phonetics* 22. UCLA.

Papçun, G., S. Krashen, and D. Terbeek (1971) "Is the left hemisphere specialized for speech, language, or something else?" *Working Papers in Phonetics* 19, 69-77.