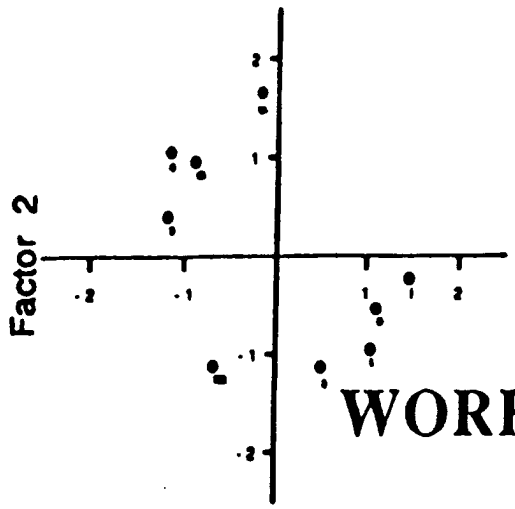
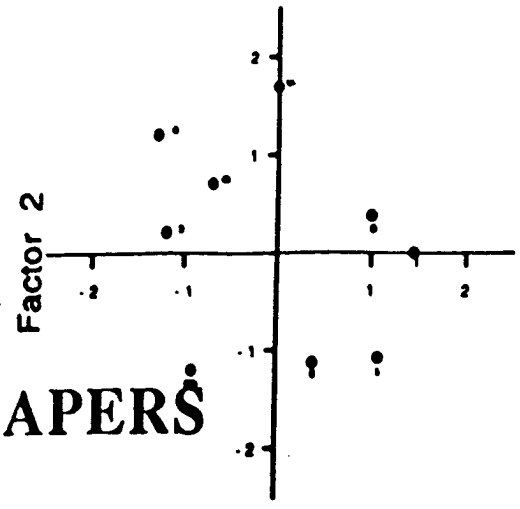
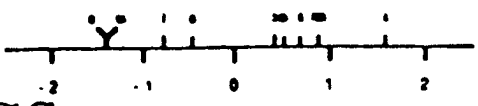


Akan vowels: 2-factor solution



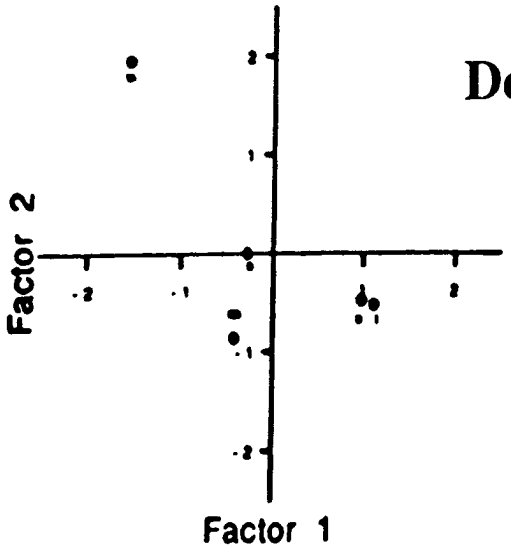


Akan vowels: 3-factor solution

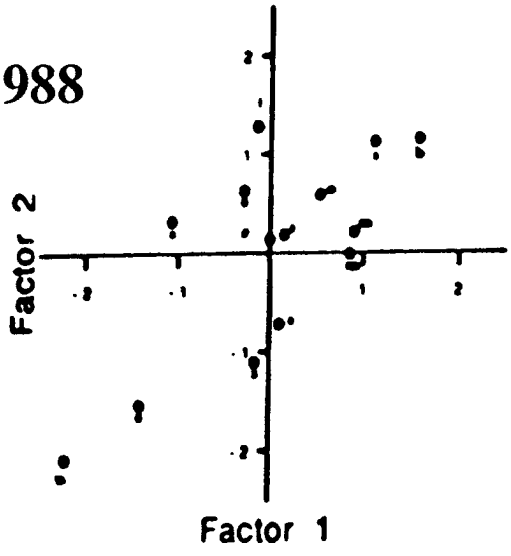
UCLA WORKING PAPERS IN PHONETICS

71

December 1988



Chinese: 2-factor solution



French: 2-factor solution

Phonetic Theory and Cross-Linguistic Variation in Vowel Articulation

Michel T.T. Jackson

The UCLA Phonetics Laboratory Group

Stephen R. Anderson
Alice Anderton
Vernon Austel
Sue Banner-Inouye
Carol Carey
John Choi
Abby Cohn
Sarah Dart
Sandy Disner
Vicki Fromkin
Bruce Hayes
Susan Hess
Susan Hess
Cao Jianfen
Abigail Kaun
Pat Keating
Paul Kirk
Jenny Ladefoged
Peter Ladefoged
Mona Lindau
Ian Maddieson
Bonny Sands
Henry Teheranizadeh
JulieAnna Thaxter

As on previous occasions, the material which is presented in this volume is simply a record for our own use, a report as required by the funding agencies which support the Phonetics Laboratory, and a preliminary account of research in progress for our colleagues in the field.

Funds for the UCLA Phonetics Laboratory are provided through:

IBM grant 538916
NSF grant BNS 87-04361
NSF grant BNS 87-20098
USHHS grant NS-22726
and the UCLA Department of Linguistics.

Correspondence concerning UCLA *Working Papers in Phonetics* should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles, CA 90024-1543
(U.S.A.)

UCLA *Working Papers in Phonetics* is edited by Ian Maddieson.

UCLA Working Papers in Phonetics 71

December 1988

Phonetic Theory and Cross-Linguistic Variation in Vowel Articulation

by

Michel Tah Tung Jackson

Table of Contents

List of Figures

Chapter 1

Chapter 2

Appendix to Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 4

Chapter 5

Chapter 6

ACKNOWLEDGMENTS

Thanks go first of all to my committee: Peter Bentler, Eric Holman, Patricia Keating, Ian Maddieson, and - last but not least - Peter Ladefoged. This dissertation owes much to their guidance.

Many other people have contributed to this work with encouraging words and constructive criticism. A partial list: André Bothorel, Sarah Dart, Richard Harshman, Bruce Hayes, Mona Lindau, Björn Lindblom, Margaret Lundy, Ignatius Mattingly, Terry Nearey, Péla Simon, Johan Sundberg, Eddy Vasile, François Wioland, Jean-Pierre Zerling. There are also the teachers who first inspired my interest in linguistics and phonetics: Louis Goldstein, Alvin Liberman, Samuel Martin. Thanks to all of you and any others I have forgotten.

Marie King Huffman has provided untold support and companionship for the years over which this project unfolded.

Portions of this work will appear in the *Journal of the Acoustical Society of America* 84, under the title "Analysis of tongue positions: language-specific and cross-linguistic models", by Michel Tah Tung Jackson. The dissertation will appear in its entirety in *UCLA Working Papers in Phonetics*. Portions of this work have been supported by NSF Grant BNS-8704361, to Peter Ladefoged and Ian Maddieson; NIH Grant USPHS 5 R01 NS 22726-02, to Peter Ladefoged and Mona Lindau; NSF Grant BNS-8713522, to Peter Ladefoged and Michel Tah Tung Jackson; and a one-quarter fellowship from the UCLA Department of Linguistics.

Chapter 1: Introduction

The scope of the problem

The aim of this dissertation is to investigate and develop an explicit articulatory model of vowel production. Although there has been a fair amount of work on articulatory models since the groundbreaking publication by Stevens & House (1955), there are still many unanswered questions about vowel production. Few workers have considered data from more than one language (Wood 1982 is an exception; Rossi 1983 qualitatively reviews several models based on data from American English and French). Many do not include any measure of how well the model proposed fits the data it is derived from, or any measure of the generalizability of the model. Many are ad hoc constructions based on convenience or some unverified intuition about how vowel articulation ought to be.

There is a fairly broad concensus in the phonetic literature that the traditional classification of vowels in terms of the putative height and backness of the tongue is in fact acoustic or perceptual, rather than articulatory (see, for example, Ladefoged 1976, Nearey 1978, Rossi 1983, and Wood 1982; see Fischer-Jørgensen 1985 and Lindau 1978 for a different view). The most serious criticisms of the traditional height-backness classification seem to be:-

- the notions of 'height' and 'backness' themselves are poorly defined. Various authors speak of 'height of the highest point on the tongue'; others of the 'highest part of the tongue'; others use 'degree of openness'. Similar problems occur in discussion of backness. There is a clear kinesthetic or proprioceptive intuition, but it has resisted rigourous formulation.

Stevens & House (1955) reject height and backness, replacing them with a parameterization in terms of degree of vocal tract constriction and the location of constriction relative to the glottis. This description is claimed to be more directly related to the acoustic properties of vowels, since constriction of the vocal tract cavity dictates the resonances that are characteristic of vowels. This replacement changes our picture of the vowel chart.

For instance, Jones (1956) describes Cardinal vowel 1 (**i**) as "the sound in which the raising of the tongue is as far forward as possible and as high as possible consistently with its being a vowel" and Cardinal vowel 5 (**ɑ**) as "a sound in which the back of the tongue is lowered as far as possible and retracted as far as possible consistently with the sound being a vowel" (p.31). He then goes on to describe Cardinal vowels 2, 3, and 4 (**e**, **ɛ**, and **a**) as "vowels of the 'front' series" and Cardinal vowels 6, 7, and 8 (**ɔ**, **o**, and **u**) as "vowels of the 'back' series". Articulatorily, Jones states that

in passing from vowel No.1 [**i**: mttj] to vowel No. 2 and then to No. 3 and then to No. 4 [**a**: mttj], the tongue is lowered through approximately equal intervals. Also that in passing from No. 5 [**ɑ**: mttj] ... and then to No. 8 [**u**: mttj], the tongue is raised through approximately equal though smaller intervals. (p.32).

Thus, Jones regards the vowels [**i**] and [**ɑ**] as differing along two parameters, height (**i** is high, **ɑ** low) and backness (**i** is front, **ɑ** back). But in Stevens & House (1955) these vowels are regarded as differing along only one parameter - [**i**] has its constriction far from the glottis, [**ɑ**] has its constriction close to the glottis. Stevens & House report that average formant values for the vowels [**i**] and [**ɑ**] as produced by adult male English speakers are produced by setting (i) the mouth opening parameter *A/l* to approximately 3 cm, (ii) the radius of the vocal tract at the point of maximum constriction to 0.3 cm, and (iii) the point of maximum constriction to approximately 11 cm from the glottis for [**i**] and 5 cm for [**ɑ**] (p. 491). The parameters of this articulatory description thus put axes through the space of possible vowels that are rather different from the traditional height and backness parameters.

- the traditional ordering of vowels by height is often contradicted by x-ray data, even under a relatively sympathetic interpretation of height (see Ladefoged 1976, Nearey 1978, Russell 1928, and Wood 1982). Fischer-Jørgensen (1985), attempting to minimize the importance of these contradictions, suggests that height should be interpreted relative to tenseness and rounding, i.e. within the tenseness and rounding categories. Following this conception, the traditional high-to-low ordering of [(tense) **i**, (lax) **ɪ**, (tense) **e**, (lax) **ɛ**], should be relativized to the orderings (tense) [**i**, **e**] and (lax) [**ɪ**, **ɛ**]. Similarly, she suggests that the traditional heightwise ordering of [(rounded) **o**, (rounded) **ɔ**, (unrounded) **ɑ**] should be relativized to (rounded) [**o**, **ɔ**], with the height of (unrounded) [**ɑ**] only considered relative to other unrounded back vowels. With such flexibility in the interpretation of the supposed articulatory property of height, it is hard to see how height is useful at all for describing vowels.
- other classificatory dimensions associated with the height-backness framework, notably the dimension of tenseness, have also escaped explicit formulation in articulatory terms. Jones (1956) remarks

the terms 'tenseness' and 'laxness' probably do not describe accurately the action of the tongue in differentiating certain vowels ... It is generally advisable to apply the terms *tense* and *lax* only to the case of close vowels ... The 'tenseness' or 'laxness' of a vowel may be observed mechanically in the case of some vowels by

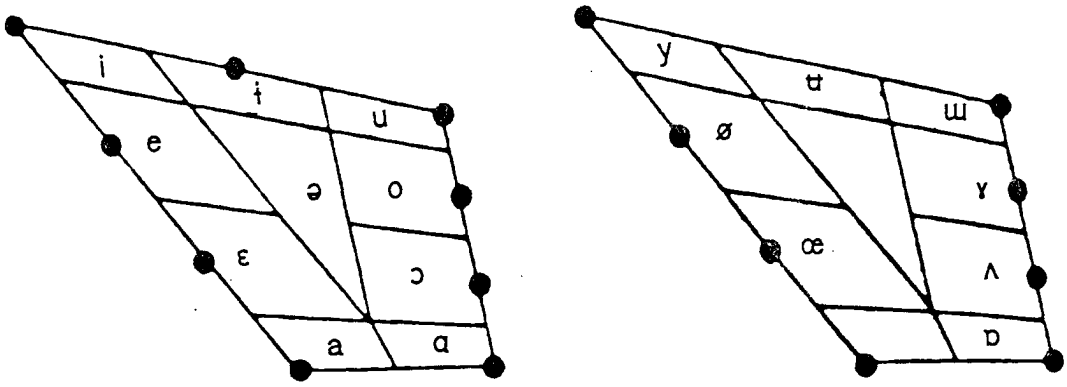


Figure 1.1a. The IPA vowel chart (after IPA 1949).

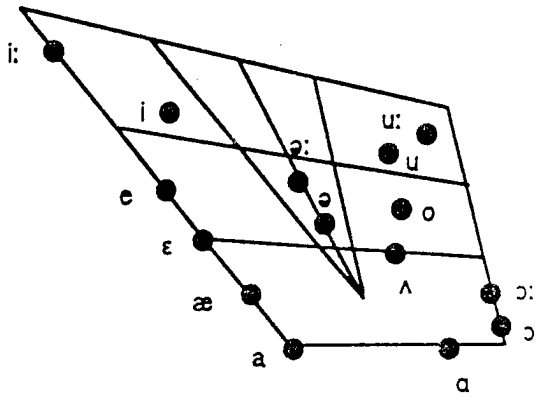


Figure 1.1b. The vowels of Southern British English (after Jones 1956, p. 64).

placing the finger on the outside of the throat midway between the larynx and the chin. When pronouncing for instance, the English **i** (as in *sit*), this part of the throat feels loose, but when pronouncing the corresponding tense vowel (the **i:** in *seat*), the throat feels considerably tenser and is somewhat pushed forward. (p. 40).

Wood (1982) and Fischer-Jørgensen (1985) follow this tradition and suggest that tense vowels have a somewhat more bunched tongue shape and a larger pharyngeal cavity than lax ones. Both of these differences are supposedly due to there being more muscular activity in tense vowels. However, Fischer-Jørgensen acknowledges that electromyographic (EMG) recordings comparing tense and lax vowels are not completely consistent with the hypothesis that tense vowels are produced with greater muscular activity.

- as Wood (1982) emphasizes, the traditional classification ignores pharyngeal maneuvers entirely. To the extent that pharyngeal maneuvers are independent of other articulatory gestures, (as Lindau 1978 suggests), this is an undoubted problem for the height-backness framework. However, if pharyngeal cavity shapes are largely predictable on the basis of the position of the tongue, as Fischer-Jørgensen (1985) suggests, this may not be a great problem.

There are thus a number of ways in which the traditional height/backness articulatory description of vowels is inadequate. However, there is no generally accepted replacement articulatory model or framework for vowel description. An overview of the vowels that an articulatory model must provide an explicit description of follows.

The array of vowel species

The IPA vowel chart (Fig. 1.1a), which has changed very little in the last 50 years, symbolizes about 20 potential phonemic vowel qualities, differentiated by height, backness, and rounding. The vowel symbols are meant to be phonemic - as stated in *The Principles of the International Phonetic Association* (IPA 1949), "There should be a separate letter for each distinctive sound; that is, for each sound, which being used instead of another, in the same language, can change the meaning of a word." The symbols are also clearly meant to represent approximately the same or similar vowel qualities, no matter what language they are being used to represent the sounds of - "When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound."

It seems not to have been felt by most practicing phoneticians that non-phonemic vowel qualities needed systematic symbolization. For instance, the symbol œ - is described by Wells (1975) (the IPA's secretary at the time) as "seldom or never ... needed

for phonemic transcription". The IPA adopted the symbol œ "to fill an awkward gap in our vowel chart" (Wells 1975, p. 52), rather than on the grounds that a symbol is needed for a previously unrecognized phonetic possibility.

The inventory of symbols in Fig. 1.1a, (augmented by œ and diacritics) thus claims to *simultaneously* be both an inventory of the possible phonemic vowel qualities and an inventory of the possible phonetic species - 'similar shades of sound'. There is an implicit claim that the possible phonemic vowel qualities map one-to-one to the possible phonetic species symbolized as regions around the various symbols on the chart (possibly with subdivisions associated with diacritics like "centralized", "retracted", etc.)

However, even a cursory review of the phonetic literature reveals that phoneticians recognize many more variations in vowel quality than provided for in the IPA vowel chart. An extensive tradition in British phonetics, associated with Daniel Jones and his students, recognizes the continuous variability of vowels. In Jones' Cardinal Vowel (Jones 1956) framework for describing vowels, the vowels marked by the large dots in Fig. 1.1a are considered landmarks with reference to which vowels should be described, and not particularly possible vowel phonemes.

For example, Jones' (1956) description of the vowels of Southern British English (SBE) is shown in Fig. 1.1b. Jones gives some 15 possible phonetic realizations of the (about) 8 vowel phonemes of SBE. Although a few of these phonetic vowel qualities are apparently identical to the qualities of the Cardinal vowels (e.g., [ɛ]), and many appear to be on the edge of the vowel space (especially the front vowels), there is no general relation between the vowel qualities Jones recognizes and the inventory of qualities provided for in Fig. 1.1a. The prototypical vowel qualities of a language (the canonical realizations of its phonemic vowels) may occur anywhere in the vowel space.

Thus, Jones and others would maintain that the claim that a finite number of prototypical phonemic vowel qualities can be mapped one-to-one to various discrete regions of the vowel chart is not tenable. Modern instrumental phonetic work has largely substantiated the continuously variable nature of phonemic vowel qualitites, showing both that there are numerous variations in "shade of sound" and that phonemic vowel qualities may (apparently) be arbitrarily close.

For instance, Disner (1983) has investigated some of the claims in the phonetic literature about differences between Germanic languages, generally substantiating them with acoustic data. Among the claims she investigated were

- Dutch [ɪ] is slightly more open than English [ɪ] (Koolhoven 1968).
- Norwegian [ø] resembles German [ø] but is also less rounded (Haugen 1935).
- Swedish [i] is closer than English [i] (McClean 1969).

- [ɛ] is somewhat more open in Dutch than in German (ten Cate, Jordens, & van Lessen Kloeke 1976).
- Danish [ø] is closer than Swedish [ø] (Nielsen & Hjorth 1971).

Not only did Disner show that these claims (reinterpreted in acoustic, rather than quasi-articulatory terms) are correct, she also showed that the prototypical qualities of the various phonemic vowels of these languages vary greatly from language to language. What counts as an /ɛ/ in English might well count as an /æ/ in Swedish. The prototypical qualities of the phonemic vowels are not clearly classifiable.

Other claims about vowel variation within and between languages may be culled from the literature at will. Many of variations are cited relative to the vowels of one language or another, given the difficulty of finding landmarks in the vowel space (the Cardinal vowels notwithstanding). Some remarks relevant to the description of the vowels of languages that will concern us in this study follow.

- Danish & Swedish have 'abnormal' rounding (Henderson's 1971 edition of Henry Sweet's writings).
- Scandinavian and Russian front vowels are sharper than English front vowels (Fant 1960).
- Arabic uses a prepalatal place of constriction for [i], whereas English uses a midpalatal one (Wood 1978).
- in Spanish (chiefly Castillian) pronounciation, "on entendra ... un *e* qui seri moins fermé que l'*e* fermé du sud de la France, et un *e* qui sera moins ouvert qu'un *e* ouvert italien." (Josselyn 1907)
- "l'*i* castillan est plus relâché que l'*i* français." Colton (1909)

These variations are claimed to be *systematic* differences between languages. It is clear from claims like these, and especially from the acoustic phonetic work of Disner (1983) that the species often presumed to exist within the traditional height-backness framework are not very clearly defined. The phonetic space (even of phonemic, 'prototypical' qualities!) is *continuous*, not discrete. It is this continuous articulatory vowel space that phoneticians have been concerned with, and that this dissertation investigates.

Some previous approaches

The investigation of vowel articulation has long been tied to investigation of the acoustic and dynamic properties of speech. A number of investigators concerned with these properties have constructed explicit models of vowel (and speech) production. Typically, modelers have concentrated on reproducing the shape of the vocal tract in the midsagittal plane, determined from x-rays. We can crudely classify the models by their approach to this kind of data: some approximate midsagittal profiles of the vocal tract

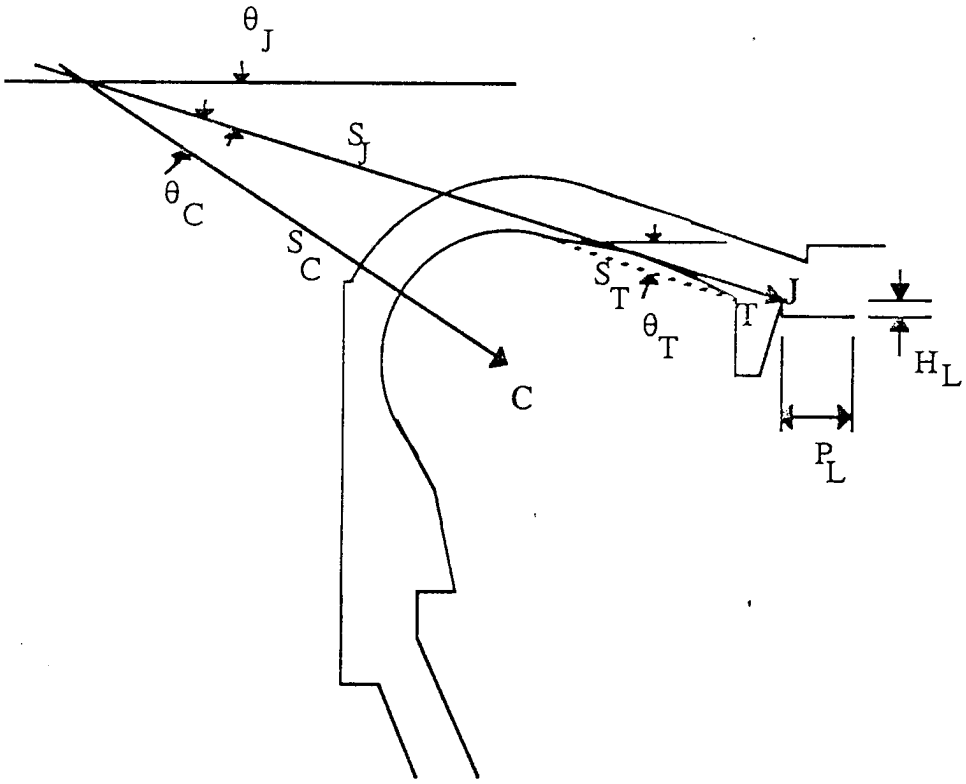


Figure 1.2. Elements of Mermelstein's (1973) model.

with various kinds of simple geometric elements; some use a small set of x-ray data as a basis for generating a range of vocal tract profiles; some pursue models that incorporate biomechanical constraints; and some adopt phenomenological procedures for parameterizing the data.

Despite the diversity of modeling techniques used, there do seem to be a few recurring conclusions about the articulatory organization of vowel production. One important result that we will refer to again is that a family of tongue shapes that ranges from an [i]-like position to an [ɑ]- or [ɔ]-like position is found in data from several different languages. Biomechanical models, taken together with electromyographic recordings, suggest that this family of tongue shapes is primarily due to variation in the degree of activity of the posterior and medial portions of the genioglossus. Perhaps because of its simple physiological implementation, this parameter of vowel articulation appears to be stable across languages and speakers in models that attempt to fit articulatory data directly.

Stevens & House (1955) devised one of the first models of any kind, assigning every vocal tract shape three parameters, based on the distance from the glottis and size of the maximal constriction of the vocal tract (their parameters d_0 and r_0), and the acoustic impedance of the oral opening (their ratio A/l). Strictly speaking, their model does not represent articulatory positions, but rather concentrates on modeling the radius of the vocal tract, using a parabola to interpolate between the fixed radius at the glottal end of the vocal tract, the radius at the point of maximal constriction, and the radius at the lip end of the vocal tract (determined from the A/l ratio).

A widely used model based on work done at Bell Labs by Coker & Fujimura (1966) and subsequent work (Coker 1967, 1968; Coker, Umeda, & Browman 1973; Mermelstein 1973; and Rubin, Baer, & Mermelstein 1981) attempts to model the midsagittal profile of the vocal tract with a set of geometric elements - arcs and lines. Fig. 1.2 (after Mermelstein 1973) shows some of the major elements of this model. In this form, the model has nine parameters, representing positions of the hyoid bone, velum, tongue body, tongue tip, lips, and jaw. None of these models has had explicit goodness-of-fit measures reported.

Nearey (1978) used one of the major elements of this model - a fixed-radius circular arc representing the body of the tongue - in an attempt to test the validity of the height-backness description of tongue positions in American English vowels. Nearey plots the range-normalized x- and y- coordinates of the center of the circle for three speakers (Fig. 1.3a). As can be seen by comparing Fig. 1.3a with the IPA-style chart of American English vowels in Fig. 1.3b, there are some discrepancies between the

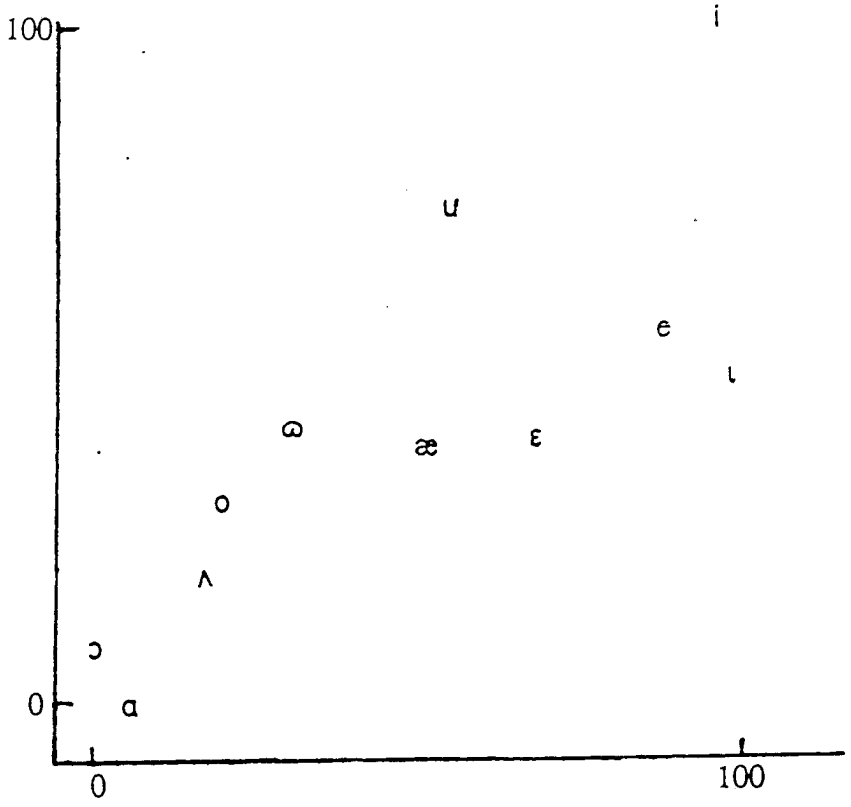


Figure 1.3a. Range-normalized plot of tongue body positions in American English vowels (after Nearey 1978)

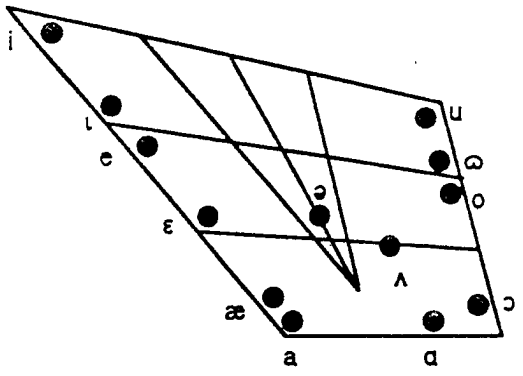


Figure 1.3b. The vowels of General American English (after Jones 1956, p. 356; Jones 1963, p. 206; Kenyon 1956, p. 61).

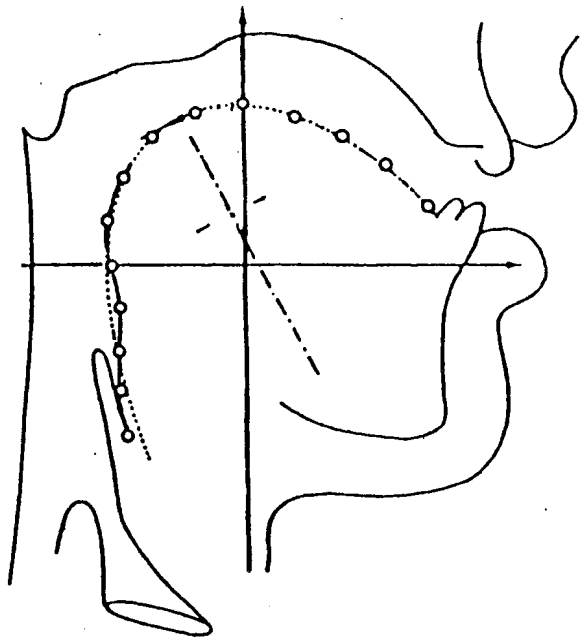


Figure 1.4. Hashimoto & Sasaki's (1982) General Quadratic model for tongue shapes.

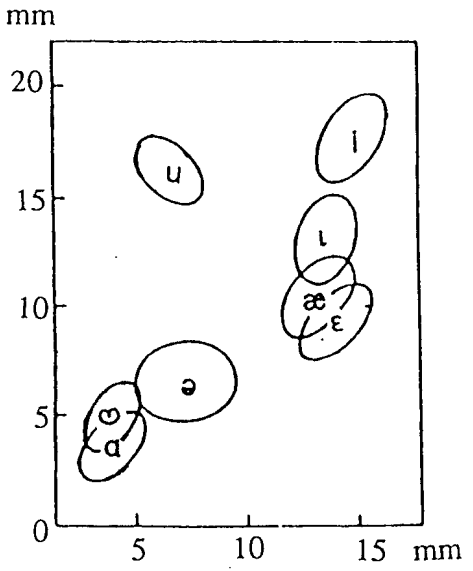


Figure 1.5. Hashimoto & Sasaki's (1982) plot of tongue center position for the vowels in Perkell (1969).

arrangement of vowel according to this articulatory criterion and their arrangement in the traditional height-backness system. For instance, [ɔ] is quite close to [æ] on Nearey's chart. The vowels are strung out on a roughly diagonal line (i.e., the x- and y- coordinates of the tongue body circle are moderately well correlated), with [**i**] at one extreme and [ɑ] at the other.

Fig. 1.3b is based largely on Kenyon's (1956) description of 'General American' but also takes Jones (1956, 1963) into account. In particular, Jones (1956) recognizes two variants of the vowel in *bad*, which he renders [**bad**]: "**a** is often about Cardinal No. 4, but a higher variety resembling the Southern British raised **a** (**æ**) is also common" (pp. 356-357.) (It should be noted that modern practice would use '**æ**' for the low front vowel - Jones' '**a**' - and reserve '**a**' for a low central vowel between Jones' '**a**' and '**ɑ**').

Hashimoto & Sasaki (1982) used quadratic curves (i.e., circles, ellipses, hyperbolas, and parabolas) to represent tongue positions in vowels in Perkell's (1969) x-rays of one speaker of American English (Fig. 1.4). The parameters of this model are the radius, eccentricity, tilt angle, and position of the average center of curvature of the tongue. They report a rms fitting error of 1.5 mm to their data. They remark that their model is not compact: the number of parameters of the model could be reduced, since the tilt angle can be modeled as a polynomial in x- and y- coordinates of the average center of curvature (multiple correlation coefficient $R = 0.92$). They do not plot the vowels with respect to the radius or eccentricity of the fitted quadratic, though apparently the radius varies from about 12 mm to something near 20 mm with a mean value of 18.2 mm, and the eccentricity varies from under 0.8 (giving an ellipse) to greater than 1.0 (giving a hyperbola) with a mean of 0.85. Their plot (Fig. 1.5) of the position of the average center of curvature of the tongue is not unlike the traditional vowel chart, though the placement of [ə] near [ɔ] and of [æ] higher than [ɛ] is somewhat surprising.

A model proposed by Ishizaki and coworkers (Ishizaki & Nakajima 1977, Ishizaki, Fuchi, & Nakajima 1977) also relies on geometric elements - in this case, two fixed-radius circle in the sagittal plane, intended to represent the back and the front of the tongue independently (Fig. 1.6). The model also includes the position of the glottis, which is allowed to move vertically, and the aperture at the lips. Ishizaki & Nakjima (1977) cite errors of between 1 and 3 cm^2 for the Japanese vowels [**i, e, a, o, ɯ**], though they do not say what their source of articulatory data is. The error values they report are very large; cf. the 1.7 mm rms error reported in Sekimoto, Imagawa, & Kiritani (1978) below.

Less geometric and more naturalistic in inspiration are the models proposed by Lindblom and his coworkers (Lindblom & Sundberg 1971, Lindblom, Paauli & Sundberg

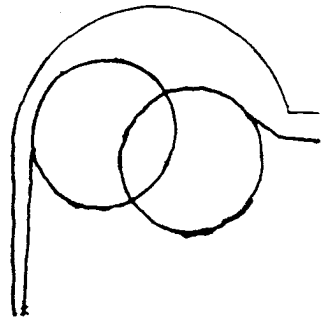


Figure 1.6. Ishizaki's (1977) model.

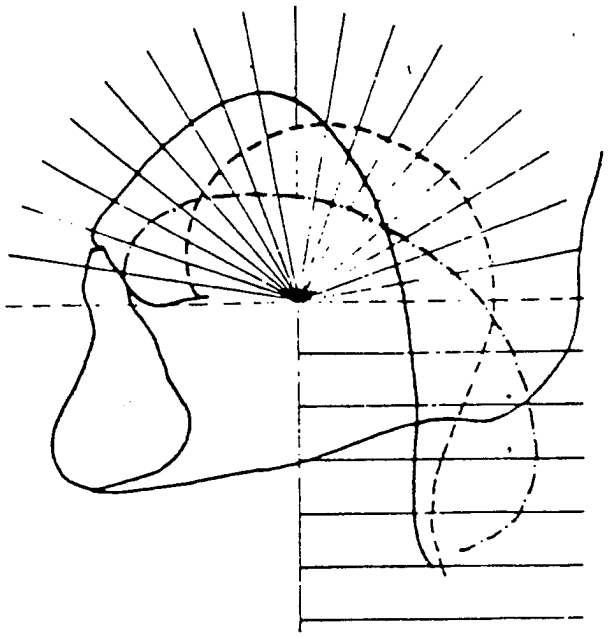


Figure 1.7. Lindblom & Sundberg's (1971) model; tongue shapes are generated by interpolationg along the gridlines.

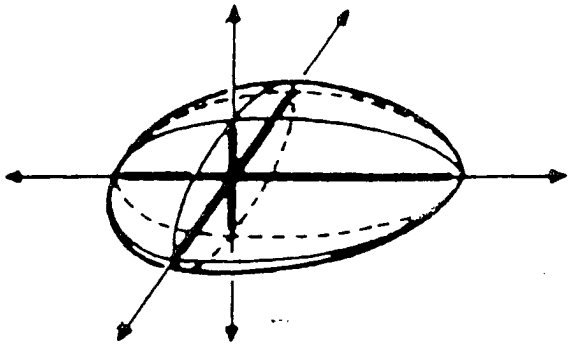


Figure 1.8. Hiki & Oizumi's (1974) model.

1974), Nakajima (1978), and Wood (1982). In all of these models, the articulatory space is modeled by interpolating between various extreme shapes selected from x-ray data.

Lindbolm & Sundberg (1971) model vocal tract shapes for Swedish vowels by interpolating between a neutral tongue shape, and the shapes for [**i**], [**a**], and [**u**]. As Fig. 1.7 suggests, they use a jaw-relative coordinate system for describing tongue shapes. The model generates a range of shapes by interpolating between the given tongue shapes along the gridlines shown in the figure. Lindblom, Pauli, & Sundberg (1974) added lip width, vertical midsagittal lip separation, tongue tip protrusion, tongue tip elevation, and larynx height to the model in their investigation of vowel coarticulation. The model is based on x-rays from two speakers of Swedish, and no fit figures have been published.

Wood's (1982) text proposes a model in which there are four places of articulation (or directions of tongue movement), namely palatal, palatovelar, pharyngovelar, and lower pharyngeal (Wood also calls these palatal, velar, upper pharyngeal, and lower pharyngeal constrictions. Wood (1982) subdivides the palatal place of constriction into pre- and mid- palatal places). However, for the purpose of acoustic modeling, he uses interpolation between observed vocal tract shapes. He has not provided any explicit formulation of the model he proposes.

Henke's (1966) model is widely considered a landmark in dynamic studies of speech. Within the framework of a highly interactive computerized graphic representation, Henke used biomechanically reasonable constraints on tongue deformation to model the x-ray data (later published) in Perkell (1969). The control parameters of this model were mandible position and the position of a large number of points on the surface of the tongue and lips; however, there is also a substantial number of covert parameters that represent tissue elasticity, characteristic articulator response times, articulator-specific maximum velocities, and so on. These parameters determined overall system behaviour and values for them are determined *ad hoc*, but they were not regarded as control parameters. Henke himself remarks that the model was a kind of experimental piece of software that continually changed as new data were considered. It was never frozen and quantitative goodness-of-fit figures are not given.

Hiki & Oizumi (1974) straddle the line between biomechanical inspiration and geometric implementation with their articulatory model. The body of the tongue is modeled as an ellipsoid of constant volume. Forces imposed on the body of the tongue by three intrinsic and five extrinsic muscles distort this ellipsoid in various ways under the constant volume constraint, as suggested by Fig. 1.8. The superior and inferior pharyngeal constrictors form a group of muscles assumed to act together, and the jaw and lips are actuated by six more muscle groups. Electromyographic data, estimated area

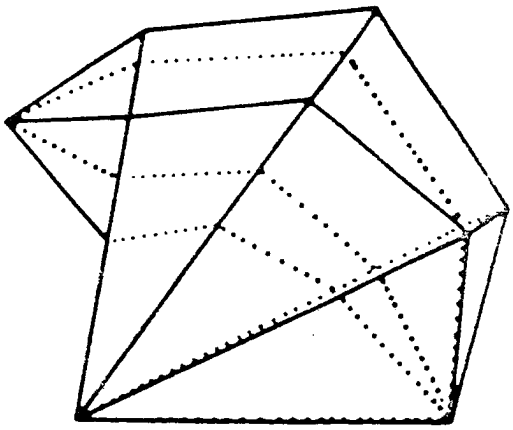


Figure 1.9. Side view of Kakita et al.'s (1985) solid model of the tongue body.

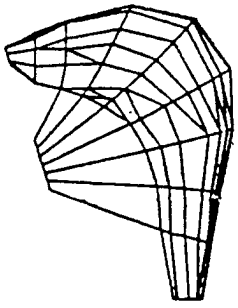


Figure 1.10. Side view of Hashimoto & Suga's (1986) solid model of the tongue.

functions, vocal tract profiles derived from x-rays, and acoustic output were all used to estimate the degree of contraction of each muscle or muscle group. No fit figures are given.

Another biomechanically inspired model is represented by the work of Fujimura and others, also proceeding at Bell Labs (Kiritani, Miyawaki, & Fujimura 1976; Kakita & Fujimura 1977; Fujimura & Kakita 1978; Kakita, Fujimura, & Honda 1985). This approach attempts to model the muscular forces and structure of the vocal tract by using finite-element methods adopted from mechanical analysis and numerical simulation. The tongue body is represented as a collection of tetrahedra that are acted upon by muscular forces and also by each other. Fig. 1.9 (redrawn from Kakita et al. 1985) shows the major functional groupings of the tetrahedral elements. Despite the computational sophistication of the method (the tongue body is modeled with 86 tetrahedra that are acted upon by 18 muscles), model fit values have not been reported. Kakita et al. (1985) simply report "A trial-and-error procedure obtained a suitable consistency among (1) EMG magnitudes reported [by Alfonso, Honda, Baer, & Harris (1982): mttj], (2) tongue shapes expected from x-ray data, and (3) formant frequencies ..."

A similar model is discussed by Hashimoto & Suga (1986). Their model consists of 492 tetrahedral elements connected at 170 nodes acted upon by 13 muscles (Fig. 1.10). They fit the model's midsagittal contour to x-ray tracings of the vowels [**i, ɛ, a, ɔ, u**] from Perkell (1969, 1975) and report the fit by a log data/error ratio analogous to the more usual signal/noise ratio. The data/error ratio, expressed in dB, ranges from about 40 to about 25 for various vowels. They also compare the muscular tensions inferred from the model to EMG measurements of the Japanese vowels /**i, e, a, o, ɯ**/ (from Miyawaki, Hirose, Ushijima, & Sawashima 1975), after normalizing each value to a percentage of the highest observed value. The comparison is rather disappointing, with rms errors as high as 30%.

The more phenomenological approaches to the problem of vocal tract modeling can all be understood to treat individual vocal tract shapes as some sort of vector of coordinates on a measurement grid that is generally aligned with reference to anatomical landmarks. Given a sample of such vectors, the researcher then attempts to find a vector or function space that approximately contains them.

Liljencrants (1971, 1985) has modeled the tongue positions in the Lindblom & Sundberg (1971) data by Fourier cosine series approximation of the position of the surface of the tongue, measured along 30 gridlines spaced approximately 0.5 cm apart. He reports that the first two terms of the Fourier series leave an rms error of approximately 0.12 cm for one speaker, and 0.14 cm for another (cf. Liljencrants' Figs.

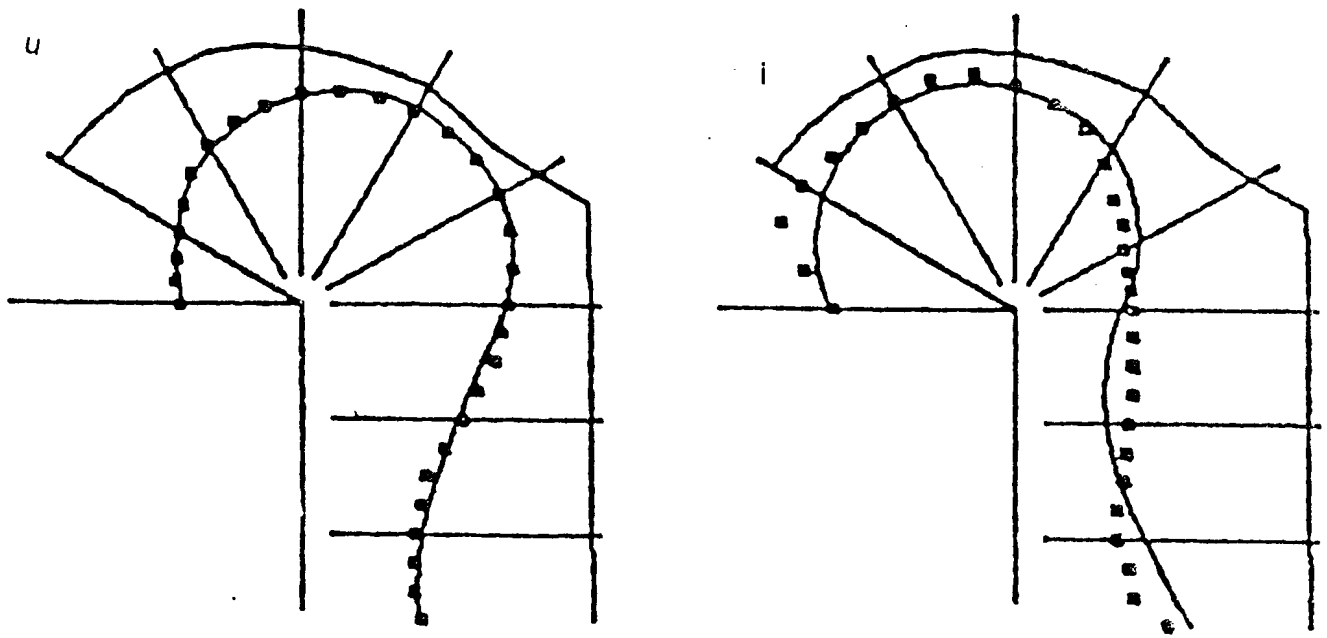


Figure 1.11. The vowels [u] and [i] in Liljencrants' (1971) model.

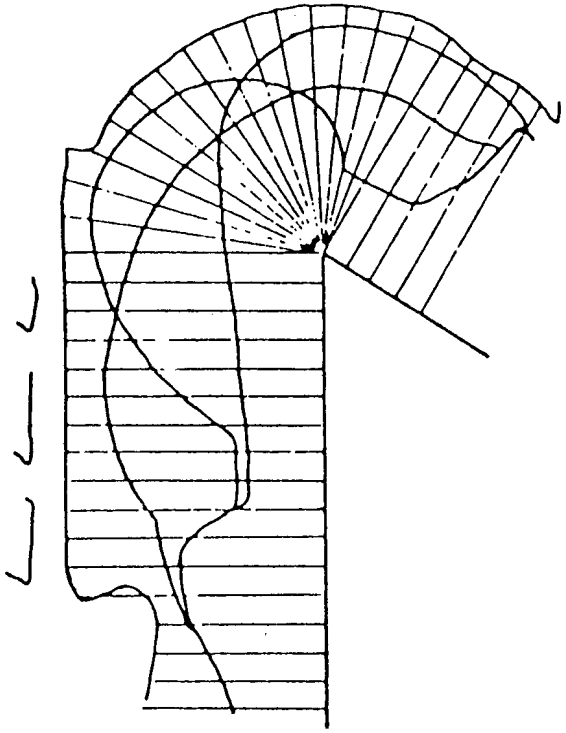


Figure 1.12a. The three basic profiles used by Jospa (1974).

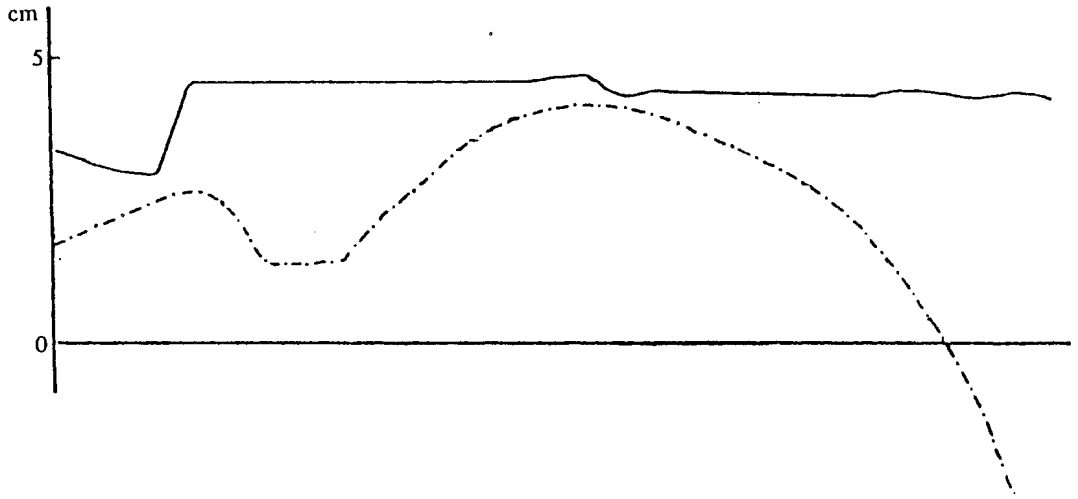


Figure 1.12b. The profile for [u] plotted as a function of the distance from the glottis.

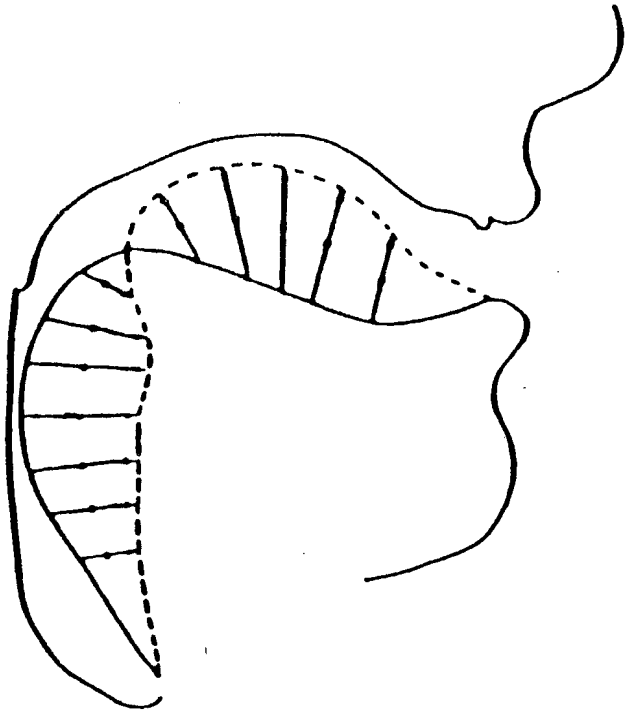


Figure 1.13a. The English Front Raising factor (after Harshman et al. 1977).

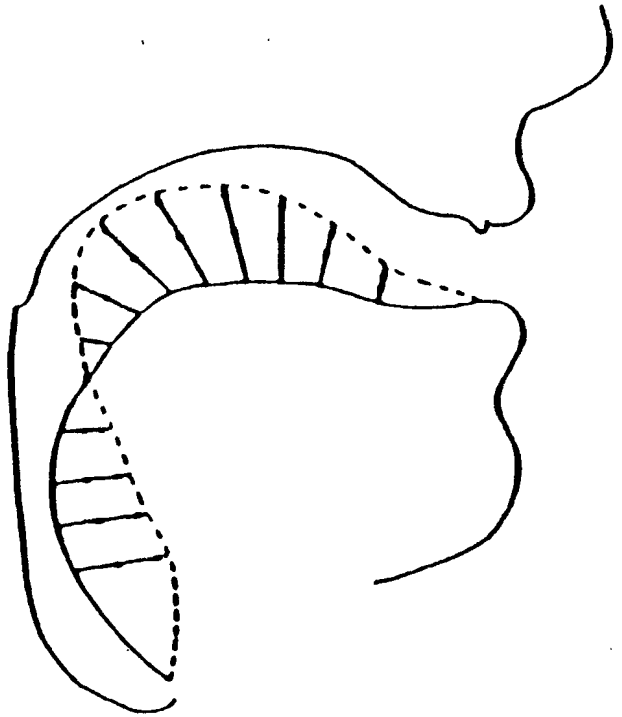


Figure 1.13b. The English Back Raising factor (after Harshman et al. 1977).

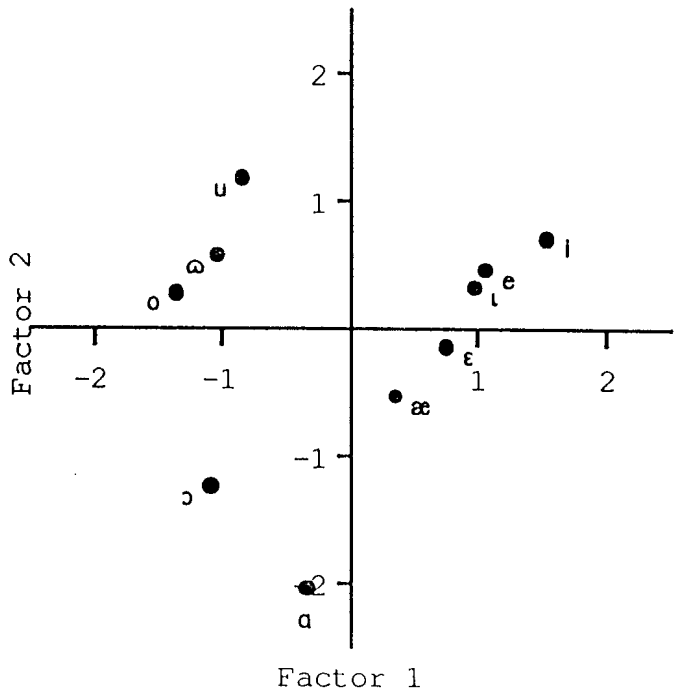


Figure 1.14. English factor loadings - vowels
(after Harshman et al. 1977).

II-A-4 and II-A-5). The zeroth term of the Fourier series description of a particular vowel is the (within-vowel) mean position of the tongue along the gridlines; the next term contributes both a magnitude and a phase, making this a three-parameter model. Liljencrants reports that the higher-order terms appear to vary much more between speakers. He also suggests that the zeroth term can be eliminated, since it is somewhat correlated with the first term's magnitude and phase. Fig. 1.11 shows tongue position profiles for [**u**] and [**i**] and Fourier-series model of them.

Jospa (1974) uses similar Fourier cosine series or alternatively Chebyshev polynomials to fit vocal tract area functions. These area functions are derived from Lindblom & Sundberg's (1971) model with some modifications. Fig. 1.12a shows the entire vocal tract profile for /**i, u, a**/ from the Lindblom & Sundberg data; Fig. 1.12b shows the 'straightened' vocal tract, with the articulatory profile for /**u**/.

The UCLA Phonetics Lab has been the site of a number of applications of the PARAFAC factor-analytic model, which relies on assumptions that we will discuss in Chapter 2, to generate a set of articulatory factors. In the present context, we may assume that the articulatory factors represent basic gestures that individuals may invoke to differing extents to produce specific vowels. The actual displacements generated by the articulatory factors along the measurement gridlines may be viewed as samples from a continuous vocal tract displacement function associated with changes in the activity of functionally coordinated groups of muscles. The parameterization of each vowel - a linear composite of the actual measurements - represents this inferred change in activity.

Harshman, Ladefoged & Goldstein (1977) applied the PARAFAC model to x-ray data on ten vowels produced by five speakers of American English. Two parameters account for over 96% of the variance in the data: one represents a tongue front-raising gesture, with an [**i**]-like posture at one extreme and an [**o**]-like posture at the other; the second parameter represents a back-raising gesture, with an [**u**]-like posture at one extreme and an [**ɑ**]-like posture at the other. Fig. 1.13 shows the articulatory displacements associated with the parameters; Fig. 1.14 shows the vowel space. Again, comparision with Fig. 1.1b or with Fig. 1.3b shows some differences: the order of [**ɪ**] and [**e**] is inverted, and [**æ**] is not nearly as 'low' as [**ɑ**].

Jacobson (1978) and Lindau (1986) investigated languages in which the position of the tongue root has been claimed to be phonologically important, using procedures similar to that of Harshman et al. Jacobson (1978) investigated 8 speakers of DhoLuo, finding that at least four factors were necessary to represent his articulatory data in the PARAFAC model. One of these factors was roughly similar to the Harshman et al. front-raising factor; Jacobson says that two more of the factors are needed are needed to

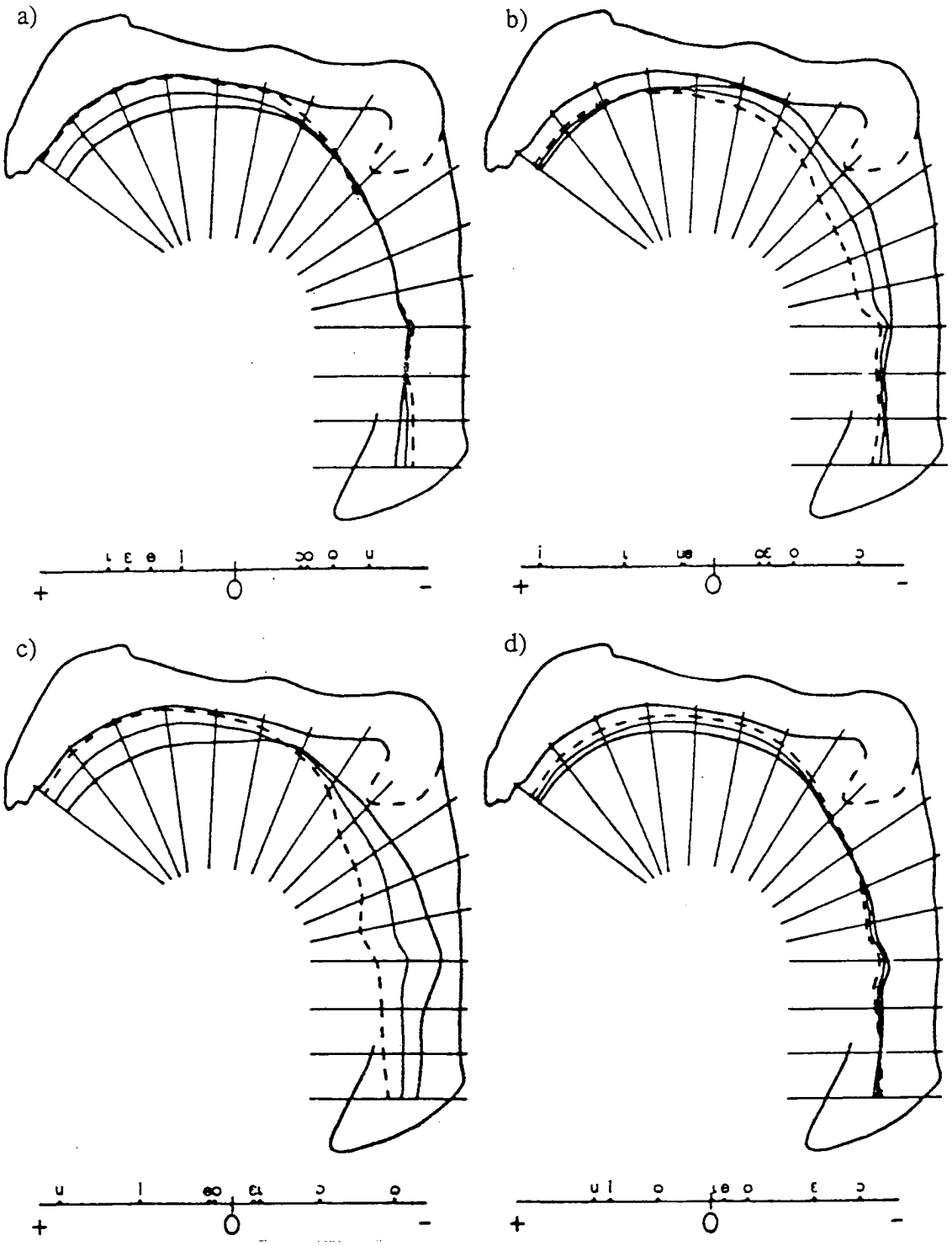


Figure 1.15. Tongue displacements and vowel loadings in Jacobson's (1978) study of DhoLuo.

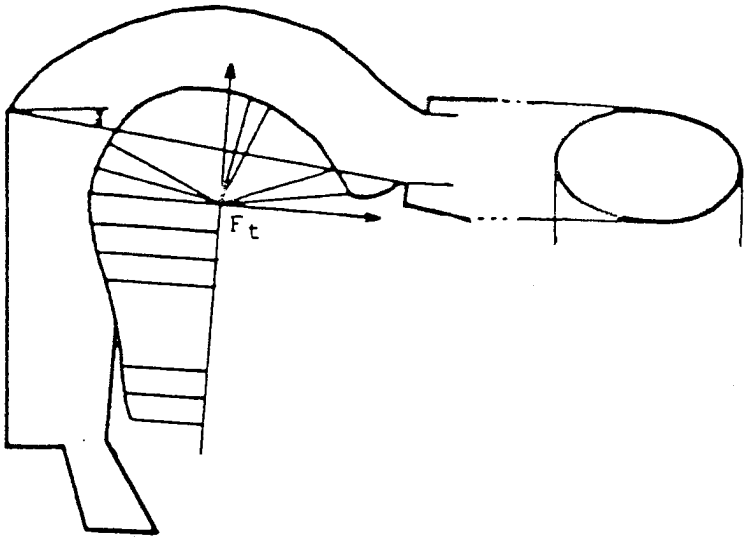


Figure 1.16. Shirai & Honda's (1978) articulatory model.

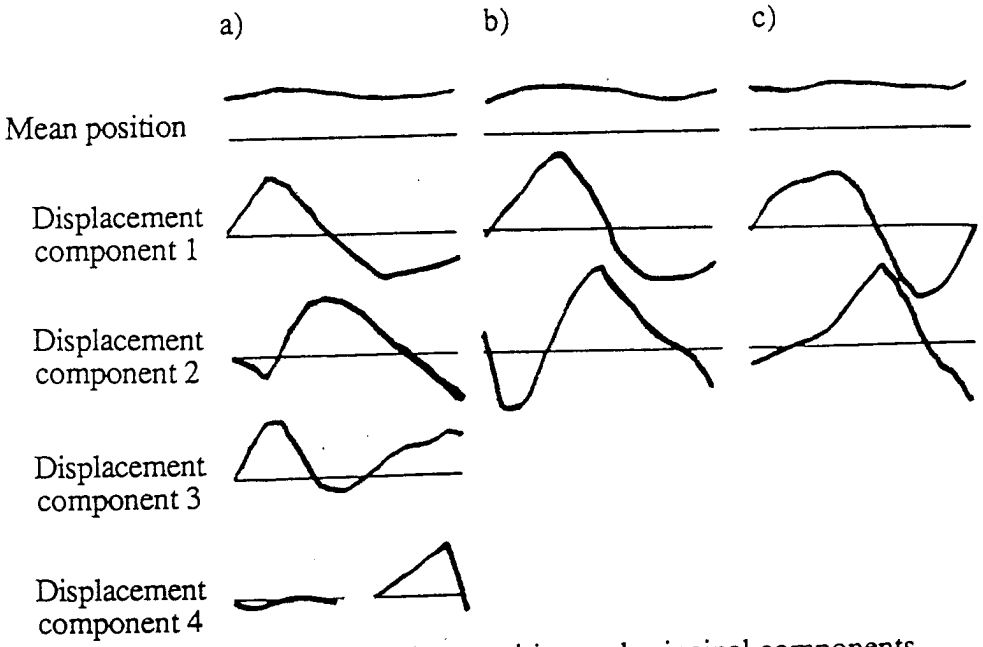


Figure 1.17. Mean position and principal components of tongue position in a) Perkell (1969); b&c) two Japanese speakers (after Shirai & Honda 1978).

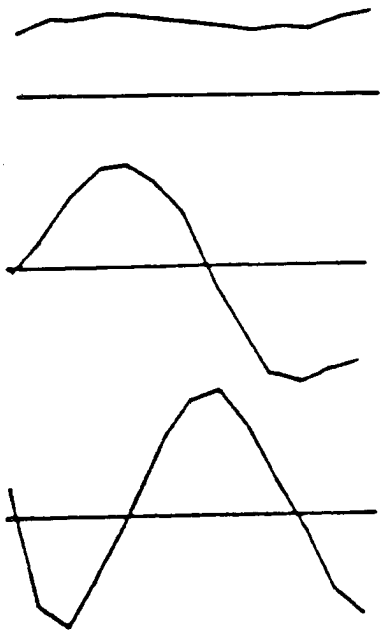


Figure 1.19. Shirai & Honda's (1978) standard model.

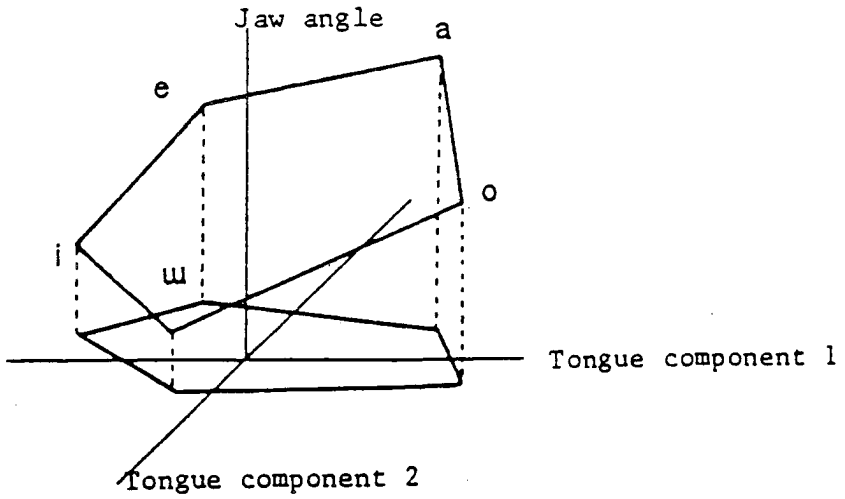


Figure 1.20. Japanese vowels in jaw angle-tongue component space (after Shirai & Honda 1978).

account for speaker differences in vowel height variation. The last factor, which he interprets as the Advanced Tongue Root factor, involves mainly contraction of the lower pharynx. Lindau (1986), working on data from four speakers of Akan, finds only two factors. Again, one is much like front-raising; the other appears to represent some sort of general contraction/expansion of the pharynx. Fig. 1.15 - Jacobson's 'analogical representation' shows tongue displacements proportional to those generated by his four factors.

Shirai & Honda (1978) used the measurement scheme shown in Fig. 1.16 to measure x-rays of one American English speaker (again, the data from Perkell 1969) and ten adult male (presumably Japanese, but the source is not cited) speakers. They derived the mean positions and principal components from tongue positions in vowels. For the English speaker, the principal components associated with the vowels were subtracted from the measurements on the x-ray profiles from consonants, and the residual was again subjected to principal component analysis (Fig. 1.17). A minimization procedure was used to adjust the center of the tongue coordinate system (F_t in Fig. 1.16) so as to bring the ten speakers' principal components into agreement, yielding the 'standard model' shown in Fig. 1.19. A plot of the five phonemic vowels of Japanese, according to the parameters of the tongue model and the jaw angle, is shown in Fig. 1.20.

Sekimoto, Imagawa, & Kiritani (1978) applied regression techniques and principal component analysis to the positions of radio-opaque pellets attached to the tongues of three speakers of Japanese. Data from one speaker is shown in Fig. 1.21. They proceeded by first removing the portion of the tongue position measurements that was linearly related to jaw position by regression (Fig. 1.22a). The residuals were then analyzed into principal components (Figs. 1.22b&c). The approximate error using the jaw component alone was 6 mm rms. The residual after the removal of the first tongue component was 1.7 mm rms; after the second, 0.6 mm. They decided that the second component was not significant, and retained only the jaw and first tongue components. The five vowels of Japanese are plotted according to their loadings on these components in Fig. 1.23.

Other workers who have performed this kind of regression and principal component decomposition of x-ray profiles are Maeda (1978, 1979) and Zerling (1979). Fig. 1.24 shows the results of Zerling's (1979) study, the first two principal components of tongue position in French vowels.

Maeda (1978) performed principal component analysis of tongue position measurements based on cineradiographic data on French vowels from one speaker (Brichler-Labaeye 1970), retaining three components. The first two components are very

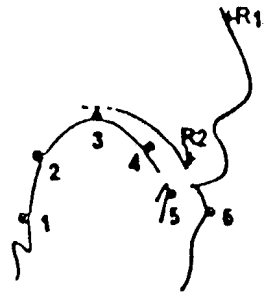


Figure 1.21. Tongue pellet location for the first of Sekimoto et al.'s (1978) Japanese speakers.

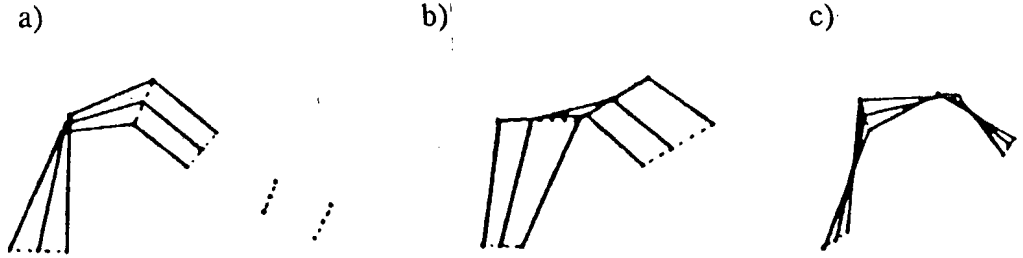


Figure 1.22. a) Tongue position as a linear function of jaw position; b&c) First and second principal components of the residual tongue position for the first of Sekimoto et al.'s (1978) Japanese speakers.

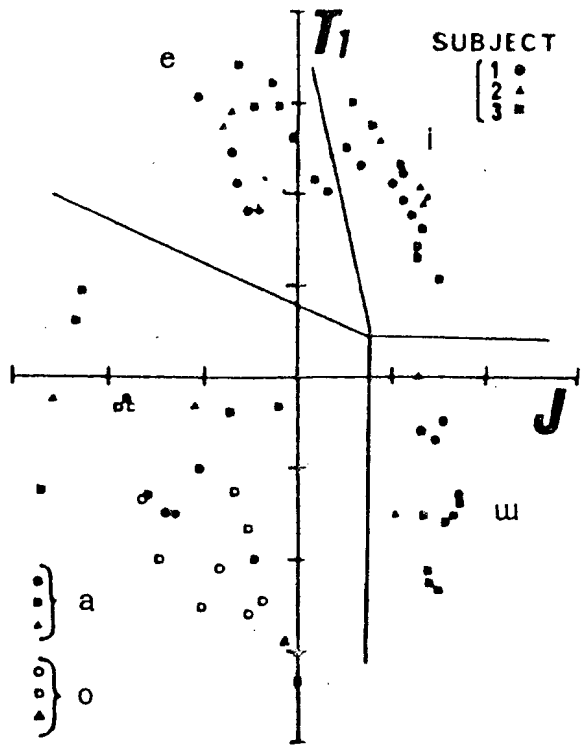


Figure 1.23. Japanese vowels in jaw position-tongue component space (after Sekimoto et al. 1978).

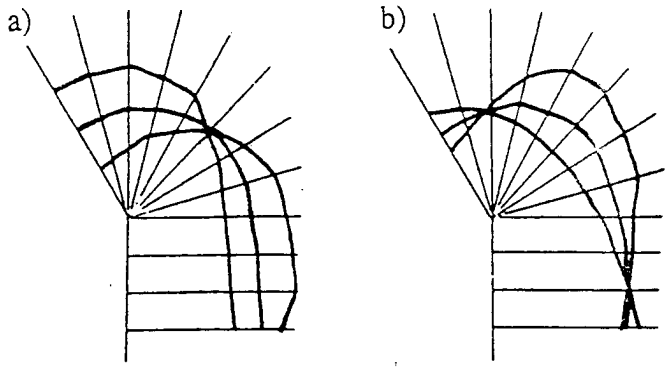


Figure 1.24. Zerling's (1979) principal components of tongue position in French.

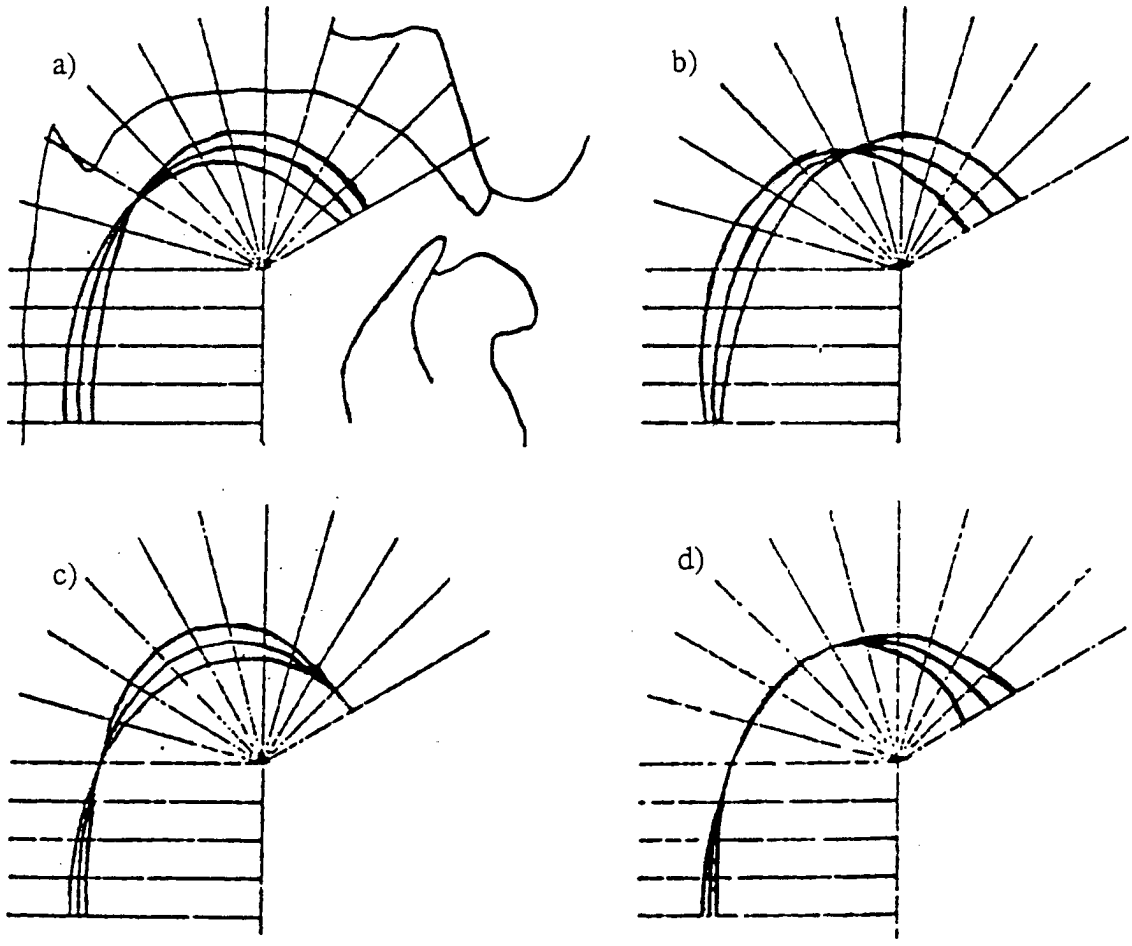


Figure 1.25. a) Tongue position as a linear function of jaw positions; b,c&d) Tongue body, blade, and tip factors in Maeda's (1979) model of Brichler-Labaeye's (1970) French vowels.

similar the components in Fig. 1.24 from Zerling (1979); the third component appears to represent tongue tip motion. Maeda (1979) analyzed the same data using a diagonal factoring (Gorsuch 1983, p.74 ff.; Overall 1962) procedure similar to that of Sekimoto et al. (1978). The first component is defined as the change in articulatory configuration linearly related to jaw position; the second component is defined as the change in articulatory configuration linearly related to the next most important articulatory measurement, etc. The change in tongue position due to jaw position is shown in Fig. 1.25a; the (very similar) first component of tongue displacement is shown in Fig. 1.25b; and the next two components are shown in Figs. 1.25c&d. It is evident that Zerling's (1979) first component (Fig. 1.24a) and Maeda's (1979) jaw and first tongue component (Figs. 1.25a&b) are quite similar. Some French vowels are plotted in Fig. 1.26.

Some trends are evident in these studies. First, the first tongue position component (Figs. 1.13a, 1.15a, 1.22a, 1.24a, 1.25b) appears to be robustly replicated in many languages and across different analysis techniques. This component generates a range of tongue positions that varies from an [**i**]-like position at one extreme to an [**o**] at the other extreme (Figs. 1.14, 1.20, 1.23).

The second component of tongue position is more variable (Figs. 1.13b, 1.15b, 1.19, 1.22bc, 1.24b, 1.25c). It is not clear how these components are related. In some languages, they appear to involve large movements of the tongue towards the velum. But the direction and relative magnitude of the movement appears to vary.

How to interpret and substantiate these observations is the topic of the next chapter.

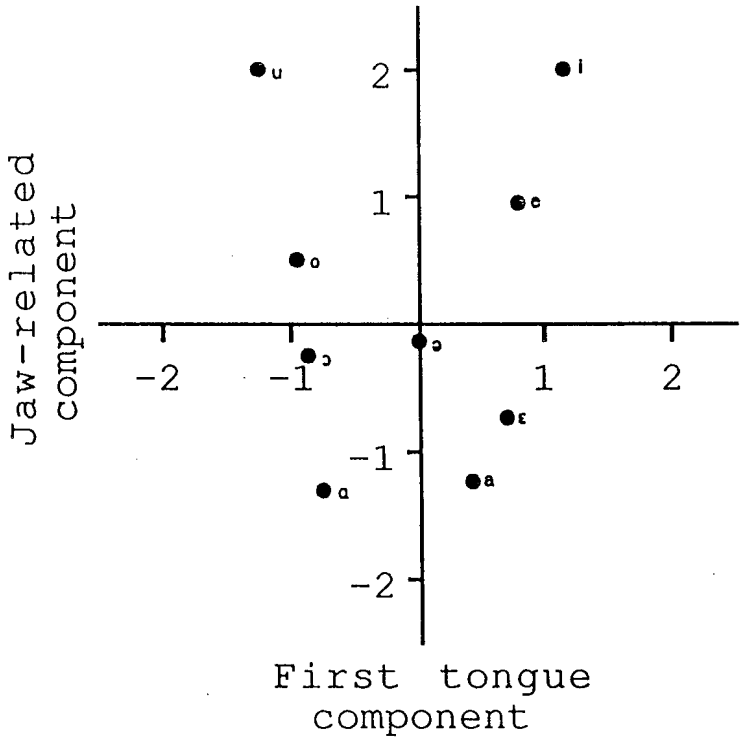


Figure 1.26. Selected French vowels in jaw position-tongue body space (after Maeda 1979).

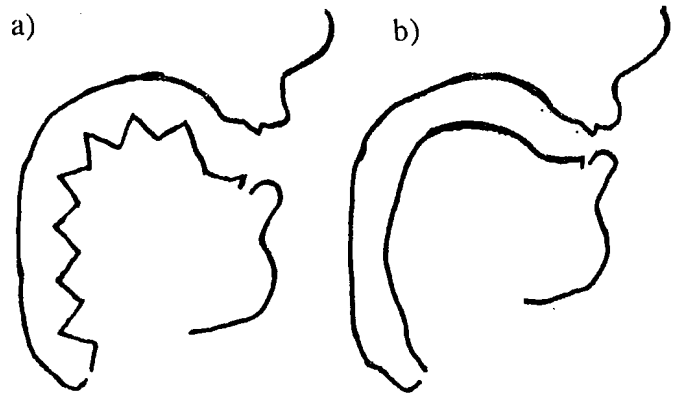


Figure 2.1. a) An unrealistic vocal tract shape that could be produced by Harshman et al.'s (1977) model. b) A realistic vocal tract shape.

Chapter 2: A model for vowel articulation

Rationalizing previous results

Despite the apparent replicability of one articulatory factor, examining the results of comparable analytic procedures in different languages suggests that further results do not generalize in any simple way. For instance, Harshman et al. (1977) and Lindau (1986) applied identical statistical analysis procedures, but arrived at different and non-overlapping sets of parameters for describing the vowels of English and Akan. Descriptions of phonological contrasts and processes - phrased in terms of the distinctive features - need not take this cross-linguistic variation into account (Ladefoged 1982, Keating 1984), but ultimately it must be accounted for in describing the phonetic entities of languages, i.e., in the phonetic component of grammars. Just as stop aspiration is not distinctive in English, is not referred to in phonological rules, but nonetheless is used systematically and must be accounted for in a description of the phonetic entities of English; so also non-distinctive, systematic variation in vowel production must be accounted for.

Thus, we must recognize several levels of description of the speech production process, each of successively greater abstraction. At the first level of articulatory description are raw measures of articulator state, e.g., the position of the highest point of the tongue, the distance between the tongue blade and the hard palate, etc. Measures of articulator velocity, acceleration, etc., may also be relevant to the characterization of an articulator's state in particular contexts, but we will not be concerned with those measures here.

These measures of articulatory state are in general related to several elementary articulatory gestures. For instance, a high tongue position in the midsagittal plane could be the result of intrinsic tongue muscle contraction (since lateral compression of the tongue by the lateralis tends to push up the center of the tongue); extrinsic tongue muscle activity (e.g., contraction of the genioglossus), or an essentially passive effect of jaw raising. Indeed, a high tongue position could be a result of all three of these articulatory gestures. Some articulatory gestures may be transparently related to raw articulatory measures; for instance, the jaw is generally considered to have only one (rotational) degree of freedom in positioning. Thus it is easily characterized in terms of an angle or a distance away from some reference position.

But this is not the general case. We must recognize that most articulatory measures are in fact composite measures affected by several articulatory events. Lower lip position is generally a composite measure, affected by both jaw position (i.e., the substantial part-whole correlation between lip position and jaw position) and intrinsic lip activity.

Thus, in general, what we think of as independent articulatory gestures are not directly observable or measurable. As will become clear below, this is a matter of principle as well as being an effect of the inadequacy of our observational capabilities.

Nonetheless, (potentially) independent articulatory gestures occupy an important place in the description of articulatory events, and for this reason, I will refer to them as *articulatory primes*. I will make a number of assumptions about articulatory primes. First, they are universal. Insofar as all speakers share the same gross anatomical structures, and are capable of employing that anatomy in parallel ways to produce parallel articulatory states, all speakers of all languages have the same space of potential articulatory states. Thus, although articulatory primes in general are not independently observable, they span the space of articulatory possibilities.

Second, articulatory primes are the correct objects with which to state systematic phonetic descriptions of articulatory events. In this, my position is similar to that of Browman & Goldstein (1986), though I differ with them in other ways. Most forms of phonological description are too abstract for this task; raw articulatory measures may reflect several conceptually independent effects.

On top of this substrate of articulatory primes, I recognize a level of *coordinative structure*. Evidence that potentially independent articulatory primes may be functionally linked is widespread. To cite but two examples, Riordan (1977) has shown that larynx lowering and lip rounding are often employed together during the production of rounded vowels; and Lindau (1978) has shown that larynx raising goes along with tongue root retraction to produce [-advanced tongue root] ([-ATR]) vowels in Akan. These learned, language- (or even segment-) specific functional linkages are examples of coordinative structures. That the articulatory primes under consideration are potentially independent is shown, for example, by Jacobson (1978). Jacobson shows that DhoLuo speakers do not use larynx raising along with tongue root movement in the production of [-ATR] vowels. Where it exists, this kind of coordination may be teleologically motivated - both Riordan (1977) and Lindau (1978) suggest that the functional linking between larynx movement and lip protrusion or tongue root retraction is due to the fact that the larynx movement accentuates the acoustic and perceptual effects of lip raising and tongue root retraction. Stevens, Keyser, & Kawasaki (1986) illustrate several other language-specific, optional linkings that enhance perceptual contrasts.

The nature of these and other cases suggests that the components of this kind of coordinated gestures are of two kinds:

- articulatory primes that are always involved in the production of a segment or phonological contrast, e.g., lip rounding, tongue root retraction.

- articulatory primes that are susceptible of language- or speaker- specific variation, e.g., the larynx-movement correlates of lip rounding and tongue root retraction.

In my view, coordinative structures find natural expression in terms of functional dependencies between *core articulatory primes* such as tongue root retraction and *helping primes* such as larynx raising. In a sense, I am suggesting that the universal linking conventions proposed by Stevens et al. (1986) should be described at the (phonetic) level of coordinative structure rather than at the (phonological) level of distinctive feature specification. The amount of larynx movement associated with these different coordinative structures appears to be gradient - much more larynx movement is seen in [-ATR] Akan vowels (Lindau, personal communication) than in rounded vowels in French (Riordan 1977). The degree of larynx movement seen in rounded vowels in French is also highly variable from speaker to speaker. As argued in Ladefoged's (1982) discussion of prime features, and in Keating's (1984) discussion of the phonology and phonetics of stop voicing, the gradient and variable nature of the use of helping primes like larynx movement may be taken as an argument that the description of these coordinative structures is properly phonetic and not phonological.

A description in terms of the coordinative structures of a language provides the most parsimonious description of the articulatory goals associated with the vowels (and other segments) of that language. The expansion of a coordinative structure into its component core and helping articulatory primes introduces language-specific redundancies that are definitionally excluded from distinctive-feature representations. The overall structure of this model of speech production can be seen as a series of successive 'unfoldings' of representations with successively greater numbers of (potential) degrees of freedom. In the phonetic component of a grammatical model, phonological representations unfold to language-specific coordinative structures, coordinative structures unfold to articulatory primes, and these primes in turn unfold to deformations of the vocal tract. Others, e.g. Kelso, Saltzman, & Tuller (1986), see this from another perspective, as a "'compression' ... from a microscopic basis of huge dimensionality to a low-dimensional macroscopic description" (p. 32).

In Kelso et al.'s (1986) view, each step upwards in description represents a set of constraints on the degrees of freedom that lower levels of articulatory description make available. At the lowest level of physiological description, there is a large number of degrees of freedom implicit in the varying levels of neuromuscular activity in various articulatory organs. However, neuromuscular units do not act independently - they act in concert, producing different families of articulator position (or state) reified as different levels of expression of articulatory primes. The family of raised larynx positions is the

overall result of a large number of microscopic physiological events. In particular languages, the potential degrees of freedom implicit in the inventory of articulatory primes may not be realized due to the functional dependence of a helping prime on a core prime that a coordinative structure produces. A constricted pharynx in an Akan [-ATR] vowel is the result of larynx raising, tongue root retraction, and some movement of the dorsal wall of the pharynx as well. Language-specific coordinative structures realize phonological goals: [-ATR] vowels in Akan include larynx raising; [-ATR] vowels in DhoLuo do not.

In addition, it seems at least plausible that dynamic properties of speech production should be stated at this level, in order to ensure that all the component articulatory primes of a coordinative structure act synchronously (see, e.g. Kelso, Tuller, & Harris 1983; Saltzman & Kelso 1987). However, these properties are not an issue in this study.

If it were clear that speech made independent use of each possible physiological degree of freedom, i.e., if it were the case that phonological features mapped directly to muscular activities, then we would not need to distinguish all these levels of description of speech production. All we would need to do would be to list the independent muscle fiber groups of the vocal tract. Their states (degree of contraction, speed of contraction, etc.) serve as the dimensions of both the linguistic and the articulatory specifications of articulatory configurations.

Though this position is certainly a logical possibility (see Halle (1983) for a position close to the one just sketched), it seems unlikely to be the correct one. Cross-linguistic evidence like that discussed above suggests that languages have variable patterns of recruitment of helping articulatory primes together with some core articulatory primes common to all language-specific implementations of phonological specifications. Thus, phonological specifications stated directly on articulatory states would have much within-language redundancy and obscure cross-linguistic similarities.

However, the existence and language-specificity of coordinative structures mean that measures of vocal tract shapes in any one language cannot distinguish intrinsically correlated variation (due to the fact that several measures may reflect the activity of a single articulatory prime) from functionally correlated variation that is peculiar to the phonology or the language-specific phonetics of that language. *Observations in single languages therefore cannot reveal the structure of articulatory primes.* Thus, cross-linguistic studies are essential for identifying articulatory primes and resolving the indeterminacies inherent in the phonological underdetermination of vowel articulation.

Because coordinative structures link articulatory primes, the contributions of some pairs (or larger sets) of articulatory primes to vowel articulation in any particular

language may be correlated. If two articulatory primes are members of the same coordinative structure, then the contribution of one of them to vowel articulation should be expresible as a function of the contribution of the other, and thus correlated with it. Since the contributions of these primes will be correlated, they will not in general be uniquely identifiable. The situation is entirely analogous to multiple regression with correlated independent variables - the multicollinearity problem. Just as multiple regression methods cannot uniquely partition the sum of squares when the regressors are correlated, the articulatory primes and their contributions to specific articulatory states cannot in general be identified in data from any one language. Rather, the articulatory correlates of language-specific coordinative structures are identifiable.

We can now rationalize several aspects of the studies discussed above. Harshman et al. (1977), Lindau (1986) and Jackson (1988) report success (R^2 of over 0.9) decomposing tongue shape data from English, Akan, and Icelandic into tongue shape factors using the PARAFAC model. The first tongue shape factor ('Front Raising') from each of the three languages are fairly similar, but subsequent factors are not. Similarly, Linker (1982) uses PARAFAC to decompose lip position data from English, French, Swedish, and Cantonese into lip position factors. Again, there is one factor that seems to occur in each of the languages (roughly related to the overall area of the lip opening), but subsequent factors are much more variable.

The conceptual model proposed above provides an explanation for both the within-language success of these models, and their weak cross-linguistic replicability. By hypothesis, coordinative structures vary from language to language, and so the PARAFAC factors found in each language vary. However, the within-language factors should have invariant 'hidden' structure, corresponding to articulatory primes, within them. The factors that replicate well (e.g., the Front-Raising like factors in English, Akan, and the other studies cited in Chapter 1) across languages are presumably the articulatory correlates of core articulatory primes or a coordinative structure that is shared by all the languages investigated.

The PARAFAC model and coordinative structure

I argued above that coordinative structures organize articulatory primes in a language-specific manner during the production of vowels. Undiscussed was the question of what kind of task the coordinative structures facilitate the performance of. Are the goals of the speech production system articulatory, acoustic, perceptual, or some kind of perceptuo-articulatory gestalt? Whatever the *ultimate* nature of these goals, it seems clear that each particular situation (word) eventually produces simple articulatory goals, since token-to-token articulatory variability within single speakers is generally reported to

be small in magnitude (e.g., Nearey 1978, p 42). This implies that speakers do not freely use possible articulatory tradeoffs that generate equivalent acoustic outputs, and therefore that the immediate goal in any particular situation is not acoustic.

This does not preclude the possibility that different speakers might in fact systematically differ with respect to their utilization of the possible tradeoffs. For instance, some speakers might consistently round and protrude their lips greatly, with little larynx lowering; others might use less lip protrusion and proportionately greater larynx lowering. Such (near) proportional differences across speakers bring us to the question of what kind of model might represent this inter-speaker variation and yet still represent the underlying unity of coordinated gestures.

The PARAFAC model (Harshman 1970, 1976) appears to represent this aspect of articulatory data well (see the Appendix to this chapter for motivation and discussion of PARAFAC). In the current context, the factors that result from a PARAFAC analysis of data from a single language may be characterized by:

- characteristic modes of articulator displacement associated with the factor, represented by 'articulator loadings'. The articulator loadings represent normalized speaker-independent articulatory displacements.
- the contribution of each mode of articulator displacement to each vowel (the vowel's loading on each factor, in traditional factor-analytic terminology). This is a speaker-independent map of the vowel space.
- a speaker-specific scale for each factor showing the magnitude of that factor's contribution to the measurements from that speaker. This scale reflects both overall differences in size and differences in relative proportion (e.g., wide versus narrow pharynges) between speakers.

These speaker-specific scales may be thought of as inducing a mapping from the speaker-independent vowel space and characteristic articulator displacement modes to an approximation of the articulatory displacements actually observed in in that speaker. The speaker-independent articulator displacement modes, when weighted by their contributions to specific vowels and their scales for individual speakers, generate the space of observed articulatory configurations.

Algebraically, the PARAFAC model as applied to vocal tract position measurements is formulated as follows. Let y_{ijk} be the ith measurement of articulator position in the jth vowel as produced by the kth speaker. Let a_{ir} be the change in the ith measure produced by the rth mode of articulator displacement (factor). Let v_{jr} be the contribution of the rth displacement mode to the jth vowel. And finally, let s_{kr} be the kth speaker's scaling

constant for the rth factor. Form the matrices $\mathbf{A} = (a_{ir})$, and $\mathbf{V} = (v_{jr})$. Form a matrix of the raw data $\mathbf{Y}_k$ for each speaker, (y_{ijk}) (k constant). Let $\mathbf{S}_k$ be matrices with the kth column of (s_{kr}) on the main diagonal and zero elsewhere. Each of these matrices thus has one speaker's scaling constants on the diagonal, and is zero elsewhere. Then the PARAFAC model, as applied to vocal tract measurements, may be written as in (2.1). This model and its applicability is discussed in greater detail in the Appendix to this chapter.

$$\mathbf{Y}_k = \mathbf{A}\mathbf{S}_k\mathbf{V}^T \tag{2.1}$$

Background to the proposed model

What exactly does PARAFAC model? How are PARAFAC models related to the other models that have been prposed in the literature? This discussion will draw on notions from functional analysis, the mathematical study of abstract spaces and mappings; Strang (1976) and Kreysig (1978) contain most of the background to this section.

We have seen that some notion of a *vowel space* underlies most of the models discussed above. This is a space containing the goals associated with prototypical phonemic vowel qualities - whether these goals are perceptual, articulatory, or neutral between perception and articulation is generally not explicitly stated. Lindblom, Paul, & Sundberg (1974), for instance, are concerned with modeling the ways in which articulatory goals are realized in various phonetic contexts. This space is a continuous space, even if it sometimes appears as though there is a finite and discrete set of goals in each particular language. It is often assumed to be parameterized in some fashion related to the usual phonological classification of vowels, as with Mermelstein's (1973) definition of tongue body height and backness as a model control parameter. However, Lindblom et al. (1974), Liljencrants (1971,1985), and Wood (1982) adopt parameterizations not so closely related to the traditional phonological classification.

There is also a notion of an *articulatory space*, i.e., the space of possible postures of the vocal tract. This space is also continuous, since the organs of the vocal tract are not given to discontinuous movement or disconnecting themselves.

The articulatory space can be characterized as a kind of function space. We can define a set of (vector-valued) functions $f_{\text{tongue}}, f_{\text{jaw}}$, etc., that represent the positions of the various organs of the vocal tract (with reference to the mean position) at a particular instant in time. For instance, we may parametrize the anterior surface of the pharynx and

the upper surface of the tongue in terms of the normalized distance from the glottis, so that the tongue's entire surface is represented as a (vector-valued) function on the interval from zero (the glottal end at the bottom of the pharynx) to one (at the tip of the tongue). This function may then be approximated in various ways - it has become customary in the acoustical modeling world to use a staircase function to approximate the tongue contour (or the vocal tract radius). Mermelstein's model uses a piecewise approximation; Liljencrants' model uses a Fourier-series approximation. Other more appropriate approximations could be considered - in particular, approximation by using the Chebyshev polynomial series (cf. Jospa 1974). Unlike the Fourier cosine series that Liljencrants (1971) used, which concentrates the error at the edges of the modeled interval, Chebyshev approximation has the attractive property of distributing the modeling error uniformly over the modeled interval. (See Ramsey (1982) for further comments on techniques for approximating continuous functional data in numerically tractable ways). Whatever the approximation we use, we can consider its parameters to be a kind of measure of vocal tract shape.

Once an approximation to a particular articulatory posture is constructed, the parameters of the construction (the position of the center of the tongue body circle in Mermelstein's model, the coefficients of the Fourier series in Liljencrants' model, or the coefficients of the Chebyshev polynomials), may be concatenated into a vector. The articulatory space may be represented as a vector space, each point of which is characterized by as many coordinates as there are parameters of the model. Since there are 'ceilings' in articulatory data (e.g., the tongue cannot penetrate the hard palate), these approximations are *not* valid for general articulatory data. However, we may assume that for vowel production, where there is in general less contact between articulators, these non-linear effects are minimal. At most, they introduce a boundary beyond which the linear approximation is no longer valid.

It is also important to note that these models of articulatory postures contain many possibilities which could not be realized by a real vocal tract. For instance, within the space of functions described by Liljencrants' truncated Fourier series, it is quite possible to set the coefficients of all but the highest-frequency component to zero. This would give a shape which is almost certainly not attainable by a real vocal tract, but is nonetheless generable by the articulatory approximation. Similarly, Harshman et al. (1977) point out that the vocal tract in Fig. 2.1 ought not to be allowed by a realistic model, even though it is a formal possibility allowed by approximating vocal tract shapes by measurements of tongue position along gridlines. Properly constraining the space generable by the articulatory model is an issue to be addressed below.

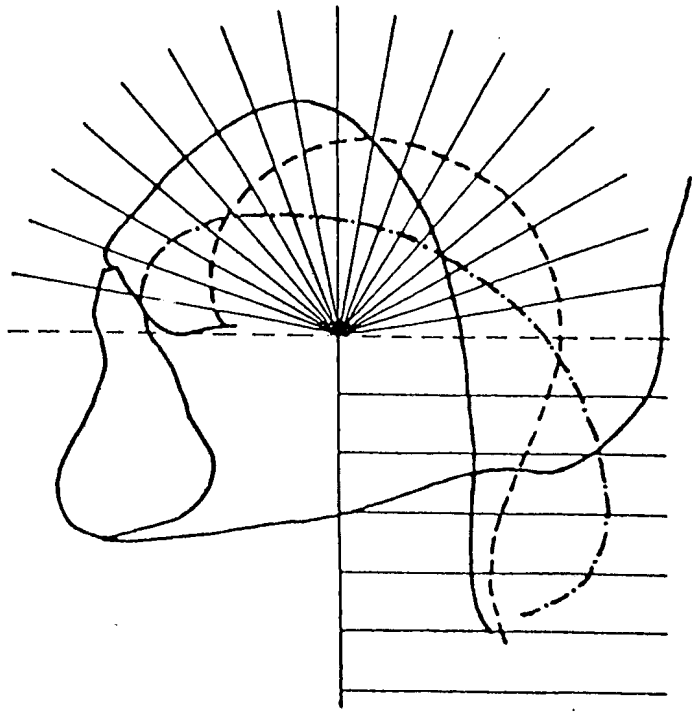


Figure 2.2. Tongue shapes for the Swedish vowels [**i**, **u**, **a**].

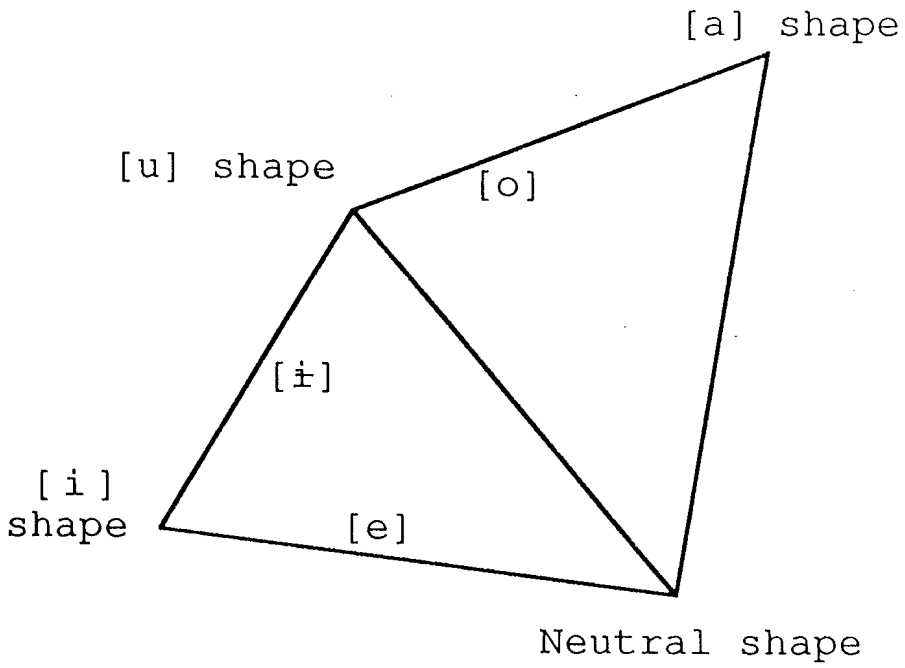


Figure 2.3. Modeling the Swedish vowel space in terms of [**i**, **u**, **a**] and a neutral vowel by interpolating between given configurations.

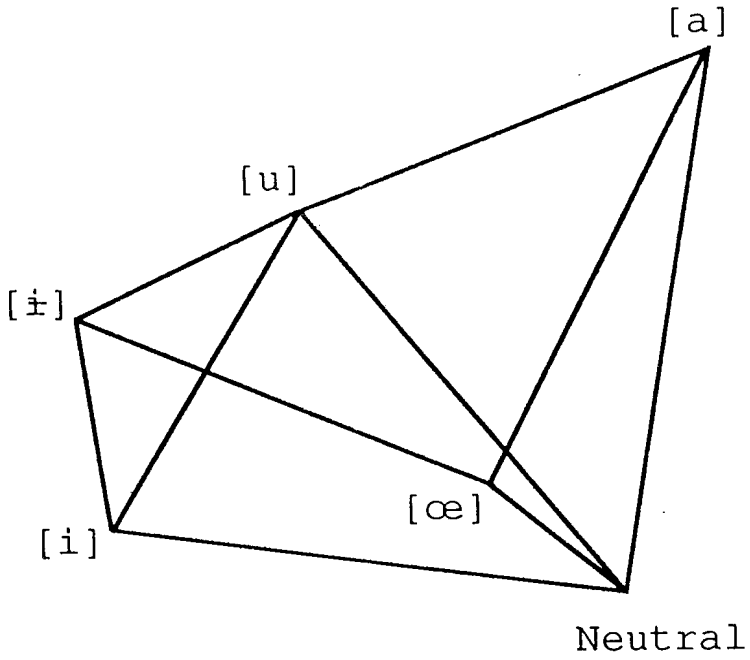


Figure 2.4. An extended model of the vowel space.

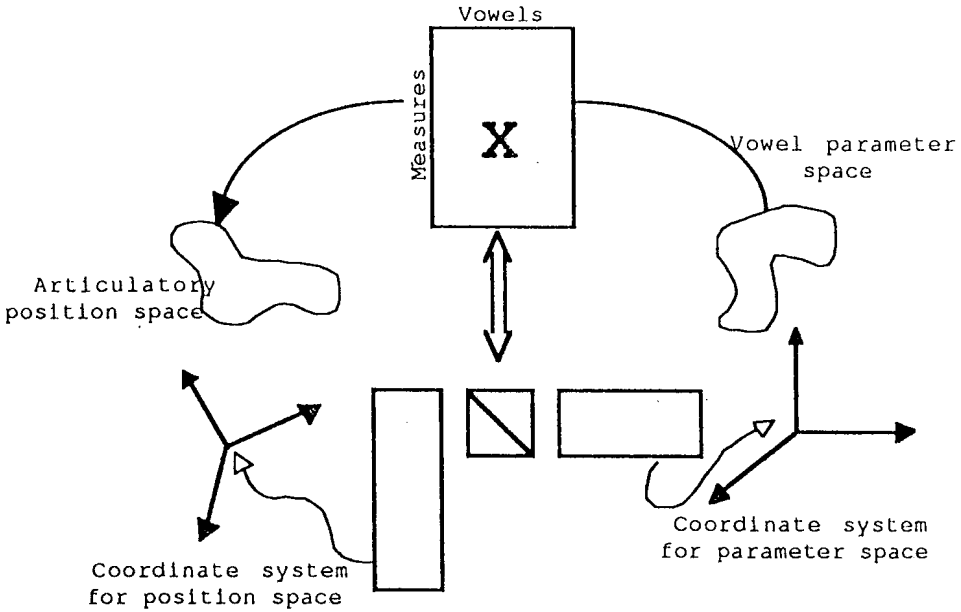


Figure 2.5. SVD of a Lindblom & Sundberg-like model.

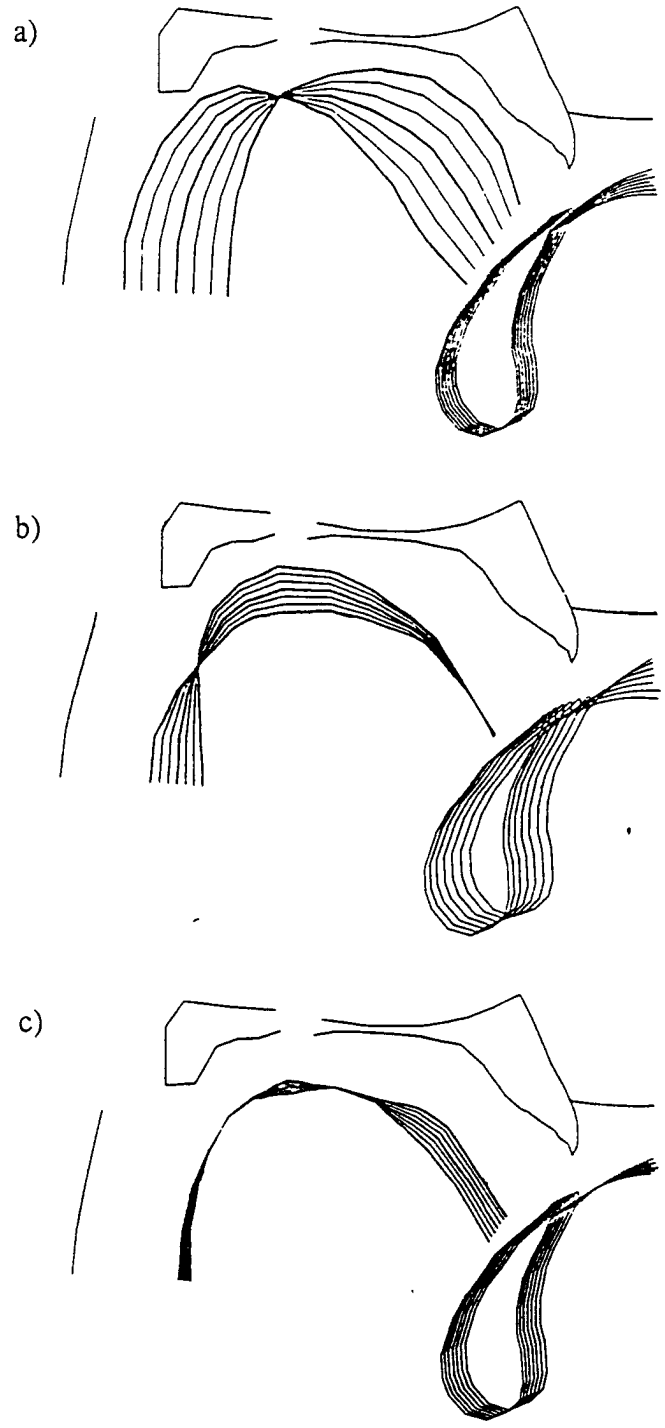


Figure 2.6 Principal factors underlying tongue positions for one speaker's Icelandic monophthongs (after Jackson 1986).

A model of vowel articulation can be considered an operator that maps parameter sets from the vowel space into the articulatory configuration space. For instance, Lindblom and his co-workers model the space of possible tongue shapes during vowels by interpolating between four given tongue shapes - a neutral shape, and the ones observed for the vowels [**i**], [**u**], and [**a**] as produced by a speaker of Swedish (Fig. 2.2). This scheme is motivated by Lindblom & Sundberg's observation that tongue shapes during the articulatory extremes of vowels seem to fall into a few distinct families, exemplified by the given vowels. Other vowels are naturally represented by less extreme members of these families, as in Fig. 2.3. This representation treats the vowel space as a set of planar patches, with the given vowel shapes at its apexes, and other vowels in between.

One can, however, legitimately ask whether the particular vowels used exhaust the families of tongue or vocal tract shapes. It would be simple enough to extend the model by adding more and more vowels to the sample, thus enclosing a vowel volume in some complicated polyhedron (Fig. 2.4). This model would be implemented by a larger matrix **X** with columns constructed from the various vowels in the sample, and larger parameter vectors **p**, with each parameter specifying some vowel's contribution to the interpolated shape. But then we would want to know what constituted a sufficient set of landmarks in the vowel space, what the articulatory relationships between the various vowels were, and how well the model generalized across speakers.

The singular value decomposition (SVD) can be used to answer some of these questions. The SVD of a matrix that represents a linear operator consists of three matrices: a matrix that provides a basis for the domain of the operator (the vowel parameter space); a matrix that provides a basis for range (the articulatory space); and a diagonal matrix that has weights proportional to the contribution of the corresponding element of the basis to the data sample. The decomposition of a linear operator such as the one described above is schematized in Fig. 2.5. (For further discussion of the SVD, see Press, Flannery, Teukolsky, & Vetterling 1986 and references contained therein). The SVD of a variance-covariance matrix (rather than a raw data matrix) produces a principal component analysis; minor variations lead to principal factor analysis, image analysis, etc. Jackson (1986) carried out a principal factor analysis on tongue positions in one speaker's Icelandic monophthongs (from Pétursson 1974b). The three factors (Front Raising, Tongue Arching, and Blade Raising) derived from the data are illustrated in Fig. 2.6 as estimated displacements, and two dimensions of the resulting three-dimensional vowel space is illustrated in Fig. 2.7.

The PARAFAC procedure (Harshman & Lundy 1984ab) is a multispeaker generalization of the SVD (Carroll & Chang 1970, Kruskal 1984). Fig. 2.8 shows the

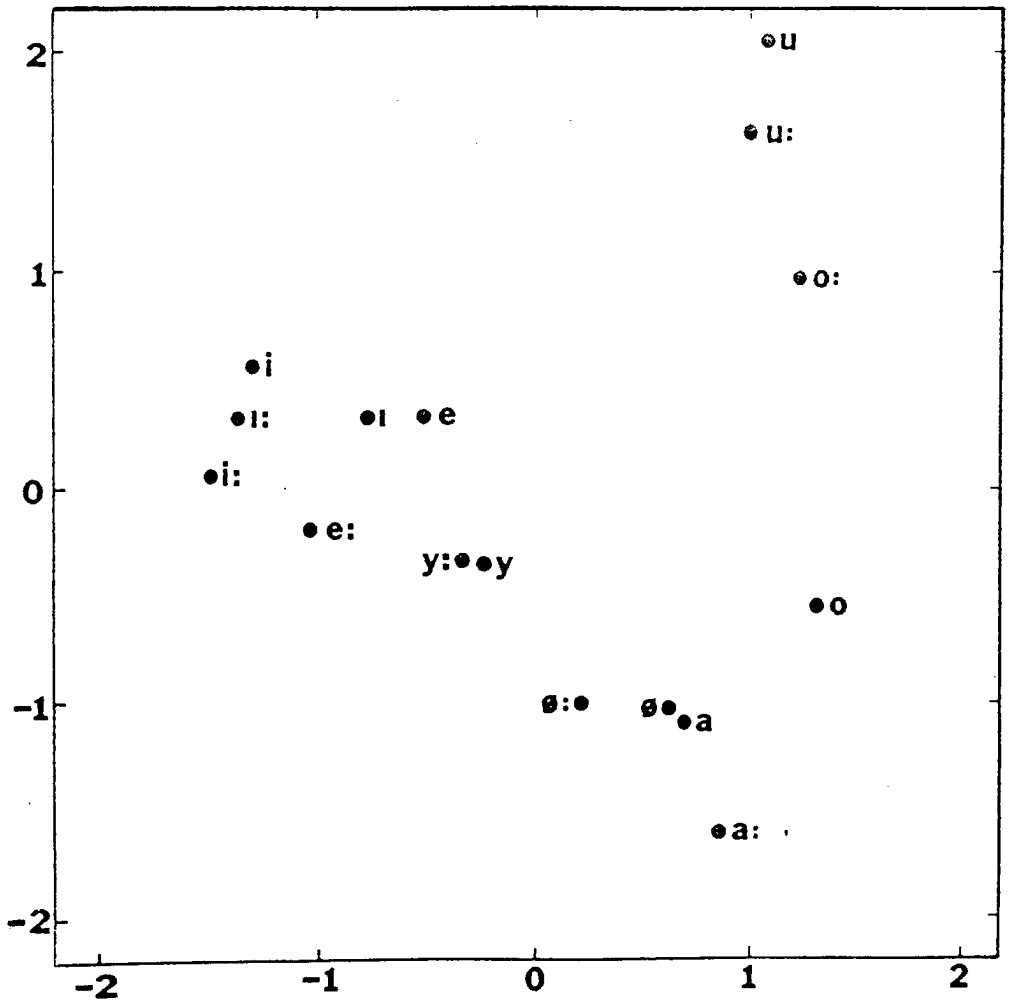


Figure 2.7. Icelandic monophthongs in the first two dimensions of the Icelandic vowel space (after Jackson 1986).

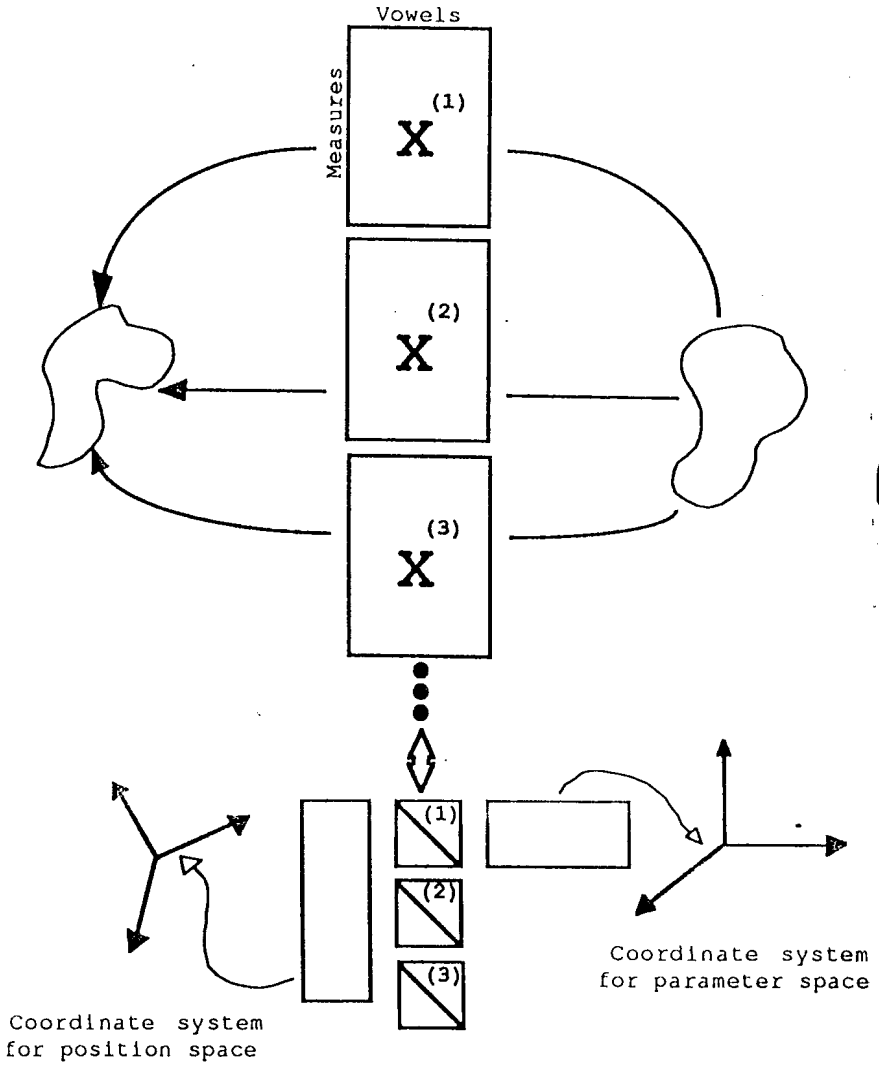


Figure 2.8. Design of a typical PARAFAC study.

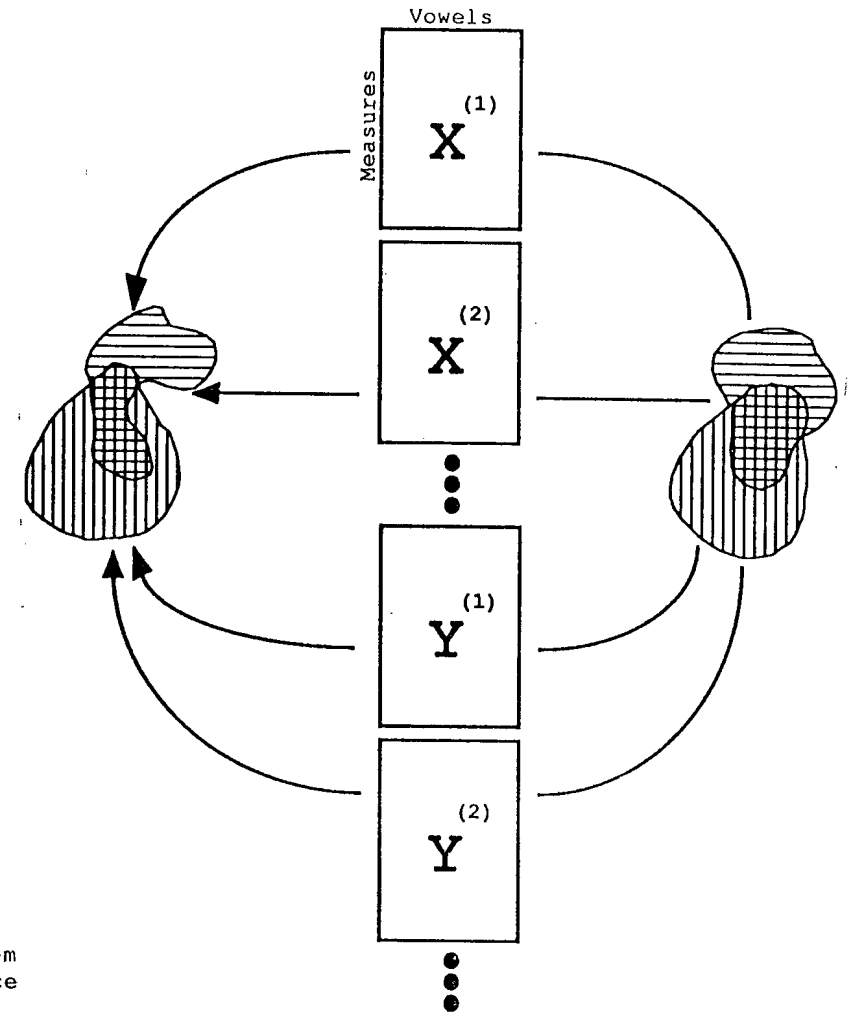


Figure 2.9. Design of a cross-language study.

design of a typical PARAFAC study. Given several speaker's mappings from a specific set of vowels to a set of articulatory postures, PARAFAC produces three kinds of matrices. First, it produces a basis or coordinate system for the domain of these mappings, a vowel parameter space (**V** in 2.1). Second, it produces a basis for the range space of articulatory postures used by the speakers (**A** in 2.1). Third, PARAFAC produces a set of diagonal matrices, one per subject, that contain weights indicating the contribution of the corresponding element of the basis to the data set ($\mathbf{S}_k$). Since the other aspects of the data are the same for all speakers (i.e., they all produce the same vowels, and the same measurements are made on all of them), these weights may be interpreted as reflecting variation in speaker size and proportion. The bases for the vowel space (**V**) and articulatory configuration space (**A**) are thus normalized across speakers; the speaker-dependent terms are isolated in the $\mathbf{S}_k$ matrices.

These columns of these matrices are known as factors, since their product estimates the mapping. The number of columns in the **V** matrix is spoken of as the *dimensionality* of the vowel space. The range basis and subject weight terms can be lumped to form an articulatory model. Just as in Lindblom & Sundberg's model, the product of this matrix and a vector of vowel parameters yields an articulatory posture.

There is no *mathematical* necessity that the subject-weight matrices be diagonal, and supposing that they are constitutes a strong hypothesis about the nature of the data. There are statistical models which allow more general mappings between the domain and range spaces of operators (Tucker (1966) and Kroonenberg & de Leeuw (1980) discuss such models), but they do not appear to be necessary for this kind of data. Harshman et al. (1977) report that the two-factor PARAFAC solution for their English tongue position data accounted for over 96% of the variance in the raw data.

A model for cross-language articulatory investigation

PARAFAC, however, is not sufficient for cross-linguistic data. The problem, as is schematized in Fig. 2.9, is that there is no guarantee that we can identify the vowels of any particular language with the vowels of any other particular language. The vowels may come from different regions of the parameter space, because the languages have different phonological systems. Similarly, speakers of the different languages may exploit different articulatory possibilities, due to the language-specificity of coordinative structures. In the terms of Harshman & Lundy (1984a), a cross-language sample of articulatory data may show not only *system variation*, but also *object variation*. In the current context, the system variation postulate is that all speakers have the same articulatory organization, i.e., the same coordinative structures; and that the expression of these coordinative structures varies in a parallel manner across vowels for each speaker.

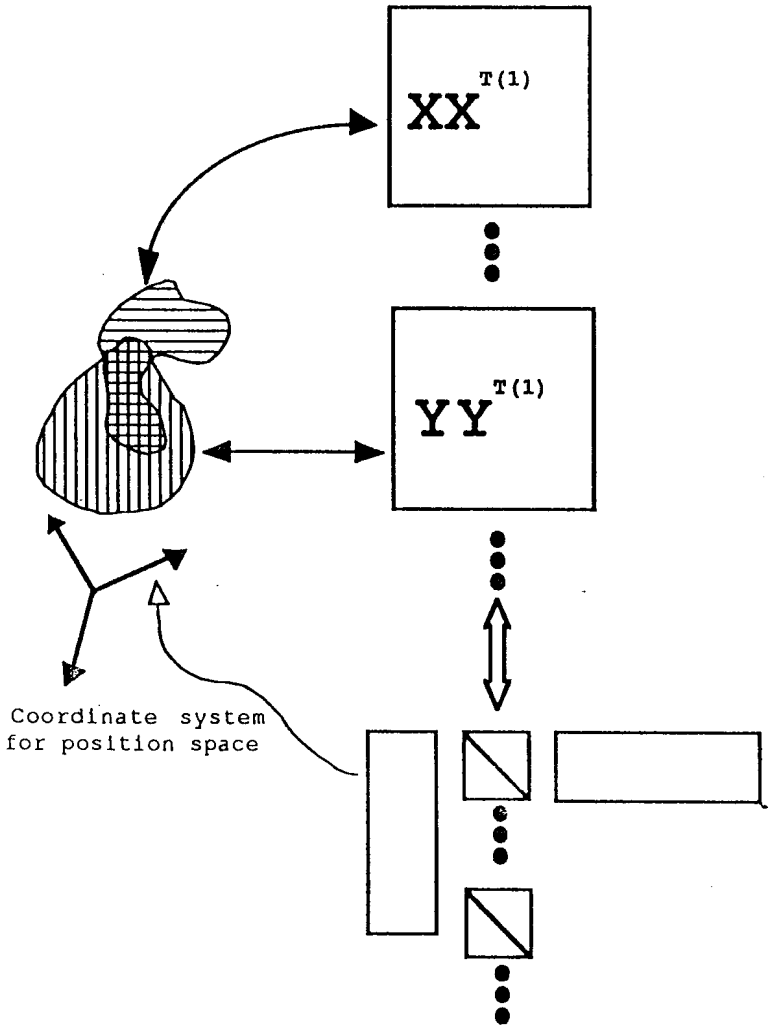


Figure 2.10. Modeling covariance matrices via PARAFAC.

Object variation implies that the coordinative structures do not vary in a parallel manner across objects (vowels) for all speakers. Thus, the vowels under investigation may differ from speaker to speaker.

One way around this problem is to model the covariances between measurements of articulatory positions rather than the raw data. As Fig. 2.10 suggests, each subject's variance-covariance matrix is related to the product of the raw data matrix with its transpose. Formally, this matrix can be characterized as an operator mapping the space of possible articulatory measurements into itself. We lose some information, in that the variance-covariance matrix no longer contains information about the parameter space, but on the other hand, we no longer have to cope with possible language-specific variation in that space. Language-specific constraints have no explicit expression in the model, except as they are reflected in individual differences.

Fig. 2.10 also shows how we can get a simultaneous decomposition of these mappings from PARAFAC. In this case, PARAFAC yields a sort of simultaneous modal analysis of the data. We get a basis set of articulatory modes analogous to the eigenvectors of a single covariance matrix; we also get a set of weights analogous to eigenvalues for each subject. Again, we should only retain modes or factors that contribute non-vanishing variance to the modeled data. An algebraic formulation of the model implicit in this treatment of the data is given in (2.2); the notation $\mathbf{Y}_{k(l)}$ indicates the data matrix formed from measurements on the *k*th speaker of the *l*th language's vowel productions. Thus, (2.2) models the indicated variance-covariance matrix for each speaker ($\mathbf{Y}_{k(l)}\mathbf{Y}_{k(l)}^{T}/N_{vowel(l)}$) in terms of a general matrix containing the articulatory primes ($\mathbf{A}$) and speaker-specific scaling coefficients ($\mathbf{S}_{k(l)}$).

$$\mathbf{Y}_{k(l)}\mathbf{Y}_{k(l)}^{T}/N_{vowel(l)} = \mathbf{A}\,\mathbf{S}_{k(l)}^{2}\,\mathbf{A}^{T} \tag{2.2}$$

Harshman & Lundy (1984a) show that this approach will yield the same factors that a PARAFAC analysis of raw data does *if* the vowels in the sample of languages are distributed orthogonally (evenly) in the vowel space. However, one could well argue that the assumption that vowels are distributed evenly in the vowel space is unwarranted. Jackson (1988) performed a PARAFAC analysis measurements of tongue position from Icelandic monophthongs, then factored covariance matrices calculated from those monophthongs together with covariance matrices derived from Harshman et al.'s (1977) measurements of tongue position in American English vowels. The tongue position factors resulting from this procedure were indeed somewhat different from the tongue

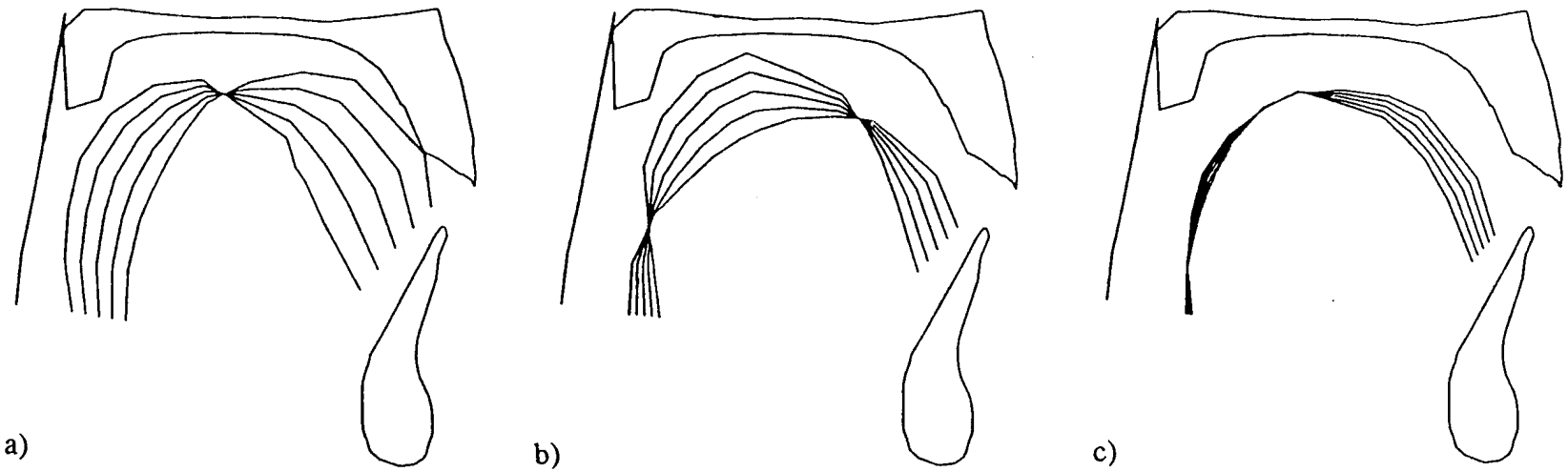


Figure 2.11. Tongue position factors in Jackson's (1988) PARAFAC model of Icelandic.

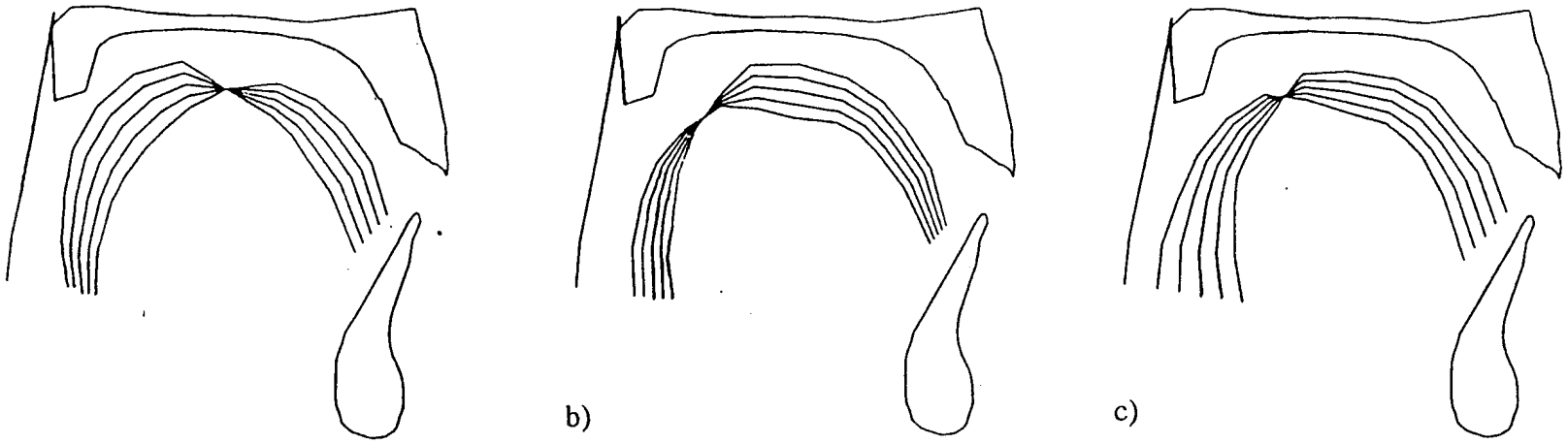


Figure 2.12. Tongue position factors in Jackson's (1988) joint English-Icelandic covariance analysis.

position factors arrived at by applying PARAFAC to the raw data within each language. Figs. 2.11 and 2.12 show the tongue position factors resulting from analysis of the raw Icelandic data and the covariance analysis of both Icelandic and English data respectively. Comparing Fig. 2.12 with Fig. 2.11 and Fig. 1.13, we can see that there are some differences in the profiles of the articulatory expression of the factors.

Because (2.2) does not have any term corresponding to the **V** matrix (the coordinates of the vowels in the vowel space) in (2.1), it also does not represent the vowels produced by the *k* speakers of the *l*th language as all having the same coordinates in the vowel space. The lack of such a constraint means that a model of the form in (2.2) may detect individual differences along with systematic behaviours. However, I will pursue this approach in Chapter 4, to see how well it does.

Another way of approaching this data is to use what has been called a 'joint-mode' PARAFAC model (Harshman & Lundy 1984b). Recall that we have postulated an underlying layer of articulatory primes common to all speakers (simply because they normally share the same gross anatomical structures), which may receive language-specific organization into coordinative structures. We are thus postulating an underlying joint, or common, set of articulatory primitives which may receive speaker-dependent scaling and language-specific organization for the purposes of vowel production. The model can be stated algebraically as in (2.3).

$$\mathbf{Y}_{k(l)} = \mathbf{A}\,\mathbf{S}_{k(l)}\,\mathbf{V}_l^T \tag{2.3}$$

In (2.3), $\mathbf{Y}_{k(l)}$ is the matrix of articulatory measurements (or articulatory approximation parameters) from the *k*th speaker of the *l*th language. The articulatory measures for the *j*th vowel form the *j*th column of $\mathbf{Y}_{k(l)}$, so that $\mathbf{Y}_{k(l)}$ has the dimensions $N_{measures}$ rows by $N_{vowels(l)}$ columns. The matrix **A** contains the effect of the *r* factors, or dimensions, on the articulatory approximation; $\mathbf{V}_l$ contain the loadings of the *l*th language's vowels on *r* factors. As in (2.1), $\mathbf{S}_{k(l)}$ is a diagonal matrix containing scale factors that account for variations in size and proportion specific to the *k*th speaker of the *l*th language. The model can be considered a kind of nested ANOVA design, except for the fact that we do not have a fixed mode (the classification, or design, variables in an ANOVA are fixed, and the parameters of the ANOVA model are estimated relative to the fixed variables).

This model assumes that the articulatory space represented by **A** is the same for all speakers. Similarly, the sample of vowels in language *l* is assumed to be drawn from a space $\mathbf{V}_l$. All of the *k* speakers of the *l*th language produce vowels from the space $\mathbf{V}_l$.

However, the vowels of language l and the vowels of language m are allowed to be different, i.e., in general $\mathbf{V}_l \neq \mathbf{V}_m$.

It is worth considering for a while the nature of the vowel space represented by the $\mathbf{V}_l$ matrices above. Is it possible, as some of the investigators in Chapter 1 have suggested it might be, to relate the elements of **V** to physical quantities like muscular tensions or neural commands? The answer, unfortunately, is no. The relationship of muscular contraction to neural stimulation is known to be complex and non-linear. The muscular state corresponding to a given level neural input is a non-linear function of the prior degree of muscular contraction, the load on the muscle, and the rate of contraction. The muscular response is also non-stationary on a time scale of fractions of a second, due to fatigue. It is also non-homogenous, in that there are several distinct populations of muscle fibers (slow vs. fast twitch) with distinct response functions.

Muscular tensions are not directly relatable to this model for a different kind of reason. The problem is that any given vocal tract position is compatible with a range of different muscular tensions. The fact that musculature tends to be organized in terms of agonist-antagonist pairs means that the tensions in any particular agonist-antagonist pair could be quite high without a difference in posture, as long as the applied forces cancel. Direct measurements lacking, we do not know if this situation occurs in natural speech, but we cannot rule it out. The same arguments apply to EMG activities.

So the vowel space of (2.3) is not a space of physiological commands along the lines of Halle (1983). It is rather an abstract space of goals, which may receive variable physiological implementation, depending on the prior state of the vocal tract. It represents an invariant linguistic-phonetic abstraction away from variable physiological implementation.

The model in (2.3) can be written in an explicit multisubject form as in (2.4). We begin by constructing supermatrices **Y*** and **S*** by putting the $\mathbf{Y}_{k(l)}$ and $\mathbf{S}_{k(l)}$ side-by-side. Finally, we construct **V*** by putting one copy of $\mathbf{V}_l$ for each speaker of language l in it corner-to-corner and setting the rest of it to zero. We can then write **Y*** in terms of the other matrices as in (2.4). (2.5) represents (2.4) in terms of the component matrices.

$$\mathbf{Y^*} = \mathbf{A}\,\mathbf{S^*}\,\mathbf{V^{*T}} \tag{2.4}$$

$$(\mathbf{Y}_{1(1)}\,\mathbf{Y}_{2(1)}\ldots\mathbf{Y}_{k(l)}) = \mathbf{A}\,(\mathbf{S}_{1(1)}\,\mathbf{S}_{2(1)}\ldots\mathbf{S}_{k(l)}) \begin{pmatrix} \mathbf{V}_1^T & & \phi \\ & \mathbf{V}_1^T & \\ & & \ddots \\ \phi & & \mathbf{V}_l^T \end{pmatrix} \tag{2.5}$$

The formulations in (2.4) and (2.5) allow the estimation of $\mathbf{A}$, $\mathbf{S}_{k(l)}$, and $\mathbf{V}_l$ to be treated as a problem in covariance-structure analysis. For instance, the matrix $\mathbf{Y^{*T}Y^*}/N_{measures}$ - the speaker-vowel-language covariance matrix - can be estimated from the right side of (2.4). This kind of estimation can be performed by the EQS program (Bentler 1985).

Each column of $\mathbf{Y^*}$ contains the articulatory measures characterizing a particular vowel produced by a particular speaker. The corresponding column of $\mathbf{Y^{*T}Y^*}$ contains 'similarity indexes' that relate the positions of the articulators in that particular speaker-vowel combination to all the other speaker-vowel combinations in the sample. Vowels with similar articulatory configurations produced by different speakers will have high values of this index; dissimilar vowels will have lower values. With EQS, we can derive factors (or latent variables) underlying these patterns of similarity and dissimilarity. We can also explicitly test hypotheses about the degree to which a particular factor is used in a particular language.

We can rewrite the matrix product $\mathbf{Y^{*T}Y^*}$ as in (2.6). In (2.6), the product $\mathbf{A^TA}$ has been replaced with the factor correlation matrix Φ. By an extension of the method demonstrated in Bentler, Poon & Lee (1987), EQS can estimate the $\mathbf{V}_l$, $\mathbf{S}_{k(l)}$, and Φ terms of (2.6). The fact that Φ need not be diagonal allows us to relax the (equivalent of the) orthogonality constraint imposed in (2.2), the PARAFAC treatment of covariance matrices. The model in (2.6), when fit with EQS, might therefore produce better fits than the model in (2.2).

$$\mathbf{Y^{*T}Y^*} = \mathbf{V^*}\,\mathbf{S^{*T}}\,\Phi\,\mathbf{S^*}\,\mathbf{V^{*T}} \tag{2.6}$$

Unfortunately, it is necessary to fit the data indirectly via covariance matrices, even though it is possible to write expressions such as (2.4) that directly fit the conceptual

model developed above. The reason is pragmatic: procedures for fitting models like (2.4) are not generally available. Harshman (personal communication) has indicated to me that the joint-mode model envisaged in Harshman & Lundy (1984b) is only now in progress. Devising an algorithm to fit models like those in (2.4) is a topic for further research. Meanwhile, the languages and measurements for fitting to the expressions in (2.2) and (2.6) are the subject of the next chapter.

Interpreting the model

Estimating the **A** matrix according to (2.2) provides us with a model for the articulatory configurations in the sample **Y***. By hypothesis, the components (column vectors) of **A** represent articulatory primes, i.e., articulatory gestures that are at least potentially independently controllable although they may always be recruited into one or another coordinative structure during normal speech. (If the data are not adequate, i.e., if all the languages that happen to be in the sample are similar enough, the components of **A** will represent coordinative structures). These articulatory correlates of these gestures are represented in a speaker- and vowel- independent manner, due to the removal of speaker size coefficients in the $\mathbf{S}_{k(l)}$ matrices. Another degree of anatomical normalization is due to the measurement procedures discussed in Chapter 3. Observed positions of the vocal tract are modeled as combinations of these normalized articulatory configurations.

Estimating the $\mathbf{V}_l$ matrices via (2.6) provides us with a model for the minimal space required to contain all the vocal tract configurations observed in the sample **Y***. The *j*th column of $\mathbf{V}_l$ contains the coordinates of the *j*th vowel in this space; these coordinates are a normalized measure of the contribution of the various articulatory primes to that vowel. We may hypothesize that this is (at least isomorphic to) a *universal* phonetic representation of the articulatory goals associated with vowel production in the languages of the sample.

As remarked above, it may be the case that in a particular language a given helping articulatory prime will be 'recruited' into a coordinative structure and therefore be functionally linked with a core articulatory prime. In this case, the contributions of the two primes to each vowel should covary. In other words, there should be a high correlation between the contribution of the helping prime and the contribution of the core prime to the various vowels of the *l*th language, even if their correlation is low in the *m*th language. These correlations are the correlations between the various *rows* of $\mathbf{V}_l$.

It may also be the case that a particular articulatory prime is not systematically exploited in a particular language. Thus, the vowels of that language will lie in a

well-defined subspace of the universal vowel space. In this case, the row of $\mathbf{V}_l$ corresponding to that articulatory prime should contain only zeroes, indicating that that particular prime does not contribute to any of the vowels represented by the columns of $\mathbf{V}_l$.

Finally, we may observe that the matrix $\mathbf{V}_l$, minus the rows corresponding to helping and unexploited primes, constitutes a systematic phonetic representation of the vowels of language *l*. This systematic phonetic representation, which we may take to be stated at the level of coordinative structures, contains the non-redundant phonetic information needed to specify the articulation of the vowels of language *l*. As such, it may be regarded as an intermediate level of systematic phonetic representation. It excludes language-specific and redundant facts about helping and unexploited articulatory gestures, and includes only the correlates of the distinctive (non-redundant across the inventory), featural specification of the vowels of the language.

From the top-down perspective, this representation in terms of coordinative structures is fleshed out by a language-specific mapping to articulatory primes during the process of translating a phonological specification to articulatory reality. Each stage of this translation is accompanied by an increase in the amount information specified: from linguistic phonological abstraction to language-specific coordinative structures to universal speaker-independent articulatory primes to speaker-specific articulatory expression.

Appendix to Chapter 2: Review of the PARAFAC model

The purpose of this section is to motivate the PARAFAC model for those familiar with more generally employed modeling procedures. This discussion contains little original, and is derived largely from Carroll & Pruzansky (1984) Harshman & Lundy (1984a), Kruskal (1984), and Ramsay (1982). Standard texts on linear algebra, e.g. Strang (1976) and functional analysis, e.g. Kreysig (1978) also contain relevant material. Defects in the expostion are nonetheless the responsibility of the author.

We will conclude with a discussion of why PARAFAC is suited to the analysis of measurements of articulator positions.

Historical background

PARAFAC and CANDECOMP are mathematically equivalent formulations of a model based on the principle of parallel proportional profiles (Cattell 1947, Cattell & Cattell 1955). The two models were discovered nearly simultaneously by independent workers (Harshman 1970, Carroll & Chang 1970). Here, I will concentrate on the PARAFAC formulation, which lies in the tradition of psychometric factor analysis.

The principle of parallel proportional profiles has been suggested by Cattell as a basis for truly explanatory factor models. Cattell & Cattell (1955, p. 84) put the principle as follows:

> The basic assumption is that, *if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study.* No inorganic factor, a mere mathematical abstraction, would behave in this way ... We could then anticipate finding the "true" factors by locating the unique rotational position (simultaneously in both studies) in which each factor in the first study is found to have loadings which are proportional to (or some simple function of) those in the second ... [italics in the original]

Gorsuch (1983) observes "the argument for [proportional profiles] is that it is a more direct approach to the invariance ... invariance across samples produces greater scientific parsimony ..." (p. 236). Parallel proportional profiles are a thus theoretically desirable property; when we fit PARAFAC or related models to a given body of data, we essentially *postulate* this property of the data. The validity of this postulate can only be indirectly assessed by the success of the fitted model.

The reader is referred to Harshman & Lundy (1984a), p. 163 ff. for further discussion of the empirical validity of this principle. It suffices to observe here that PARAFAC, based on this principle, has been applied to a wide range of psychological and psychometric data with interesting and useful results.

Mathematical background

It is convenient to develop PARAFAC as an extension of the techniques of functional analysis (Kreysig 1978), in order to make apparent the relationship between PARAFAC and other data-analytic procedures. From this perspective, we can view a model as a mapping between two isomorphic spaces. Typically, the elements of one space are simple (e.g., coordinates in an n-dimensional space), and the elements of other are complex (e.g., a function space of some sort). The first space is a parameterization of the second.

As an example, consider a periodic function $f(t)$. This function could be, for example, the pressure at a particular point in a resonating hard-walled tube. Let us imagine collecting a set of such functions, while stimulating the resonating system in different ways. It is well-known that such functions can be represented as Fourier series. Fourier analysis provides:

(i) a *basis* for the set of functions $\{f(t)\}$, consisting of a set of modal displacements, each a

function of t (i.e., the mean pressure, the pressure deviation from the mean as a function of t at the first resonant frequency, the second resonant frequency, etc.) Each of the $f(t)$ observed under different conditions can be represented as a weighted combination of the elements (the sinusoidal pressure wave forms) of this basis.

(ii) a parameterization of each of the individual functions in terms of these weights. Each one of the observed $f(t)$ is parameterized in terms of its *spectrum*, i.e., the collection of weights which represent the contribution of various members of the basis to the observed waveform.

A model of this set of measured functions thus consists of two elements - a basis for the space of observed output functions, and a set of spectra that vary according to the different stimuli applied to the system under observation. Linear combinations of the elements of the basis (e.g., Fourier series summations) represent all of the possible outputs of the system (periodic waveforms from the resonating system); the basis is therefore said to *span* the (function) space containing the sample of observed functions.

A little consideration will suggest that - given an adequate sample of the space of possible stimuli - any possible $f(t)$ may be represented, not only as a linear combination of the elements of the basis, but also as a linear combination of the observed set of $\{f(t)\}$. As an example from a related field, it is well-known that any solution of a (e.g.) a second-order linear differential equation may be represented as a weighted combination of any two independent solutions of the equation. An analogous result applies here.

For functions or other data for which the basis is not known analytically (as it is in the case of periodic functions), a sample of functions from the system under consideration may be used to obtain an empirical basis for the space of functions. An adequate sample of functions may be reduced to a model: a parameter space that we may think of as underlying the observed behaviors and a mapping from this parameter space to a function space that contains the possible $f(t)$. Even samples of the sample of functions can be reduced (as is generally the case, e.g., in digital signal processing: see Ramsay (1982) for an informal overview of how this works).

Practically, it is usually most convenient to analyze a matrix of samples. Consider the matrix **O** with as many rows as we have measures of the behavior of the system (e.g., pressure samples at different times) and as many columns as there are different input conditions. This matrix is a linear operator from a domain in $\mathbf{R}^n$ (e.g., spectra associated with each possible stimulus to the system, expressed as linear combinations of the spectra in the sample) to the range space of possible measures of the behavior of the system (e.g., values of $f(t)$ for arbitrary t). Thus, we have made the same measures of the behavior of

the system for each input condition: a fully crossed experimental design. A minimal representation for **O** is obtainable via the singular value decomposition (SVD). The SVD uniquely factors arbitrary matrices into three components:
- a matrix **U**, the columns of which are a basis for the range of **O**
- a matrix **V**, the columns of which are a basis for the domain of **O**, and
- a diagonal matrix **W**.

The non-zero elements of **W** (i.e., the ones along the main diagonal) are known as the singular values of **O**, and indicate the contribution of the corresponding elements of the bases **U** & **V** to elements of **O**. The vanishing elements of **W** correspond to elements of **V** that contribute little or nothing to **O** (i.e., the nullspace of **V**). The number of non-vanishing elements of **W** corresponds to the rank of **O**, i.e., the dimensionality of its range.

The matrix product $\mathbf{UWV}^T$ is **O**; when an approximation to **O** is sufficient (e.g., when modeling), the vanishing elements of **W** may be zeroed out (see Press et al. (1986) for a discussion of when to eliminate elements of **W** (p. 54 ff.) and how to apply chi-squared tests of significance to the elements of **W** (pp. 516 ff.)). In scalar form, we can eliminate the non-diagonal elements of **W** and write the n non-vanishing elements on the diagonal as w_r. The summation in (2A.1) then represents the SVD of **O**.

$$o_{ij} = \sum_{r}^{n} u_{ir} w_r v_{jr} \qquad (2A.1)$$

The model represents two different aspects of the data: we have the coordinates (in the range space) of each of the (sample of possible) measures; and the coordinates (in the domain space) of each of the (sample of possible) input conditions. These coordinates can be regarded in other ways: for instance, they can be considered the regression coefficients of the particular item on the dimensions that underly the space that the item comes from. The item is thus represented as a linear composite of the basis vectors of the space, just as it would be in a multiple regression.

For instance, Kroonenberg & de Leeuw (1980) show how the stimuli of Miller & Nicely (1955) can be represented in a three-dimensional space whose dimensions correspond roughly to sibilancy, voicing, and nasality. The average response to a particular stimulus in a particular noise condition can be modeled by a linear transformation (i.e., an operator) of the stimulus space that depends on the noise conditions.

O can also be decomposed indirectly, by decomposing the matrix product $\mathbf{S} = \mathbf{O}\mathbf{O}^T$ (e.g., a sum-of-squares and cross-products matrix or a variance-covariance matrix). In this case, the SVD of **S** is its modal decomposition: the w_r are the eigenvalues of **S**, and the columns of **U** and **V** are the eigenvectors of **S**. The set of eigenvalues of **S** is of course its spectrum, and the eigenvectors of **S** are the modes, or principal components of whatever set of data we are modeling. In the case of continuous, rather than sampled, functions, we obtain eigenfunctions (again, see Ramsay 1982).

It is worth noting that the eigenvectors of **S** are invariant with respect to the reference point for the measurements in **O**. A change in the origin of the coordinate system containing the measured functions is reflected only in changes to the eigenvalues w_r (the *shifting theorem* for eigenvectors: see Press et al. 1986, p. 335 ff.).

PARAFAC represents a generalization of the SVD which constructs several simultaneous SVD-like representations. Like the SVD, PARAFAC is unique. PARAFAC is designed to deal with data for which there are more than the two aspects of variation mentioned above. One could, for example, apply PARAFAC to rating produced by several subjects of several stimuli on several scales. PARAFAC may thus be conceived of as applying to a three-dimensional array, rather than a two-dimensional matrix, of data. The model may nonetheless be written as a series of conventional matrix equations as in (2A.2), or in scalar form as in (2A.3). The notation $\mathbf{O}_k$ is employed to signify the *k*th slice out of the three-dimensional array of data, e.g., the *k*th subject's ratings. As before, the **W** matrices are diagonal.

$$\mathbf{O}_k = \mathbf{U}\mathbf{W}_k\mathbf{V}^T \tag{2A.2}$$

$$o_{ijk} = \sum_{r}^{n} u_{ir} w_{kr} v_{jr} \tag{2A.3}$$

In these forms, the **U** and **V** matrices are the same for all levels of *k*. **U** and **V** thus provide unique, subject-independent coordinate systems for the domain (stimulus) space and range (scale) space. Traditionally, the column vectors of these matrices are known as *factors*; the elements of each these vectors are known as the *loadings* of the stimuli or scales on that factor. The set of loadings associated with a particular stimulus (e.g., a row vector of **U**) can be thought of as the coordinates of that stimulus in a parameterized stimulus space.

Subject variation is modeled only by variation in the non-zero diagonal elements of $\mathbf{W}_k$. These subject-specific weights can be thought of as introducing subject-specific transformations of the stimulus space to produce a subject-specific perceptual space. In the subject-specific perceptual space, each stimulus is represented by (a linear composite of) that subject's ratings of the stimulus.

The subject-specifc transformations are restricted to scalings along the axes of the space - i.e., *proportional increases or reductions of each stimulus' significance with respect to a particular dimension*. Thus, PARAFAC models the data according to the principle of parallel proportional profiles advocated in Cattell & Cattell (1955).

In another parallel with the SVD, PARAFAC can be applied to generate a sort of simultaneous modal decomposition of variance-covariance matrices. Assuming that the raw data obey the form of (2A.2), the within-subject variance-covariance matrices each obey the form in (2A.4).

$$\mathbf{S}_k = \mathbf{O}_k\mathbf{O}_k^T = \mathbf{U}\mathbf{W}_k\mathbf{V}^T\mathbf{V}\mathbf{W}_k\mathbf{U}^T \qquad (2A.4)$$

If the stimuli are distributed randomly or orthogonally with respect to the underlying dimensions of the stimulus space, the product $\mathbf{V}^T\mathbf{V}$ will tend towards $\mathbf{I}$. In this case, (2A.4) reduces to (2A.5).

$$\mathbf{S}_k = \mathbf{U}(\mathbf{W}_k)^2\mathbf{U}^T \qquad (2A.5)$$

This is what is known as the PARAFAC1 model for covariances (Harshman & Lundy 1984a). This model is convenient in that the same algorithm which is used to fit raw data to the model in (2A.2) can be used to fit (2A.5) to variance-covariance matrices, just as the SVD can be used to obtain the modal decomposition of a variance-covariance matrix. It is not a strictly appropriate model to use if the orthogonality condition mentioned above does not hold. Nonetheless, it often produces factors similar to those of the model in (2A.2).

A final similarity between the SVD/eigenanalysis and PARAFAC is that the location of the origin of domain space does not affect the factors: at most, PARAFAC finds a factor that has roughly constant loadings that represents location of the centroid of the data sample (see Harshman & Lundy 1984b, p. 261 ff.; Harshman & De Sarbo 1984, p. 606 ff.). When there is a possibility that the location of the centroid is significant, it of course becomes important to determine an absolute origin for the coordinate system, so

that the centroid, or overall mean, of the data can be characterized. Nonetheless, the modes of variation in the data will be the same, no matter what the coordinate system is.

There are, nonetheless, some differences between the SVD and PARAFAC. First, PARAFAC generally does not produce an orthogonal coordinate system for the particular stimuli or rating scales used in gathering the data. The SVD by definition does. In certain situations, of course, the property of producing orthogonal coordinate systems is desirable, and PARAFAC can be constrained to produce them. However, when a study is designed to produce multiple measures of several aspects of one hypothesized underlying entity, i.e., intrinsically correlated measures, imposed orthogonality could well distort the analysis, and is therefore undesirable.

Second, PARAFAC's subject-specific weights (the diagonal elements of $\mathbf{W}_k$) cannot in general be interpreted as the contribution of the corresponding column vectors of **U** and **V** to the data. Thus, they cannot be interpreted in the same way as the singular values of the SVD. This is because of the non-orthogonality of the PARAFAC coordinate system, which means that the factors do not uniquely partition the variance of the data set. This problem is analogous to the problem of multicollinearity in multiple regression. But even if the bases produced by PARAFAC for a particular data set are orthogonal, the fact that the variance along each axis of the domain and range spaces is *hypothesized* to vary for each individual makes it difficult to interpret the subject-specific weights in terms of an overall contribution.

It should be noted that the principle of parallel proportional profiles and PARAFAC are not the only possible conceptions of the structure of psychological data. More general transformations (e.g., rotations) are possible, and are treated in models like those proposed by Tucker (1966) - see Kroonenberg & de Leeuw (1980) for more recent discussion of these models. However these models lack the uniqueness of PARAFAC in that their subject-specific transformations are indeterminate. They are also not endowed with an interpretive principle like the parallel proportional profile principle. And finally, PARAFAC appears to be adequate for the kind of phonetic data we are considering here. Indeed, it is hard to imagine that any model could substantially improve on the R^2 of over 0.96 that Harshman et al. (1977) obtained with only two dimensions.

Applicability of the model

Why should it be that PARAFAC, originally developed for a very different kind of psychological data, fits articulatory data? The answer lies in the fundamental physical nature of the data considered. Articulators such as the tongue are only capable of relatively limited deformability and displacement. In vowels, the tongue is subject to relatively little contact deformation, which would introduce non-linearities into the

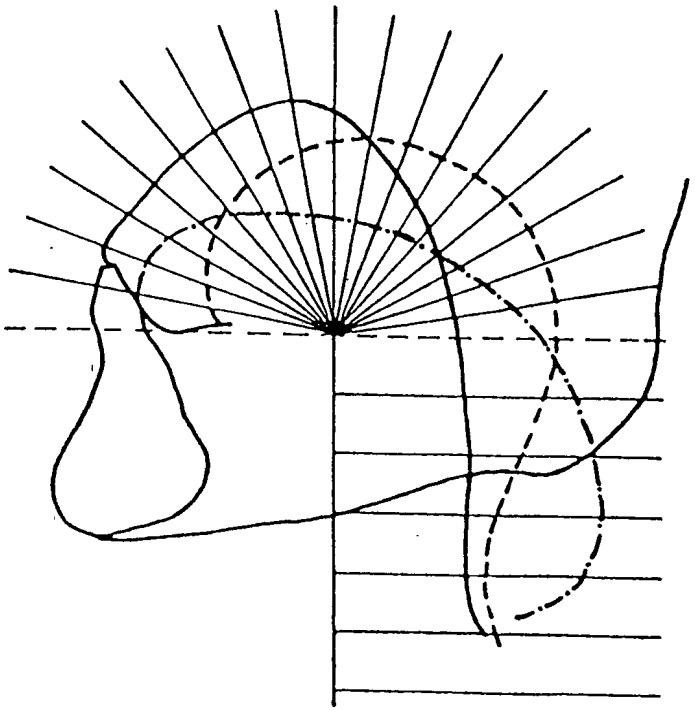


Figure 2A.1. Several tongue positions in vowels (after Lindblom & Sundberg 1971).

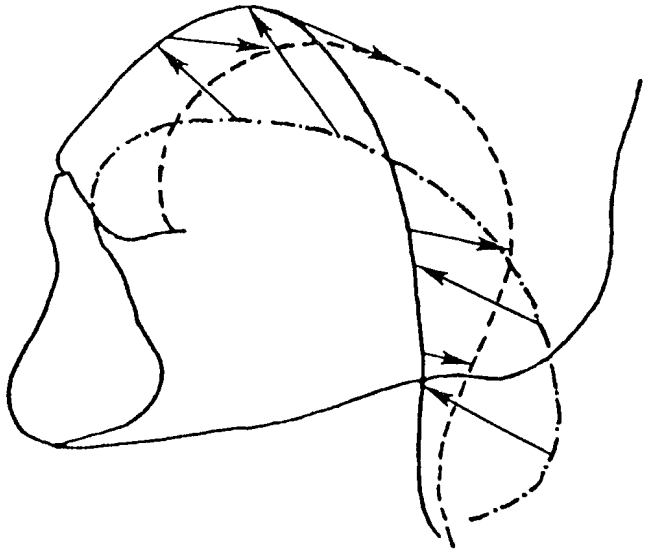


Figure 2A.2. (Hypothetical) mappings of one tongue position onto another.

measurements of tongue position. Thus the functions representing vocal tract position in the midsagittal plane tend to have a high degree of continuity and to be mathematically tractable.

Let us begin by considering the relations between articulator positions in various different vowels. Fig. 2A.1 shows several superimposed tongue positions observed during the production of some Swedish vowels (after Lindblom & Sundberg 1971). We can map one position of the tongue into another position by specifying the displacement at each point of the tongue required to bring the first position into the other position. Such a mapping must exist, since the same tongue does in fact assume all these positions!

Two such (hypothetical) mappings are schematized in Fig. 2A.2, in which the tongue position (relative to the jaw) for the vowel /**a**/ is mapped onto the position for the vowel /**i**/, which is in turn mapped onto the position for the vowel /**u**/. We can think of mapping positions onto each other in this way as specifying a *displacement vector* **y** at each point of the tongue. We can define a set of (vector-valued) functions $\mathbf{y}_{tongue}$, $\mathbf{y}_{jaw}$, etc., that represent the positions of the various organs of the vocal tract at a particular instant in time. As discussed above, the choice of reference position does not substantially affect the results of the analysis, so we may use any convenient position, such as the mean position, as a reference. We can then, for instance, describe each of the observed tongue positions in terms of a vector-valued function $\mathbf{y}_{tongue}(\mathbf{p})$, defined at each point **p** of the tongue surface.

The task of modeling these functions then becomes in part the task of finding a basis that allows each **y** to be represented as a weighted sum of the elements of the basis. Various analytically tractable approximations are possible - it has become customary in the acoustical modeling world to use a staircase function to approximate the tongue contour (or the vocal tract radius). Mermelstein's (1973) model uses a piecewise approximation based on geometrical constructions; Liljencrants' (1971) model uses a Fourier-series approximation. Other approximations could be considered - in particular, approximation by using the Chebyshev polynomial series (cf. Jospa 1974). Unlike the Fourier cosine series that Liljencrants used, which concentrates the error at the edges of the modeled interval, Chebyshev approximation has the attractive property of distributing the modeling error uniformly over the modeled interval. (See Ramsey (1982) for further comments on techniques for approximating continuous functional data in numerically tractable ways). Whatever the approximation we use, we can consider its parameters to be a kind of measure of vocal tract shape.

These approximations lead to representations of vocal tract shape in terms of linear combinations of the basis set of the approximation. Thus, Liljencrants (1971) represents

vocal tract shapes as a sum mation of cosines, and Jospa (1974) represents vocal tract shapes as a summation of Chebyshev polynomials. But if, instead of picking some *a priori* basis, we choose to represent vocal tract positions as linear combinations of 'modal displacements' of the tract $\mathbf{d}_r$, then the position of the vocal tract in the *j*th vowel can be written as a summation of modal displacements that are analogous to eigenvectors. This kind of representation is represented algebraically in (2A.6).

$$\mathbf{y}_j = \sum_{r}^{n} w_{jr}\mathbf{d}_r \qquad (2A.6)$$

We may think of these modal displacements as arising from the action of various articulatory organs, with each articulatory organ giving rise to a single mode of tongue displacement. For instance, the various parts of the genioglossus - posterior, medial, and anterior - probably each contribute a distinctive mode of tongue position displacement. Modeling studies such as Kakita et al. (1985) have found it necessary to model this one muscle as severally functionally independent units. It thus appears that some unit smaller than the entire muscle will often be relevant at this level of articulatory description.

When dealing with data from several speakers, it is necessary to recognize that different speakers have different sizes and proportions. Although normal speakers generally share the same gross anatomical structures, they vary a fair amount in overall and relative size. (In addition, the process of reproducing x-ray tracings for publication may introduce an unknown variable scale factor). As a first approximation, we should, therefore, include a scale normalization factor associated with each mode of articulatory displacement. Other more elaborate schemes, such as extracting an overall speaker-normalization factor first, are possible, but they turn out to have little substantive effect. We thus arrive at an expression like that in (2A.7) as the expected form of a multisubject sample of measurements of articulator position over a number of vowels.

$$\mathbf{y}_{jk} = \sum_{r}^{n} w_{jr}\mathbf{d}_r s_{kr} \qquad (2A.7)$$

Is this kind of model justified? Are the assumptions of linearity and scalability embodied in (2A.6) and (2A.7) justified? Studies like that Sekimoto et al. (1978) provide important evidence in favor of the plausibility of these assumptions. Sekimoto et al.

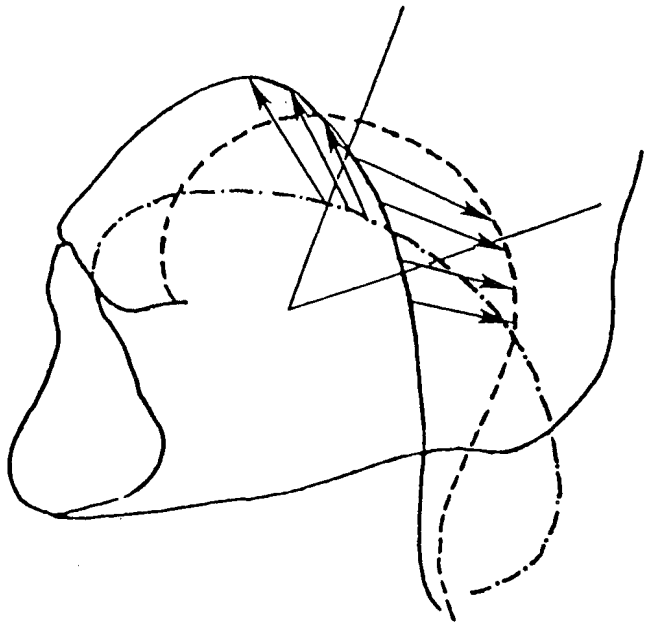


Figure 2A.3. (Hypothetical) displacement vectors and measurement gridlines.

gathered raw measures of articulator position and applied a form of factor analysis to their results with quantitatively good results. The researchers 'labelled' individual tongue points with which were then tracked with the x-ray microbeam system at the University of Tokyo. Sekimoto et al. (1978) report on data from five subjects, fitting a two-factor model to each. The first factor of their articulatory model consists of all the tongue displacements linearly related to the jaw displacement. The second factor is the first principal component of the residual tongue displacement. With each of the subjects analyzed individually, these two components suffice to bring the standard deviation of the model error to less than 1.7 mm. (For comparision, the root mean squared error reported by Harshman et al. (1977) is about 1.73 mm; the rms error of the Icelandic PARFAC analysis in Jackson (1988) is 1.68 mm). They also show that the factors derived from analyses of static vowels and from analyses of vowels in running speech are very much the same. These results show that linear models very similar to the ones that we are considering here are capable of producing very good fits to articulatory data.

As is detailed in Chapter 3, I have measured articulator positions using gridlines, since most of the available x-ray data do not include radio-opaque pellets. Geometrical considerations lead me to believe that this does not perturb the model in (2A.7). Fig. 2A.3 shows some (hypothetical) displacement vectors mapping one tongue position into another. Two sample gridlines are also shown. The difference between the two positions of the tongue, as measured by the difference between positions of the tongue along a gridline, is approximately proportional to the magnitude of the displacement vectors in the neighborhood of this gridline. In particular, if θ_i is the angle between a displacement vector that crosses the *i*th gridline and the gridline itself, then the displacement of the surface of the tongue along the gridline is approximately $\cos \theta_i \, ||\mathbf{y}||$. If we denote the point at which the *i*th gridline crosses the tongue in the reference position as $\mathbf{p}_i$ and the angle between $\mathbf{d}_r(\mathbf{p}_i)$ and the gridline as θ_{ir}, then the displacement d_{ir} along the *i*th gridline is approximately $\cos \theta_{ir} \, ||\mathbf{d}_r(\mathbf{p}_i)||$. Thus, (2A.8) may be derived from (2A.7) for measurements of articulator position along gridlines.

$$y_{ijk} = \sum_{r}^{n} w_{jr} \cos \theta_{ir} \, ||\mathbf{d}_r(\mathbf{p}_i)|| \, s_{kr} \qquad \text{(2A.8)}$$

θ_{ir} will in general be different for each *r* and each *i*, i.e., for each modal displacement at each gridline. Thus, the proportionality constants $\cos \theta_{ir}$ are not identifiable from data

of this sort. We may nonetheless lump the terms $\cos \theta_{ir} \, ||\mathbf{d}_r(\mathbf{p}_i)||$ and write (2A.9).

$$y_{ijk} = \sum_{r}^{n} w_{jr} d_{ir} s_{kr} \qquad (2A.9)$$

This expression is the expected form for the tongue displacement observed at the *i*th gridline during the *j*th vowel's production by the *k*th speaker. It is a PARAFAC form, as may be seen by comparision with (2A.3). Although we have only considered the tongue here, these arguments generalize to other articulators. Thus, the PARAFAC model appears to fit our conception of the nature of articulatory data well.

We have postulated that coordinative structures may be described in terms of functional dependencies governing the usage of (pairs of larger sets of) articulatory primes. In (2A.9), we would express this (in at least the first approximation) by introducing linear dependencies between the *r*th and *s*th articulatory primes, i.e., by setting $w_{jr} = kw_{js}$ for all j. The introduction of linear dependencies - correlations - like this into (2A.9) makes d_{ir} and d_{is} no longer jointly identifiable (again, the analogy with multicollinearity is appropriate). Rather, only their combination $d_{ir}+kd_{is}$ and their joint contribution w_{jr} are identifiable. Thus, *PARAFAC factors generally should be viewed as representing the articulatory correlates of coordinative structures*, and not articulatory primes.

The expression of coordinative structures in this model is still assumed to be governed by the principle of parallel proportional profiles. To expand on the example given in the body of the text, if a particular speaker uses less lip protrusion and more larynx lowering than another in a particular vowel, she/he is assumed to make more use of these articulatory displacements in all vowels. If this assumption of systematic variation is valid, a unique and explanatory set of factors that obey the principle of parallel proportional profiles can be obtained from the data by PARAFAC.

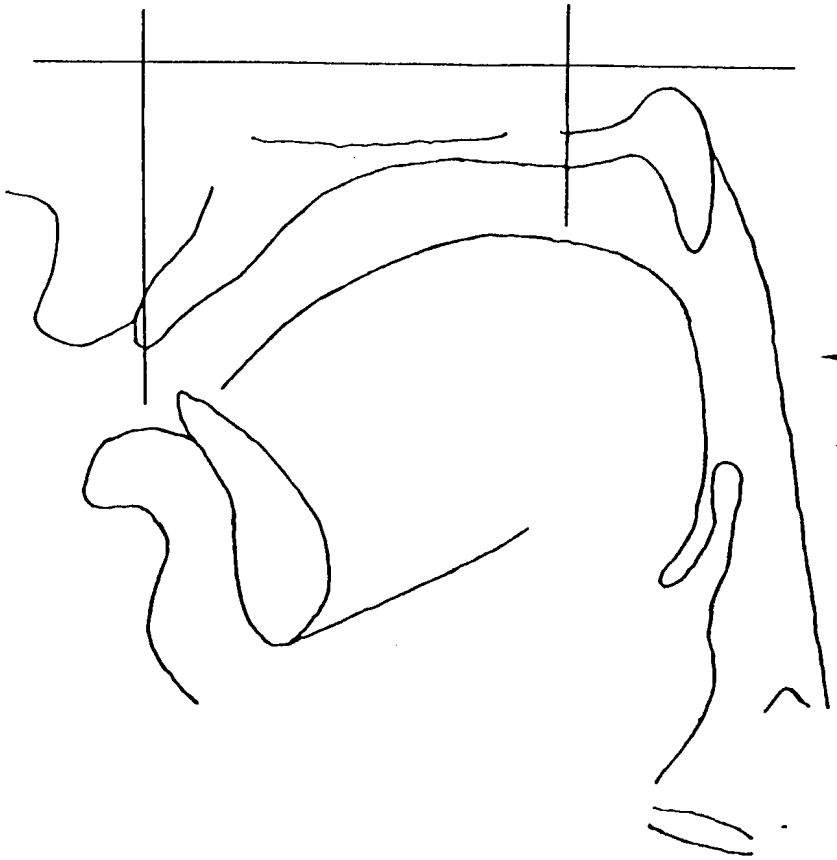


Figure 3.1. Horizontal reference line for Bothorél et al.'s (1986) speaker 1.

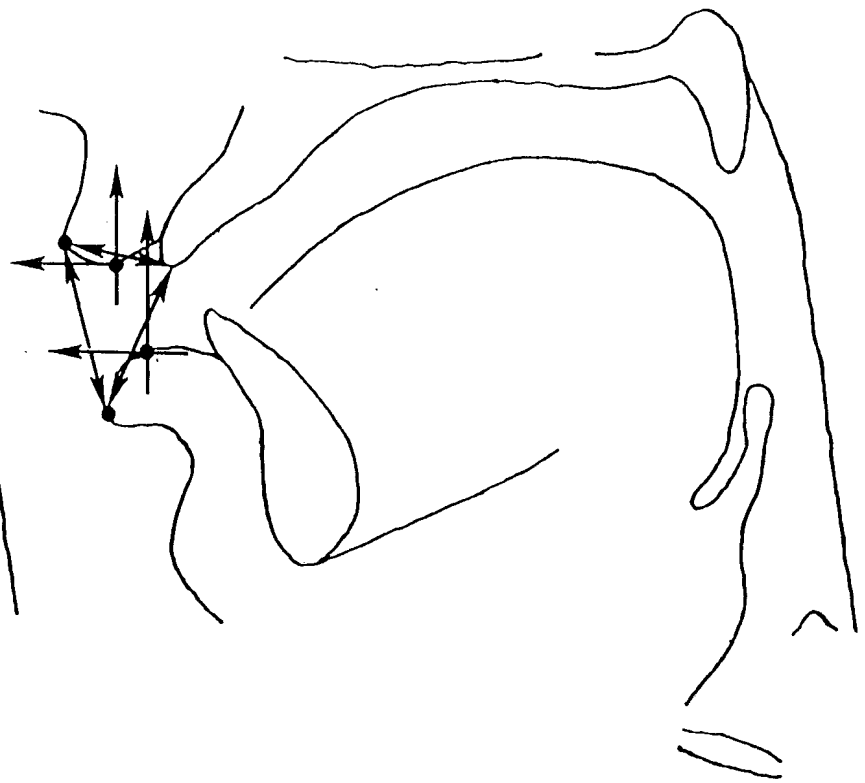


Figure 3.2. Measures of lip position.

Chapter 3: Selecting the data: considerations for this study

The aim of this chapter is to present the data that this study is concerned with, along with a methodology for verifying the models constructed. We will also survey the phonological and phonetic systems of the languages which will be used. Thus, we will survey the various aspects of content - articulatory, linguistic, and cross-linguistic - that this study is concerned with quantifying.

Sampling the space of articulatory measures

We are basically concerned with modeling the mid-sagittal profile of the vocal tract during vowel production. It is generally believed that only the mid-sagittal profile of the vocal tract is linguistically distinctive, since vowels generally involve central, and not lateral, articulations. Furthermore, there exist algorithms for approximating the overall shape of the tongue, given the mid-sagittal profile (Kakita et. al 1985, Hashimoto & Suga 1986). The scheme for approximating the mid-sagittal profile of the vocal tract is based on the tongue gridline construction procedures of Harshman et. al (1977) and Jacobsen (1978). There are, however, several additions, since we are concerned with more than the position of the tongue in this study.

The articulatory organs that are generally considered relevant to vowel production are the lips, jaw, tongue, epiglottis, velum, pharynx, and larynx. There is, of course, a long phonetic tradition of measuring the positions of these articulators during speech, and I will be drawing on it.

The first step in measuring the positions of the various articulators from an x-ray tracing for the purposes of this study was the construction of a horizontal reference line. This line was approximately perpendicular to the axis of the pharyngeal cavity in a vowel having a relatively uniform mid-sagittal diameter, as is often seen in [ɛ]- or [æ]-like vowels. As far as possible, this reference line was also parallel to the superior surface of the palate. The reference line used for one speaker (the fourth speaker in Bothorel et al. 1986) is shown in Fig. 3.1.

Perpendiculars to this reference line were dropped through the apex of the upper incisor and through the most anterior point of the soft palate that appeared to change position in different vowels.

Lips

Linker (1982) studied lip positions in English, Cantonese, Finnish, French, and Swedish. In her measurement scheme, she determined the positions of five points from simultaneous frontal and lateral photographs of the mouth. Seventeen distances between these points were then calculated and used in her PARAFAC study along with seven other distances and areas measured directly off the photographs.

The PARAFAC study suggested that certain points on the lips, and the distances between them, were strongly related to the systematic patterns of lip movement in the languages studied. On these grounds I selected the following measures of lip position:

- llx, lly. The x- and y- coordinates of the point of the lower lip at the maximum constriction of the lips. These quantities were measured using the apex of the upper incisor as the origin of the coordinate system, with the x-direction parallel to the horizontal reference line.
- ulx, uly. The corresponding coordinates, on the upper lip.

These measures are meant to approximate the points at the two ends of the distance that is Linker's measure 21, which was the vertical opening between the lips as seen from the front.

- ild. Inter-lip distance - the distance between the most anterior points of the lips (corresponding to Linker's measure 7).
- lid. Lower lip-incisor distance - the distance between the most anterior point of the lower lip and the upper incisor (corresponding to Linker's measure 3).
- uid. Upper lip-incisor distance - between the most anterior point of the upper lip and the upper incisor (Linker's measure 1).

The most anterior points of the upper and lower lips were found by constructing a line tangent to both the lips. The point at which this tangent line contacts each lip was taken as that lip's most anterior point.

I have changed Linker's procedure somewhat to avoid introducing artificial dependencies between the measures. In particular, her procedure of determining the positions of five points first (i.e., an x- and a y-coordinate for each, or ten degrees of freedom), and then calculating seventeen distances between them (i.e., seventeen supposed d.f.) seems risky and likely to magnify correlated errors. Therefore, I either retained the x- and y- coordinates of the points, or else measured the indicated distances separately. These measures of lip position are shown in Fig. 3.2.

Jaw

The lower lip's position, at least, is greatly influenced by the position of the jaw. For this reason, it seems a good idea to measure the position of the jaw as well. It is commonly assumed that the jaw moves only over a segment of a circular arc (as in Mermelstein's 1973 model, or in Kelso, Saltzman, & Tuller 1986), or even that its motion can be approximated by translation at a specific angle (as in Kakita & Fujimura 1977). However, there is experimental evidence that both of these approximations are incorrect and that both vertical and horizontal jaw movement in natural speech are independent and appreciable (Edwards & Harris 1985). Therefore, the x- and y- coordinates of the

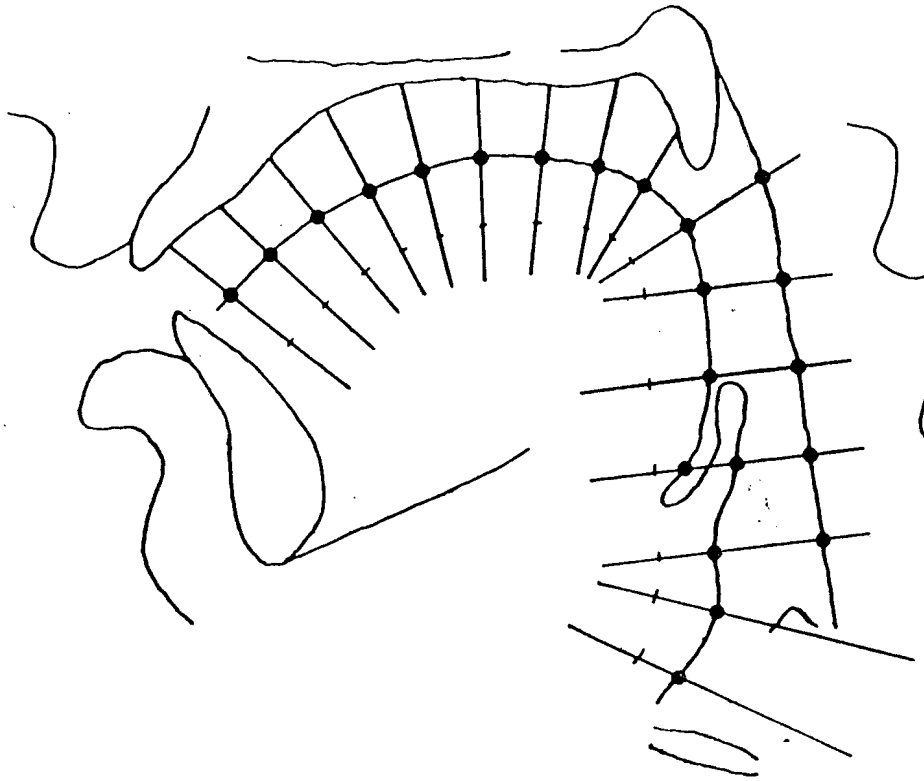


Figure 3.3. Measures of tongue, dorsal wall of the pharynx, and epiglottis position.

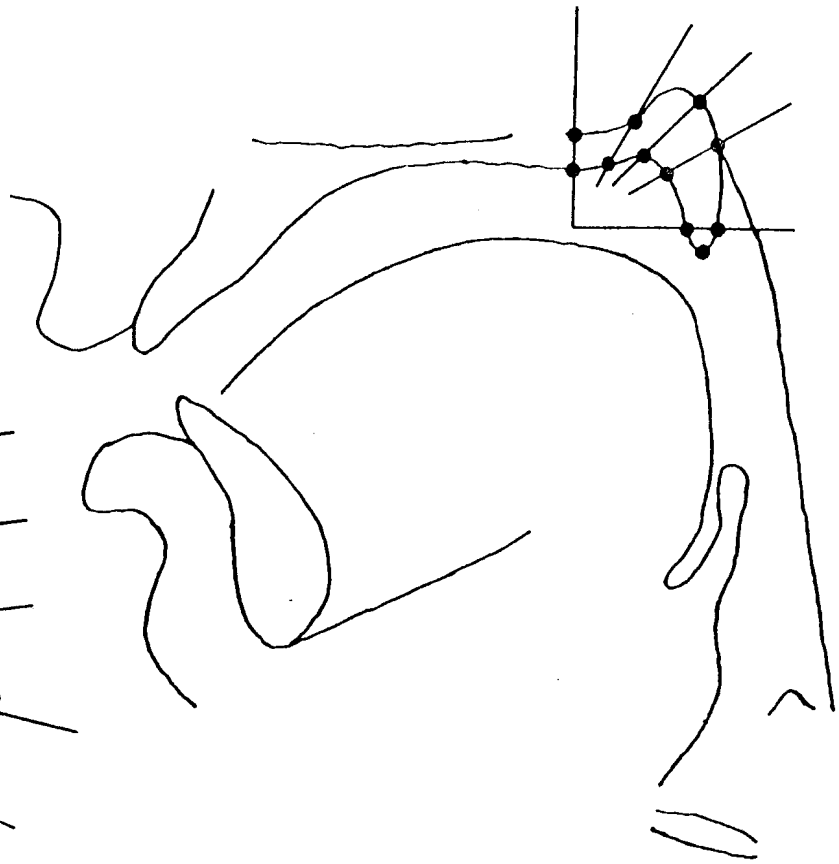


Figure 3.4. Measures of velum position.

apex of the lower incisor (abbreviated ljx, ljy), relative to the upper incisor, were also measured.

Tongue, epiglottis, and dorsal wall of the pharynx

The procedure for measuring the positions of the tongue, epiglottis, and dorsal wall of the pharynx is adopted from Harshman et al. (1977), with some modifications inspired by Jacobsen (1978). The starting point is a vocal tract in a position such that the mid-sagittal diameter is fairly uniform over the entire length of the tract. Such postures are typically found in [ɛ]- or [æ]-like vowels. The midline of the vocal tract is constructed and it is divided into sixteen sections. The sections, beginning at the laryngeal end of the tract, are distributed as follows:

- sections 1-3 are such that the center of section 3 is below the root of the epiglottis, and the sections are of equal length.
- sections 4-7 are such that the center of section 4 is opposite the epiglottis, and the center of section 7 is a few millimeters below the level of the uvula (relative to the horizontal reference).
- sections 8-16 are of equal length, with section 16 ending at the upper incisor.

These sections were constructed by beginning with sixteen equal-sized sections, and then making the adjustments necessary to satisfy the above conditions. Thus, the sections are all as similar in length as possible given the proportions of each individual speaker. The distribution of these sections is constrained by the anatomical landmarks mentioned in order to normalize for differences in pharynx length relative to total vocal tract length: males are known to have proportionately longer pharynxes than females.

At the center of each section, a gridline perpendicular to the midline was constructed, and an arbitrary origin was marked about one centimeter from the surface of the tongue toward the center of the tongue. The position of the epiglottis was measured along gridlines 1-4 (abbreviated e1-e4); the tongue, along 4-16 (t4-t16); and the dorsal wall of the pharynx, along 3-7 (p3-p7). Fig. 3.3 illustrates these gridlines.

It does not seem necessary to correct measures of tongue position for jaw position (e.g., to adopt a jaw-relative coordinate system for tongue positions). Liljencrants (1971, 1985) tried using both mandibular (jaw-relative) and maxillar (fixed) coordinate systems of the tongue, concluding

> There seem to be no tangible gains in using the mandibular position as a reference in the description [of tongue shapes: mttj] ... It is thus recommended that the mandibular position is not used as a primary articulatory parameter, but rather as a secondary parameter, dependent on the others (Liljencrants 1971, p. 18).

Liljencrants (1971, 1985) also empirically investigated the degree to which his results were dependent on the precise placement of the gridlines, concluding that as long as their placement was consistent, there was little substantive difference in the results.

> A different problem is to determine how critical the position of the coordinate system is with relation to the stationary structure ... the coordinate system has been translated .5 cm forwards and .5 cm upwards. Inspecting the polar coefficient plots ... shows that the distribution of the coefficients is very closely similar. We see that the set of data points has been translated and somewhat rotated, but the interrelations are essentially unchanged (pp. 14-15).

We might add that such is the result to be expected in a procedure that fits a model linear in the unknown parameters (e.g., Liljencrants' Fourier coefficients, Harshman et al.'s PARAFAC loadings). The space may be stretched, rotated, etc., but as long as linearity holds, the space is essentially invariant, give or take an affine transformation. Nonetheless, from the point of view of *interpretation*, it will be crucial what the rotational position of the vowels with respect to the axes of the space is. It is for this reason that is important to have a good match between our conception of the factors underlying articulator position and the model.

Velum

The velum is traditionally regarded as a completely independent articulator. It is apparently capable of acting orthogonally to the other articulators involved in vowel production. For instance, Maddieson's (1984) survey of 317 languages lists 71 languages (22.4%) that have phonemic nasalized vowels (produced by lowering the velum). By far the greatest number (35 by my informal tally) have nasalized counterparts to every one of the oral vowel qualities of the language, suggesting that nasalization (i.e., velum lowering) combines freely with all other articulatory gestures in vowels.

However, there are several kinds of evidence that might lead us to believe that the phonetic situation is not quite so simple. It has been observed in the phonetic literature that low vowels are more nasal or have greater velum lowering than high vowels (Fant 1960, p. 139; Moll 1962). A further look at Maddieson (1984) reveals that, for languages which do not have freely combining nasalization, there are two kinds of "missing" nasalized vowel.

One tendency, widely noted in the literature, is for mid vowels like [e] and [ɛ] to have only one nasal counterpart, e.g., [ɛ̃]. It is generally attributed to the difficulty of perceiving the difference between pairs like [ẽ/ɛ̃], and does not concern us here.

The other tendency is for some or all high nasalized vowels to be missing. About 10 languages (Breton, French, Seneca, Yuchi, Kpelle, Songhai, Lakka, Nambakaengo, Southern Nambiquara, and !Xũ by my informal tally), from many different language families, show this tendency. There are, however, no languages which have nasalized high vowels but lack nasalized low vowels. There thus appears to be a small but wide-spread preference for nasalized low vowels among languages of the world. If the observation that low vowels (both phonemically oral and phonemically nasal) tend to be more nasal is correct, it could be that low nasalized vowels are somewhat preferred to high nasalized vowels on phonetic, and presumably articulatory, grounds. The increased nasalization of low vowels is sometimes hypothesized to be a mechanical effect of tongue height. The palatoglossus, which runs from the soft palate, down the side of the oropharynx, to blend into the muscles of the back of the tongue, presumably pulls down on the velum when the tongue is lowered.

If this effect exists, it would be nice to quantify it. Therefore, I have constructed gridlines on which to measure the positions of the inferior and superior surfaces of the soft palate and velum. The gridlines are constructed by dropping a perpendicular from the horizontal reference line through the most anterior point of the soft palate that appears to change position from vowel to vowel. This point is generally a few milimeters posterior of the hard-soft palate juncture, when the juncture is marked on x-ray tracings. A line parallel to the horizontal reference is then constructed so that it intersects the velum two or three milimeters above the tip of the uvula. Finally, additional gridlines at angles of 30, 45, and 60 degrees to the horizontal reference are constructed, passing through the intersection of the first two gridlines.

The positions of the inferior and superior surfaces of the velum along these gridlines (abbreviated vi1-vi5 and vs1-vs5, with 1 denoting the gridline parallel to the horizontal reference and 5 denoting the gridline perpendicular to it) were measured. The x- and y-coordinates of the tip of the velum, relative to the upper incisor, were also measured (tvx, tvy). Fig. 3.4 shows how these measurements are made.

Larynx

The larynx has been implicated in several kinds of vowel articulation. Lindau (1979) points out that larynx lowering accompanies tongue root advancement for some speakers of Akan, which leads her to postulate a feature [expanded pharynx] to replace or supplant [advanced tongue root]. Riordan (1977) has shown that larynx lowering is employed by some speakers to accentuate the acoustic difference between front rounded and front unrounded vowels. Larynx position presumably might also varies with phonation type in vowels, although the data in Thongkum (1987) shows no sign of such

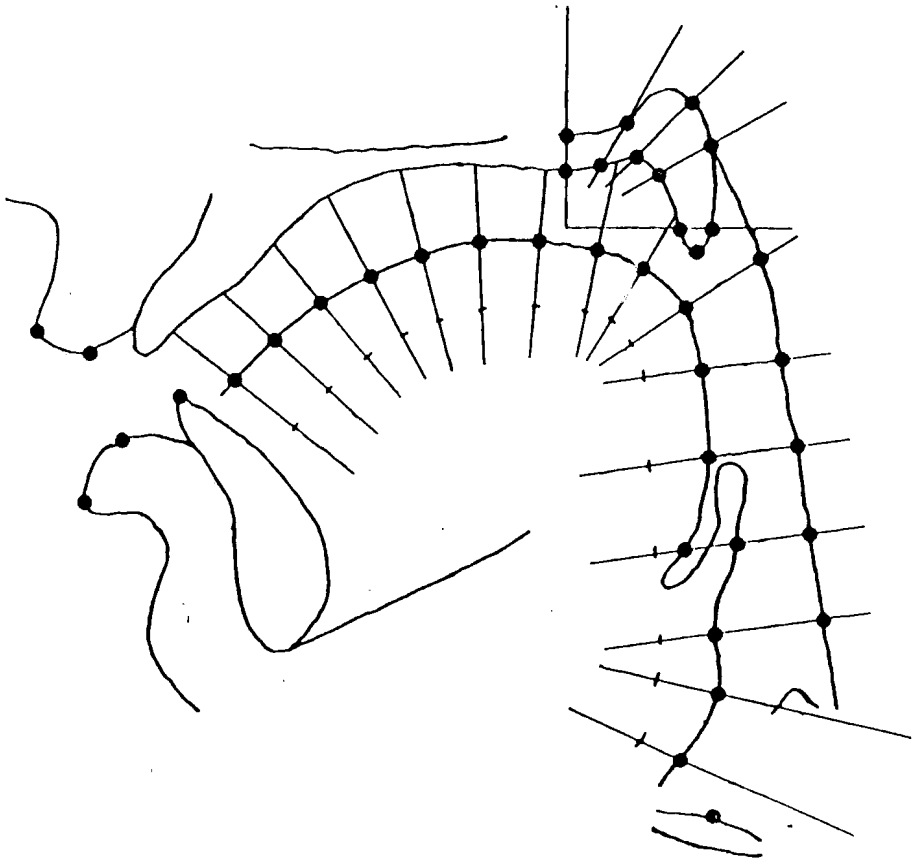


Figure 3.5. A complete measurement grid.

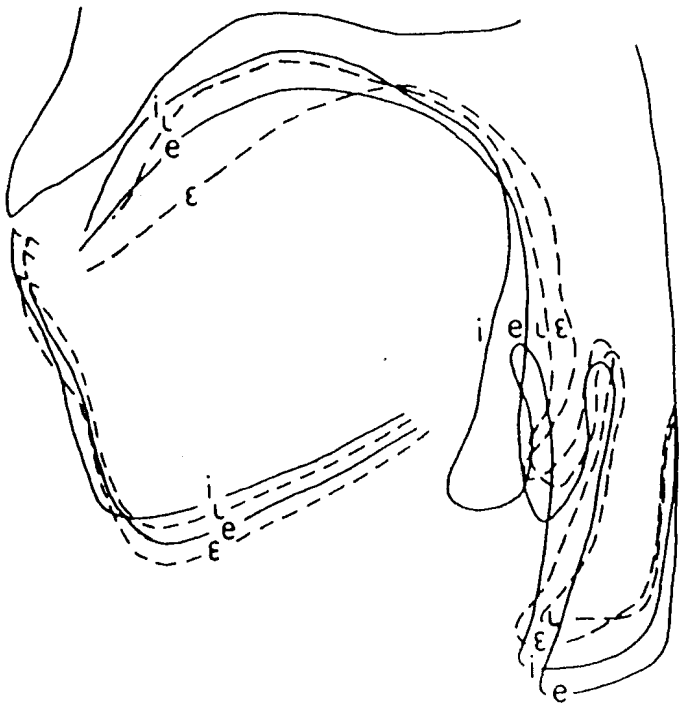


Figure 3.6. Some Akan from oral vowels (after Lindau 1978).

variation.

Relative larynx height (lyy) was measured by finding the perpendicular distance from the larynx to the horizontal reference. For some sets of x-rays, the vocal folds inside the larynx were traced, and I used the distance from the vocal folds to the horizontal reference. For others, the vocal folds were not traced, but the arytenoidal cartilages at the posterior of the larynx were. In these cases, I measured the height of the apex of the arytenoid. Since the larynx is a fairly rigid structure, and the vocal folds are attached to the arytenoid at their posterior end, the relative deviations of larynx positions on these measures should be comparable. Of course, some authors did not trace the laryngeal structures at all. For these speakers, nothing was measured.

The grids thus constructed were then checked against all the x-ray tracings of vowels for the speaker in question. The slopes of gridlines 1 & 16 were sometimes adjusted to ensure that they intersected the surface of the epiglottis and the tip of the tongue in as many vowels as possible.

A typical completed measurement grid is shown in Fig. 3.5. It is based on the vowel [ɛ] as produced by the fourth speaker in Bothorel et al. (1986). The positions measured are shown by the dots on the figure; all measurements were taken to the nearest 0.5 mm on the tracing.

Sampling the vowel space and sampling languages

Having surveyed the questions associated with properly sampling the articulatory content of x-ray tracings of vowels, we now turn to the other aspects of variation within and between various languages and their speakers. Given that different languages often have different phonologically distinct set of vowels, these aspects must be considered together. Both principled and pragmatic considerations dictate the selection of languages and their vowels for this study.

Availability of data. Although the UCLA X-Ray Bibliography (Dart 1987) lists over 280 articles which include x-ray data, the number of languages which have received comprehensive treatment is disappointingly small. The number of languages for which reliable, dialectally homogeneous, multi-speaker samples can be constructed is even smaller. Geographical coverage is extremely irregular: I have not seen any x-ray work on American, Central and Eastern African, or Australian languages.

Phonological bias. For entirely understandable reasons, most researchers have not made an effort to systematically sample the entire phonetic space, preferring to concentrate rather on prototypical, phonemic, highly-differentiated vowels. Most x-ray studies therefore do not give extensive descriptions of the phonetic variants of each vowel phoneme. For instance, in Standard Chinese (Putonghua), the phonetic

vowel qualities are [**i, ɛ, y, ə, u, ɔ, a**] (Bao & Yang 1986, Zhou & Wu 1963). However, works like Ohnesorg & Svarny (1955) only provide x-rays of the phonemic vowels /**i, y, ə, u, a**/. Similarly, Jacobsen (1978) only took x-rays of the vowels of DhoLuo that appeared to participate in the [+/- advanced tongue root] contrast. This means that the space of articulatory possibilities in many languages is underrepresented.

Genetic resemblances. The vowel inventories of languages are clearly strongly influenced by their historical affiliations. For instance, in the 317 language sample in Maddieson (1984), the modal number of vowels per language is five, the mean is about seven, and the range is from three to 24. However, the mean number of vowels among the Indo-European languages in the sample is over ten. 24% of the Indo-European languages have more than ten vowel *qualities* (i.e., discounting length, nasalization, etc.), as opposed to only 3.5% of the sample at large. At the other end of the scale, things are similarly skewed. Eight of the nineteen Australian languages in the sample have only three vowels; the continental average is barely over four. The Australian languages (5.9% of the sample) make up nearly half of the 18 languages (5.7% of the sample) with three vowels. So both the high and the low ends of the distribution are dominated by specific families. The concentration is even worse than it looks: in Indo-European, the Germanic group, and in Australian, the Pama-Nyungan group, are responsible for a disproportionate share of the cases.

There are also apparently phonological contrasts that are limited to certain subgroups or families. For instance, Fischer-Jørgensen (1985) observes that the tense/lax contrast in Dutch and German does not sound like what has been called tense/lax in English. (However, Lindau (1978) shows that tense/lax contrasts in German, for at least one speaker (Wängler 1961), are like tense/lax contrasts in English in that they involve mostly tongue height). The term tense/lax has also been used in South-East Asian languages and in West African languages, but in both cases, it seems probable that the contrast is not phonetically the same one that Germanic languages use (Thongkum 1987, Maddieson & Ladefoged 1985). Thus it seems that the Germanic tense/lax contrast is rather restricted. Not only is it unlikely to be found in other language families, but even inside Germanic, it is probably not found in English. Similarly, contrastive use of [advanced/retracted tongue root] appears to be limited to certain areas of Niger-Kordofanian (reports of its use in South-East Asia seem to be exaggerated: see Thongkum 1987). Pharyngealization, uvularization, and velarization of vowels are similarly restricted.

In view of these known familial peculiarities, it seems only prudent to control for

genetic affiliation when selecting languages. The best strategy would seem to be to sample as widely as possible with respect to genetic affiliation, in order to maximize our chances of covering the entire space of articulatory possibilities.

Phonological resemblances. It hardly seems worthwhile to include languages with what seem to be similar inventories of vowels, e.g., closely related languages. Such languages are not likely to increase our coverage of the space of articulatory postures (though they might well increase the stability of our results). Rather, it seems prudent to choose languages with as many phonological and phonetic species as possible.

Constructing the model

With these considerations in mind, I have assembled an initial set of data from three genetically diverse languages: Akan, Chinese, and French. These languages are known to differ phonologicaly in several ways. Akan has a phonological contrast argued by Lindau (1979) to be best described by a feature [advanced tongue root] (or [ATR]) that the other languages do not have. Chinese and French have distinctive lip-rounding (front vowels occur in both rounded and unrounded varieties), but Akan does not. Furthermore, Linker's (1982) results suggest that their might be some differences in the articulation of rounded vowels in these languages. Akan and French have distinctive nasalization on vowels, and Chinese does not.

Akan

Akan is a Niger-Kordofanian language spoken in Ghana. The Akan data comprises x-ray tracings of nine vowels as produced by two male speakers, and eight vowels as produced by two female speakers. The data were provided to me by Mona Lindau, and have been described in Lindau (1979) & (1986). The tracings themselves are based on cinéradiographic data collected by Lindau.

The oral vowels of Akan are often written as /**i, ɪ, e, ɛ, a, ɔ, o, ʊ, u**/, though this is not an accurate symbolization of all the phonetic qualities involved. The vowels /**i, e, o, u**/ are usually described as [+advanced tongue root]; the vowels /**ɪ, ɛ, a, ɔ, ʊ**/ are [-advanced tongue root]. Fig. 3.6, after Lindau (1979), shows the position of the tongue and other articulators in some Akan front oral vowels. No x-ray data on the nasalized vowels of Akan is available.

Chinese

'Chinese' comprises a large group of related languages of the Sino-Tibetan family. Here, however, we are concerned with Standard Chinese, which is also sometimes called called Pekingese or Putonghua. The phonemic vowels of Standard Chinese are /**i, y, ə, a, u**/. Ohnesorg & Svarny (1955) provide tracings of still x-rays of these phonemic vowels as

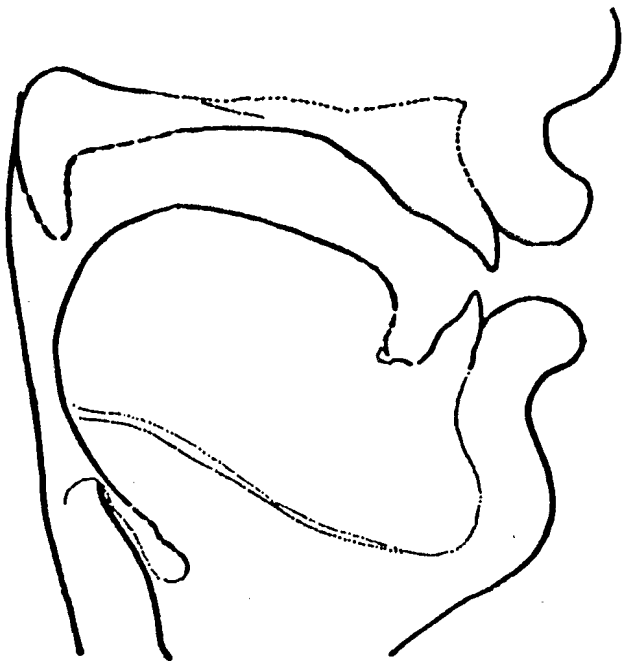


Figure 3.7. The Chinese vowel /ɘ/ as produced by Ohnesorg & Svarny's (1955) speaker B.

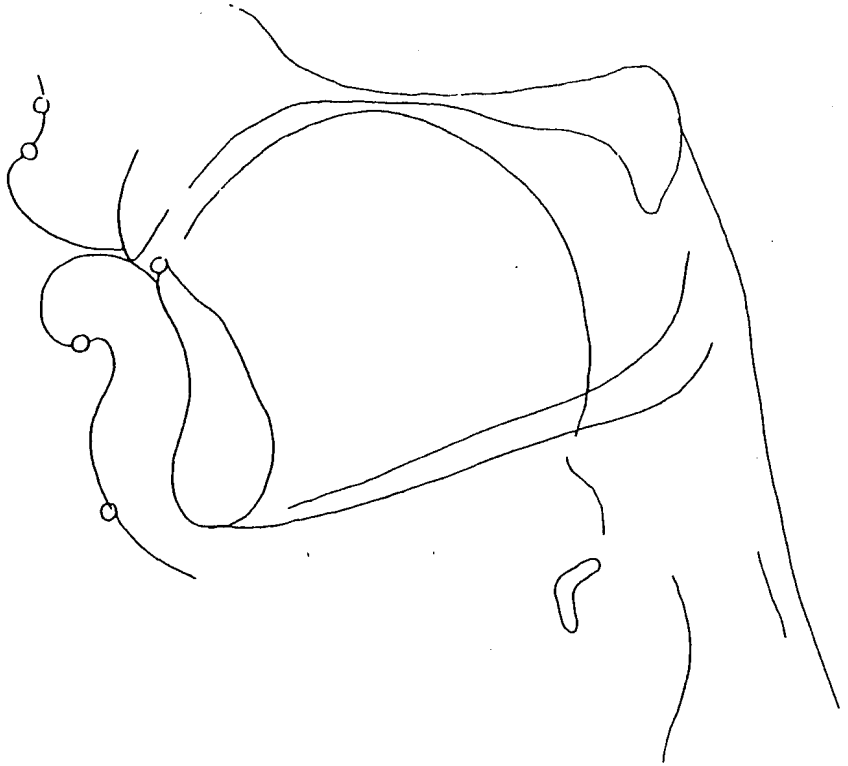


Figure 3.8. The vowel /y/ as produced by Bothorél et al.'s (1986) speaker 1.

produced by several speakers of various dialects of Chinese; I have used their speakers A (Beijing dialect) and B (Tianjin dialect). These 'dialects' are in fact spoken in close proximity (Beijing and Tianjin are only about 150 km apart) and are phonologically very similar. Zhou & Wu (1963) provide x-ray tracings of the vowels [**i, ɛ, y, ə, a, ɔ, u**] as produced by one female speaker of Standard Chinese. None of these tracings includes information about the position of the larynx. Another source, Bao & Yang (1986), is a video-tape of a cineradiographic recording of the sounds of Standard Chinese as produced in natural speech by two speakers. The video-tape includes all the phonetic vowels. It was viewed but not traced due to the technical difficulties of producing high-quality still frames from videotapes. Fig. 3.7 shows a tracing of the vowel /ə/ as produced by Ohnesorg & Svarny's speaker A.

French

French is an Indo-European language spoken in Western Europe. Although it is a Romance language descended from Latin, its vowel inventory shows clear Germanic influences. French has (i) a large number of vowels, (ii) front rounded vowels, (iii) tense/lax distinctions, and (iv) diphthongs (or glide-vowel clusters), all of which are typical features of Germanic phonologies, but not Romance. Bothorel et al. (1986) provide tracings of fourteen phonemic vowels of French as produced by four speakers described as being without regional accents. The tracings are from cinéradiographic films made by the Institute of Phonetics at the University of Strasbourg.

The fourteen vowels represented in the data are /**i, e, ɛ, y, ø, œ, u, o, ɔ, a, ɛ̃, œ̃, ɔ̃, ɑ̃**/. As is well known, many speakers of French do not actually have the vowel [**œ̃**] in their inventories any more; some of these speakers produced a vowel closer to [**ɛ̃**] instead. Fig. 3.8 shows a tracing of the vowel /**y**/ as produced by the first speaker in Bothorel et al.

In all cases, I have used a tracing that the researchers cited above have selected as being representative of the vowel. In the case of the tracings based on ciné x-rays (e.g., Lindau 1979) this is generally a tracing of the articulatory extreme of the vowel in a relatively neutral context.

Validating and extending the model

After having applied the procedures suggested in Chapter 2 to the articulatory data described above to produce an articulatory model, how is one to assess the success of this model and its validity for other languages? Can we tell if the model contains so many free parameters that it is unfalsifiable? What is the number of factors or articulatory primes attested to in the data? The explanatory validity of this study depends on the answers to these questions.

It is most convenient to answer these questions by beginning with what is perhaps the

most technical and also the most bedeviling. The problem of determining what the correct number of factors in a particular data set is has perplexed factor analysts since the inception of the technique (see, e.g. Gould 1981, pp. 234-320). We have already referred above to some of the commoner heuristics, such as only retaining factors that account for 'substantial' variance.

With this kind of articulatory data, we can appeal to external criteria to set an upper bound on the possible number of factors. For instance, Hashimoto & Suga (1986) model tongue shapes as a function of the tensions in thirteen muscles. Linker's (1982) results suggest that there are perhaps three factors involved in lip positions. We might add a few more degrees of freedom associated with jaw, larynx, and velum position. Thus, the combination of anatomical knowledge and previous modeling results suggests that the absolutely highest number of factors plausible in a model of this sort is about twenty. I hasten to add that I do not think that even half this number of factors is likely.

A lower bound is suggested by the phonological features that have been used to classify vowels. Whether or not the features have simple articulatory correlates, they can still be used to roughly evaluate the number of degrees of freedom necessary for the linguistic classification of vowels. The traditional inventory for vowel qualities includes high/low, front/back, tense/lax, advanced/retracted tongue root, and rounded/unrounded. Adding nasal/oral to these gives a total of six or seven phonological degrees of freedom to vowel production. There are a few more contrasts, not usually thought of a phonological but nonetheless needed to describe the vowels of various languages: protruded/non-protruded lips, bunched/unbunched tongue, and retroflex/non-retroflex.

However, some of these phonological features may not be associated with distinct articulatory implementations. For instance, it has been suggested (Lindau 1978) that the tense/lax phonological classification is phonetically a matter of peripherality - a tense front high vowel is simply fronter and higher than a lax front high vowel; a tense back high vowel is backer and higher than a lax back high vowel. If this kind of nested phonological categorization of articulatory phonetic continua occurs, then it is possible for the number of articulatory dimensions needed to described vowel articulation to be smaller than the number of phonological categories.

A good method for determining the number of factors that is reliably determinable from a given data set uses cross-validation. The essence of cross-validation is to compare a model derived from one sample to a model derived from an independent sample. At the point at which factors of the model derived from one sample cease to reflect systematic variation in the data (which we assume to be present in all samples) and rather reflect sample-specific noise, the factors should stop corresponding to those derived from

the other sample.

I will use this principle to evaluate the validity of various articulatory models. Both PARAFAC and EQS allow some factors in a model to be fixed, and other parameters of the model to be estimated. In the PARAFAC program, for instance, the loadings of the factors on the articulatory measurements (the elements of the matrix **A**) can be specified, and the PARAFAC program will estimate appropriate speaker-dependent terms (the $W^2_{k(l)}$ terms). PARAFAC then calculates several measures of fit (stress and R^2) that allow assessment of the model's sufficiency. EQS allows similar procedures with greater fit-testing flexibility.

Following a procedure suggested to me by Eric Holman, I will begin by using PARAFAC in an *exploratory* manner to construct models with varying numbers of factors based on half of the speakers of each language. I will then proceed with a *confirmatory* study to evaluate their validity by fitting the models to the other speakers of the same languages. At the point at which the factors derived from the first half cease to be reliable, the fit to the data from the other half should decline or at least stop increasing. Exploration and confirmation of the articulatory model based on Akan, Chinese, and French are the topics of Chapter 4.

For convenience, I distinguish another kind of cross-validation which I will call *extension*. Extension is testing the model against articulatory data from an independent sample of languages. It thus allows us to assess the generalizability of the model. Extension of the model to Icelandic, Spanish, and Swedish is the topic of Chapter 5.

Extension of the model to an independent sample of articulatory data is necessary to control for the language-specific organization of articulatory primes into coordinative structures. With only three languages in our initial sample, a coordinative structure could occur be shared by all the languages in the sample. We could not distinguish such a (in principle, language-specific) coordinative structure from a (universal) articulatory prime. Extending the articulatory model to other languages provides an opportunity to detect such cases.

Icelandic

Icelandic is a Northern Germanic language spoken in Iceland. The Icelandic data analyzed here consists of 16 vowels as produced by one female (Pétursson 1974a) and one male speaker (Pétursson 1974b). The vowels are often phonemically written as /**i, iː, ɪ, ɪː, e, eː, y, yː, ø, øː, u, uː, o, oː, a, aː**/. Phonetically, these vowels are generally considered to be [**i, iː, ɪ, ɪː, ɛ** or **æ, eː, ʏ, ʏː, œ, øː, ɔ, uː, ɔ, oː, ɑ, ɑː**], respectively. The tracings I have used are the ones selected by Pétursson from his cineradigraphic films as

typical of the vowels. He decribes the selected frames as showing the 'culminating phase' of the vowel, i.e., the part of the vowel that shows the least coarticulatory effects of the surrounding consonants.

In Icelandic (as in many Germanic languages), the long vowels (e.g., /**iː**/) may only occur in open syllables, and only short vowels (e.g., /**i**/) may occur in closed syllables. As is evident from the list of phonetic qualities above, the length, or quantity, differences are confounded with quality differences among the non-high vowels. Earlier PARAFAC analyses of this Icelandic data (Jackson 1988) have shown that at least two, and possibly three, dimensions underly the space of tongue positions used by these speakers.

Spanish

Spanish is a Romance language spoken originally on the Iberian peninsula, but also widely spoken in North and South America. It is usually described as having five phonemic vowels, /**i, e, a, o, u**/, but some phoneticians (e.g., Navarro-Tomás 1968) have described as many as ten 'matices', or allophones, of the basic phonetic qualities of these vowels. Other phoneticians (e.g., Josselyn 1907) describe several more allophonic variants, including front rounded vowels.

The modern consensus appears to be that Spanish as spoken in and around the Madrid area has seven major phonetic vowels, namely [**i, e, ɛ, a, ɔ, o, u**]. Some authors (e.g. Malmberg 1963) also recognize [**æ**] as a major allophone of /**a**/ before palatal and palatoalveolar consonants. The allophones [**ɛ, ɔ**] occur in closed syllables before [**r, l, n**]; [**e, o**] occur in open syllables; and the remainder occur freely.

I have used three sources that provide x-ray tracings of all seven of these phonetic vowel qualities: Holbrook & Carmody (1937), Navarro-Tomás (1916), and Parmenter & Treviño (1932). In addition, two other sources provide tracings of the five phonemic vowels: Quilis (1981) and Russell (1929-30). I believe that all of these authors trained their speakers to prolong the vowel of interest in a real word and then took still x-rays. Cineradiographic techniques were not available at the time that much of this work was done. A sixth source, Malmberg (1963) was consulted but not used because it appeared that the x-rays had been taken with the speaker in a position (lying on his back) which resulted in articulations rather different from those of the other speakers.

Swedish

Swedish is a North Germanic language. Like Icelandic, it has a system of long and short vowels where quantity is partially confounded with quality. The long vowels, which are all that I have used in this study, are usually described as having the phonetic qualities [**iː, eː, æː, yː, ʉː, øː, uː, oː, ɑː**].

My major source for the x-ray tracings used in this study is the data described in

Sundberg (1969). Copies of the tracings of all three speakers were provided to me by Johan Sundberg (only one speaker is illustrated in the article). Only the spoken vowels from this comparative study of spoken and sung vowels were used. In addition, I have used a selection of the tracings of Swedish vowels reproduced in Fant (1965). (Fant's article includes tracings of the short vowels in Swedish as well as tracings of the long vowels).

Chapter 4: Constructing the model

PARAFAC analyses

This section discusses the results of applying PARAFAC to both raw within-language measurements of articulator position and cross-language data. Obtaining a PARAFAC model of a given set of data is generally a multistep process. Proceeding in a stepwise manner, one-, two-, three-, etc., factor analyses are obtained until various diagnostics (described below) indicate that too many factors have been extracted. A given step consists of obtaining several PARAFAC analyses using different random starting positions with the given number of factors. The convergence, uniqueness, and generalizability of the solutions thus obtained are then verified. Finally, after solutions at a range of dimensionalities are obtained, comparision of the fit values provides a basis for selecting the correct solution for interpretation. A brief discussion (drawn from Harshman & Lundy 1984b and Harshman & De Sarbo 1984) of the criteria for judging the optimality of PARAFAC solutions follows.

Convergence

The PARAFAC program (Lundy & Harshman 1985) uses an iterative alternating least-squares (ALS) procedure to optimize random initial estimates (see Carroll & Pruzansky 1984 and references therein for a discussion of ALS algorithms). The iterative procedure is considered to have converged when all the factors change less than 0.1% between iterations. In the analyses reported below, the PARAFAC program terminated if it did not converge within 400 iterations; these solutions are unreliable.

Uniqueness

If all of the PARAFAC runs at a particular dimensionality (number of factors) converge to the same solution, despite their different random initial estimates, then there is a high probability of the solution being a unique optimum. For instance, if six sets of random initial estimates all converge to the same solution, then the probability of another solution having the same goodness of fit (i.e., equal optimality) is less than one in 2^6. Non-uniqueness may occur when PARAFAC is trying to fit non-systematic or random variation. Non-uniqueness at a particular dimensionality often indicates that too many factors have been extracted.

Degeneracy

After verfication of the convergence and uniqueness of a PARAFAC model at a particular dimensionality, the solution is checked for a condition known as *degeneracy*. Degenerate solutions are those in which two or more factors differ only trivially in their description of all aspects of the data (i.e., with respect to speakers, vowels, and measurements of tongue position). Usually, the factors that differ only trivially are fitting

slightly different patterns of subject-specific noise along with "real" factors. The contributions of the factors to the data may balloon to extremely large values, but one factor will cancel most of the other factor's contribution.

Degeneracy is usually an indication that too many factors have been extracted from the data. When this occurs, each additional factor produces only small increases in the fit of the solution. However, if degenerate solutions occur at low dimensionalities (e.g., two or three factors) and the degenerate factors appear to fit substantial amounts of variance, it can also indicate that the data are not well-suited to the PARAFAC model. In this case, recourse to some other model, or the imposition of orthogonality constraints on PARAFAC is indicated (Harshman & Lundy 1984b, p. 271 ff.).

Generalizability

Ideally, the strongest confirmation of a particular set of factors is to show that the same factors recurr in split-halves of the data. If a factor is found to replicate across split-halves, this provides strong evidence that the factor is the result of systematic effects generalizable across samples of subjects. Split-half analyses across subjects generally require a reasonably large sample of subjects, and are not practical in this study. Other resampling procedures, such as jackknifing (Mosteller & Tukey 1977, Chapter 8) are of limited applicability because of the non-independence of the jackknifed samples.

Measures of goodness-of-fit

Two measures of fit are commonly used with PARAFAC. For strictly metric data - such as our raw measurements in millimeters - R^2 is an appropriate measure. It can, moreover, be interpreted in the usual way as the proportion of variance in the raw data that is fit by the model. However, when dealing with other kinds of data, such as covariances, *stress*, defined as the square root of the ratio of the sum of squared errors to the sum of squares (i.e., root(SSE/SS); Lundy & Harshman 1985) is a more appropriate measure of fit. A low stress indicates a good fit; many factor analysts consider stress values of less than 0.1 to be good.

Selection of the best PARAFAC solution from the set of converged, unique, and non-degenerate ones is often on the basis of fit. A solution may well be converged, unique, and non-degenerate, but nonetheless include factors that account mostly for noise specific to one or a few conditions. Such factors typically increase the fit value very little, since they do not account for systematic behavior across the entire data set. Thus, examination of the fit values may allow rejection of a higher-dimensionality solution in favour of a lower-dimensionality one if their fit values are comparable. This use of the fit values is entirely comparable to the scree test of ordinary principal component or principal factor analysis.

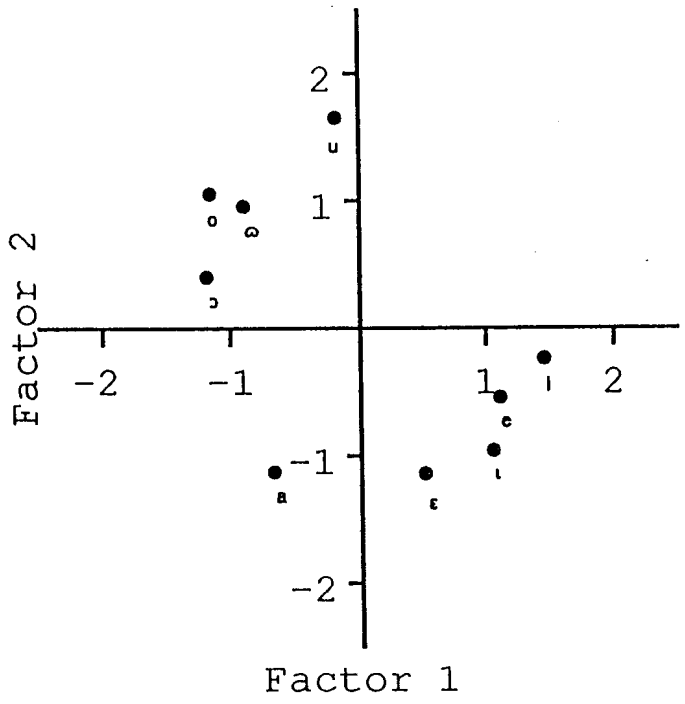


Figure 4.1. Akan 2-factor vowel space.

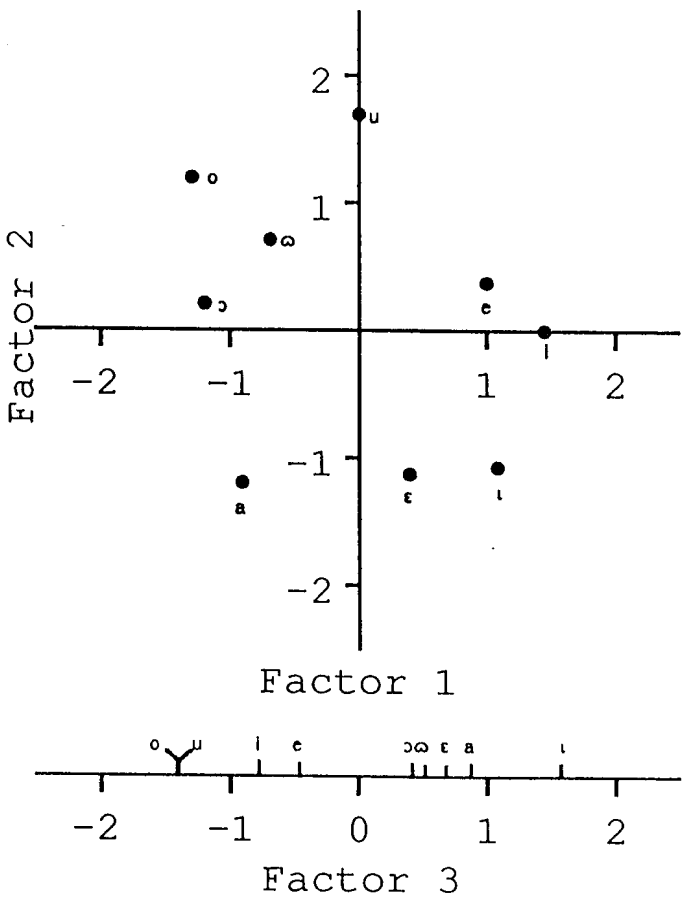


Figure 4.2. Akan 3-factor vowel space.

Non-convergence, non-uniqueness, degeneracy, and poor fit are all reasons for rejecting PARAFAC analyses at particular dimensionalities. In the case of degenerate or near-degenerate analyses that appear to be otherwise acceptable, we will discuss orthogonally constrained solutions.

Akan

The Akan data set includes 1467 data points (4 speakers x 9 vowels x 44 measures - 117 missing data points). Most of the missing data points are due to the fact that tracings were not available for the vowel /**a**/ as produced by two of the Akan speakers. The measurement grids for the Akan speakers were drawn using the tracings of the vowel /ɛ/.

Table 4.1 summarizes the fit figures and other information from the various analyses. A value for R^2 ending in "+" for a set of analyses which did not converge indicates an approximate value; since the analysis has not converged, its fit value is only provisional. As can be seen, it appears that there are two easily determinable factors in the Akan data. At three and higher dimensions, the PARAFAC program has some trouble converging to an optimum. Nonetheless, at three dimensions there is a unique optimum.

Table 4.1: PARAFAC fits for the Akan data set

r	R^2	converged	unique	degenerate
1	0.6617	yes	yes	-
2	0.8010	yes	yes	no
3	0.8376	2/3	yes	maybe
4	0.876+	no	no	yes
5	0.893+	no	no	yes
6	0.908+	no	no	yes

Fig. 4.1 plots the vowel space of Akan according to the coordinate system derived from the PARAFAC analysis at two dimensions. The vowel space roughly resembles the vowel space for English reported in Harshman et al. (1977; see Fig. 1.14). The most obvious differences are for the vowels conventionally written /ɪ, ɛ, **a**, ɔ/, i.e., most of the [-advanced tongue root] vowels. Since these vowel qualities are arguably phonetically different from the English vowels written with the same symbols, these differences in the analysis should not surprise us.

Fig. 4.2 plots the vowel space of Akan according to the converged PARAFAC analyses at three dimensions. The projection of the vowels onto the Factor 1-Factor 2 plane (Fig. 4.2a) is very much like Fig. 4.1. Fig. 4.2b shows the vowels's projection onto the Factor 3 axis. The figure shows that Factor 3 cleanly separates the [-ATR] vowels from the [+ATR] vowels. In other words, this articulatory factor provides a quantitative

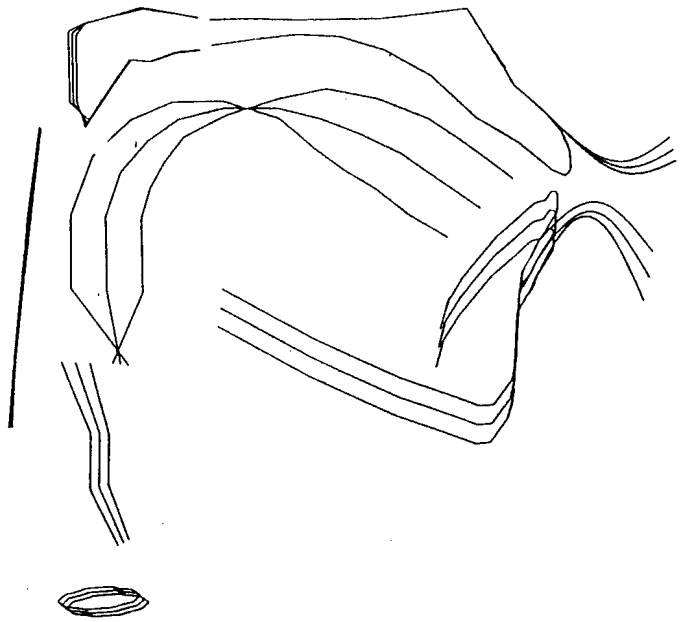


Figure 4.3a. Articulatory displacements generated by Akan factor 1.

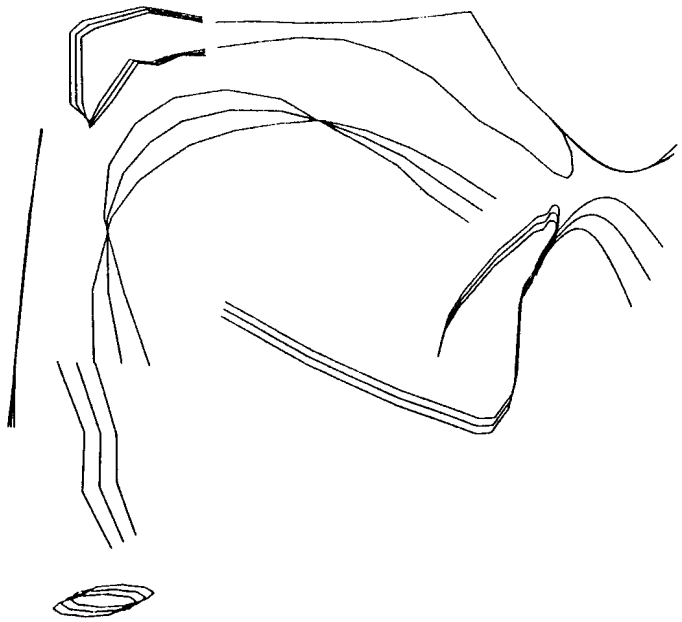


Figure 4.3b. Articulatory displacements generated by Akan factor 2.

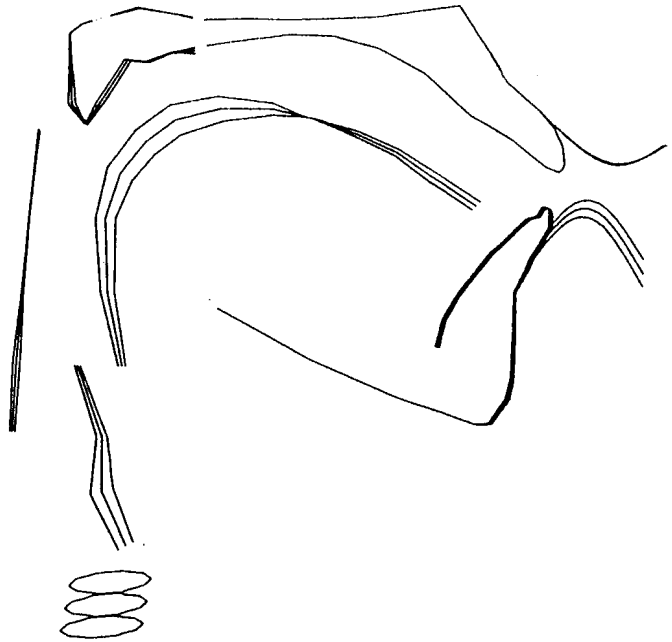


Figure 4.3c. Articulatory displacements generated by Akan factor 3.

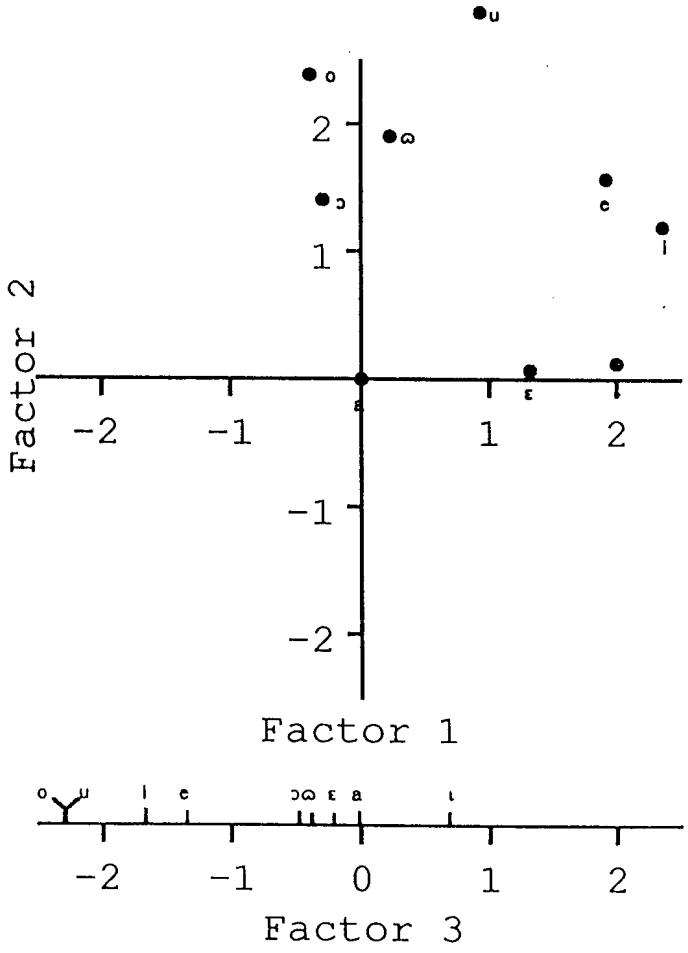


Figure 4.4. Akan 3-factor vowel space translated to put /a/ at the origin.

description of the phonological contrast descriptively known as "advanced tongue root". However, this factor is correlated with Factor 2, as is shown by the fact that the left-to-right order of the vowels on the Factor 3 axis is similar to the low-to-high order of the vowels on the Factor 2 axis. This kind of correlation between two factors produces numerical difficulties for the algorithm used to fit the PARAFAC model because it makes the solution almost degenerate.

This near degeneracy, and the associated difficulty in distinguishing the second and the third factors, may account for Lindau (1986)'s finding only two factors in tongue position data in Akan. However, this study includes additional measures of articulator position and Lindau did not include the vowel /**a**/ in her analysis, so there are several differences between the studies. The measures of larynx and epiglottis position in particular appear to strongly mark the third factor.

Harshman (personal communication) recommends orthogonality constraints in the vowel space in order to accelerate convergence to a solution in cases such as these. Three further runs from random initial positions constraining the three factors to be orthogonal in the vowel space converged to a unique solution very similar to the one obtained from the unconstrained runs. By contrast, the orthogonally constrained four-factor solutions did not converge to a solution similar to the (unconverged) unconstrained solutions. I therefore believe that the three-dimensional solution obtained is a genuine one.

The modes of articulator displacement associated with the three factors are plotted in Fig. 4.3. Factor 1 generates a range of tongue positions that is similar to the Front-Raising factor found by Harshman et al. (1977; Fig. 1.13), though it includes somewhat more displacement of the lower pharynx than Front-Raising. This confirms the findings of Lindau (1986)'s analysis of the eight non-low vowels of Akan. In addition, this factor generates larynx lowering, and jaw raising together with lip compression.

The range of tongue positions that Factor 2 generates is not very similar to the Back-Raising factor (Fig. 1.13); it rather seems more similar to the Icelandic Tongue Arching factor of Jackson (1988; see Fig. 2.15). Like Factor 1, it generates large displacements of the lower pharynx and larynx. Also associated with the tongue arching is a certain amount of raising of the soft palate. As can be seen in Fig. 4.3, this factor is involved in the production of the back round vowels of Akan; it therefore unsurprisingly also generates lip protrusion and closure that we expect to be associated with rounding.

The third factor's outstanding articulatory correlates are larynx raising and something like Back-Raising of the tongue. Rather surprisingly, this factor also generates quite a bit of lip movement. However, examination of the vowel loadings reveals that the

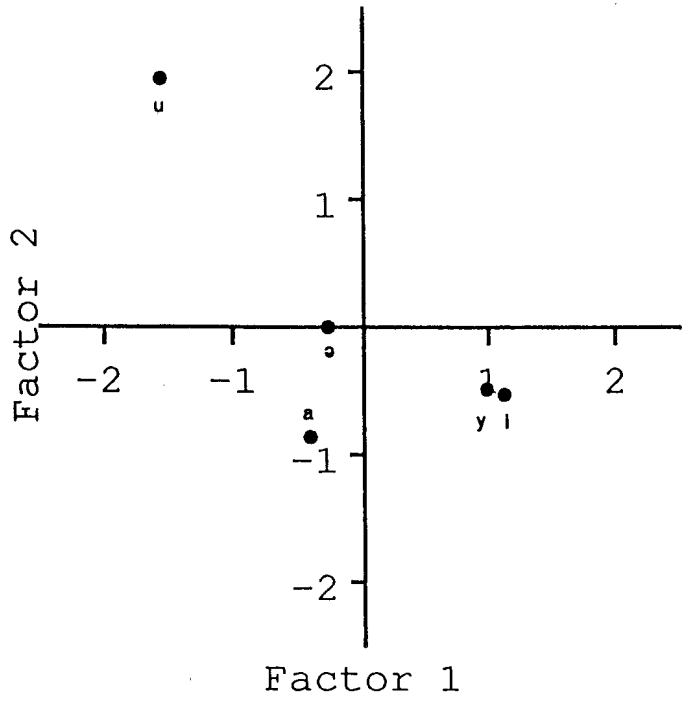


Figure 4.5. Chinese 2-factor vowel space.

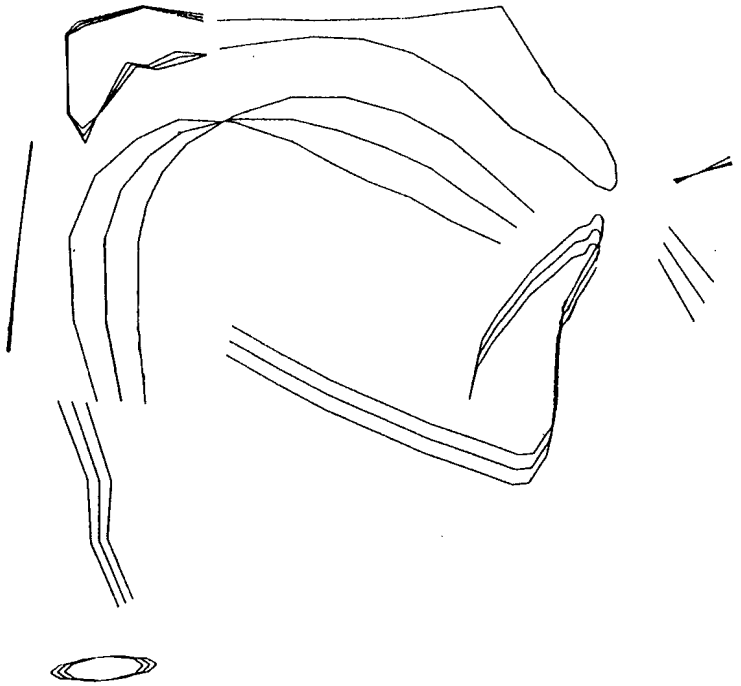


Figure 4.6a. Articulatory displacements generated by Chinese factor 1.

vowel loadings of Factor 2 and Factor 3 are negatively correlated. Thus, when Factor 2 contributes large displacements of the tongue and lips from their mean positions to a particular vowel, Factor 3 tends to cancel the excess, leaving behind reasonable tongue and lip positions.

However, this interpretation is somewhat dependent upon our choice of the mean position as a reference. If we had chosen the low vowel /**a**/ as our reference, then it would appear as though there were two "Back Raising-like" factors at work. This can be seen by translating the loadings plot in Fig. 4.2 so as to place /**a**/ at the origin of the coordinate system for the vowel space, as in Fig. 4.4. Furthermore, these two Back Raising-like factors appear to be mutually incompatible (thus the negative correlation between them). One is an articulatory gesture devoted to the production of [+ATR] back vowels, the other to the production of [-ATR] back vowels. This points up the "holistic" nature of PARAFAC factors, due to the way in which they reflect coordinative structures.

Chinese

The Chinese data set contains 643 measurements (3 speakers x 5 vowels x 44 measures - 17 missing points). Most of the missing points are due to the fact that larynx position is not indicated on any of the x-ray tracings. The grids for the Chinese speakers were constructed using the vocal tract profile for the vowel /ə/.

Table 4.2 summarizes the analyses of the Chinese data; it can be seen from the table that there are probably only two reliably recoverable factors in the data.

Table 4.2: PARAFAC fits for the Chinese data set

r	R^2	converged	unique	degenerate
1	0.5070	yes	yes	-
2	0.7651	yes	yes	no
3	0.814+	no	no	2/3
4	0.877+	1/3	no	yes
5	0.923+	no	no	yes

The vowels of Chinese according to the two-factor solution are plotted in Fig. 4.5 It is worth noting that the high front rounded vowel [**y**] is very close to the high front unrounded vowel [**i**], suggesting that this solution is dominated by tongue position variance. A look at the error analysis suggests this is correct: only about 64.5% of the variance in the lip position measures is accounted for in the model, vs. about 77% overall.

The articulatory displacements generated by the two factors of the solution for the

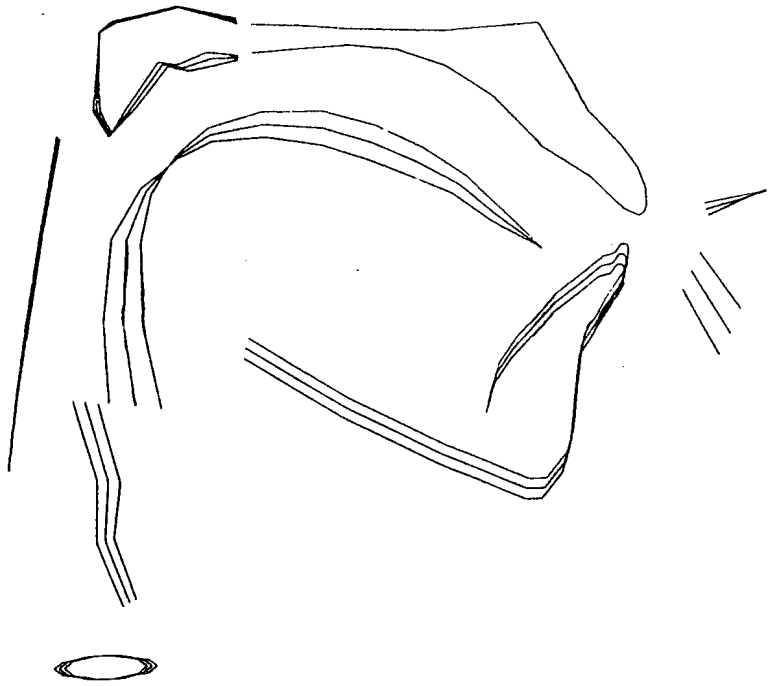


Figure 4.6b. Articulatory displacements generated by Chinese factor 2.

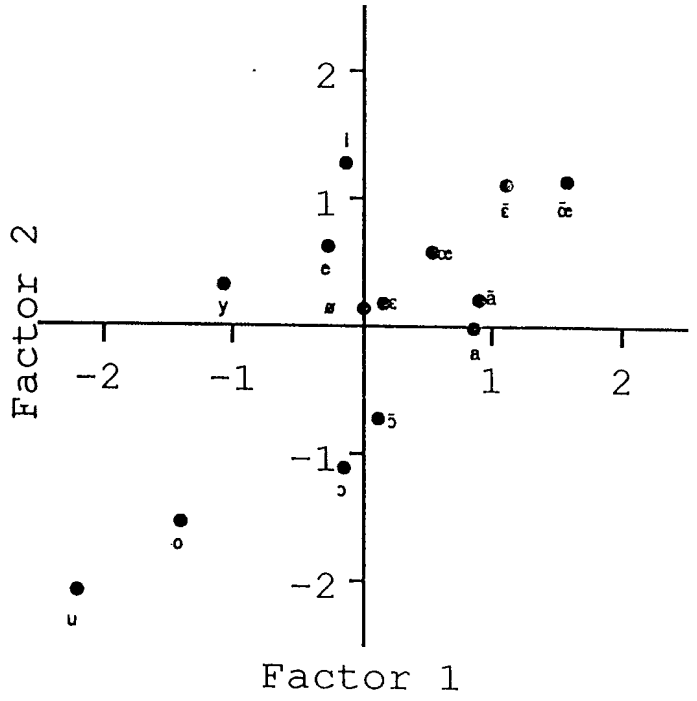


Figure 4.7. French 2-factor vowel space.

Chinese data set are plotted in Fig. 4.6. The tongue positions generated by these factors are very similar to the ones in Harshman et al. (1977). In addition, the Chinese Factor 1 includes jaw raising and apparently passive raising of the lower lip along with the jaw. Chinese Factor 2 also includes jaw raising, but in this case the lower lip raises and protrudes actively. The upper lip also lowers to generate typical rounded postures.

Given that the Chinese data include both rounded and unrounded front vowels ([ʯ] and [i], respectively), it is somewhat surprising that lip rounding does not appear as a separate factor in this stable solution. Lip rounding does, however, appear in the non-degenerate (but apparently also non-unique) three-factor solution. It would seem that the limitation of the vowel sample to only phonemic qualities, excluding the other rounded and unrounded phonetic species in the language, makes it difficult to reliably determine the articulatory organization of rounding in Chinese. By way of comparision, Linker's (1982) study of lip positions in Cantonese vowels used phonetic qualities like [œː] as well as the phonemic /yː/. Linker's study also included eight speakers: this better sample of the phonetic space of Cantonese and of speakers allowed her to reliably determine two different modes of lip-rounding in Cantonese.

French

The French data set contains 2449 measurements (4 speakers x 14 vowels x 44 measures - 15 missing). The missing data points are due to sporadic gaps in the tracing of the velum and the lower pharynx. The grids for the French speakers were constructed using the tracings of the vowel /ɛ/.

Table 4.3 summarizes the analyses of the French data. As can be seen, it is far from obvious what the correct dimensionality is. The two-factor solution is clearly a good one, but it fits the data rather poorly. Although very few of the higher dimensionality solutions converged, there is some evidence that the five-factor solutions all found the same optimum. If the convergence criterion had not been set as low as 0.1%, these solutions would have been considered to have converged.

As all of the R^2s from this dataset are relatively low, it may be the case that PARAFAC is not an appropriate model for this particular data set. This is likely to be due to object variation in the data - e.g., dialectal differences in the vowels that the speakers produced.

Table 4.3: PARAFAC fits for the French data set

r	R^2	converged	unique	degenerate
1	0.4971	yes	yes	-
2	0.6469	yes	yes	no
3	0.714+	no	2/3	maybe
4	0.762+	2/3	no	yes
5	0.8095	no	yes	maybe
6	0.8369	1/3	no	yes

The vowels in the French data set are plotted in the two-factor space in Fig. 4.7. It is worth remarking that, probably because all the nasal vowels of French are low or lower mid, Factor 1 partially confounds vowel height and nasality. The non-nasal counterparts to the nasal vowels do not score as high on Factor 1, despite the fact that these vowels are articulatorily very similar (for all the vowels except [ɛ] which is considerably backed due to the uvular trill [ʀ] that precedes it in the utterance). Indeed, one of the features of the almost-solutions at four and five dimensions is the appearance of (one or more!) factors that distinguish nasalized vowels from the others.

Due to the overall poorness of the two-factor solution, it does not seem worthwhile spending too much time on it. It suffices to note that the lip position measures are not too poorly fit by this solution, even though there is not any single factor that correlates strongly with roundedness. The reason appears to be that the front rounded vowels [y], [ø], and [œ] are not as front as the front unrounded vowels [i], [e], and [ɛ], as can be seen in the plot of the vowel space (Fig. 4.7). This is somewhat surprising, given that French is sometimes cited as a language in which [y] is very high and front. However, an inspection of the data in Bothorel et al. (1986) reveals that this feature of the PARAFAC solution is correct - for at least these French speakers, [y], [ø], and [œ] are not as front or close as [i], [e], and [ɛ]. It is therefore the case that lip rounding and backing are still correlated, and lip rounding therefore does not clearly appear as an independent articulatory gesture.

The tantalizing hint of the existence of a higher-dimensionality solution caused me to try constructing a four-dimensional Procrustean model for the French vowels. I specified loadings for the vowels, and allowed the PARAFAC program to iterate until it had found the best articulator loadings and speaker scaling coefficients with the vowel loadings fixed. The initial guesstimates of the vowel loadings on two factors were arrived at by using the values for Front Raising and Back Raising from the closest comparable vowels in the PARAFAC solution for English (Harshman et al. 1977). The other vowel loadings

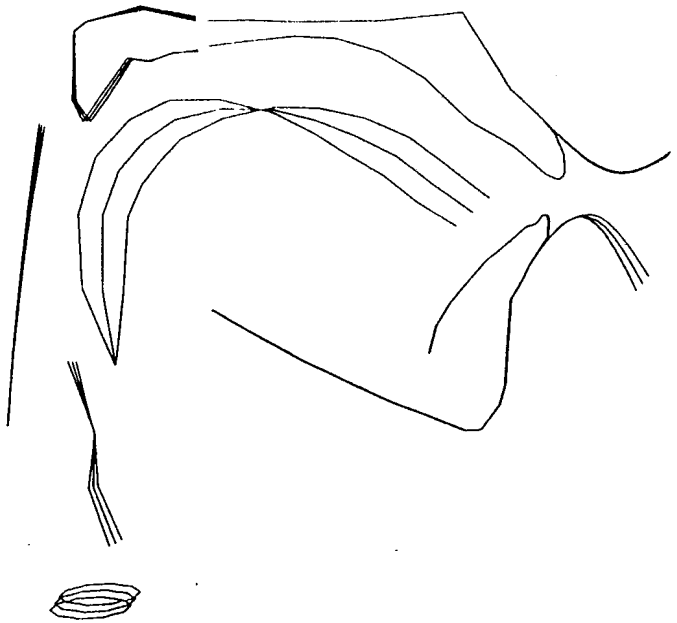


Figure 4.8a. Articulatory displacements generated by the Procrustean Front-Raising factor.

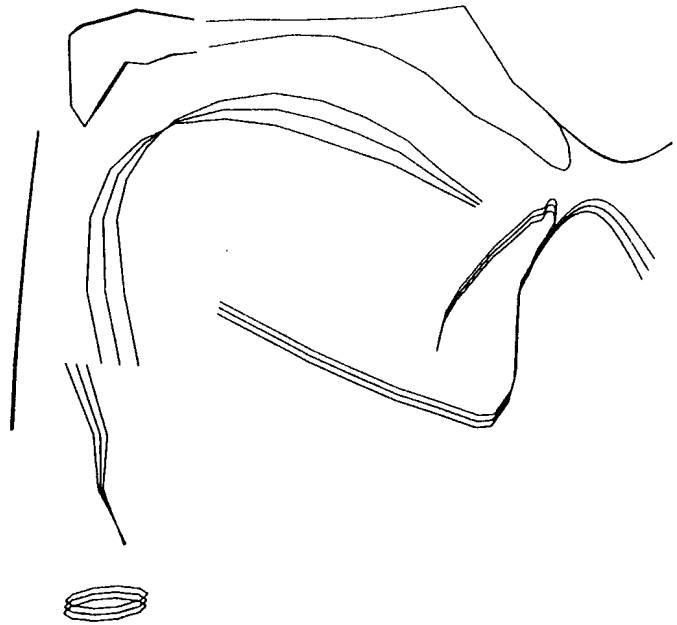


Figure 4.8b. Articulatory displacements generated by the Procrustean Back-Raising factor.

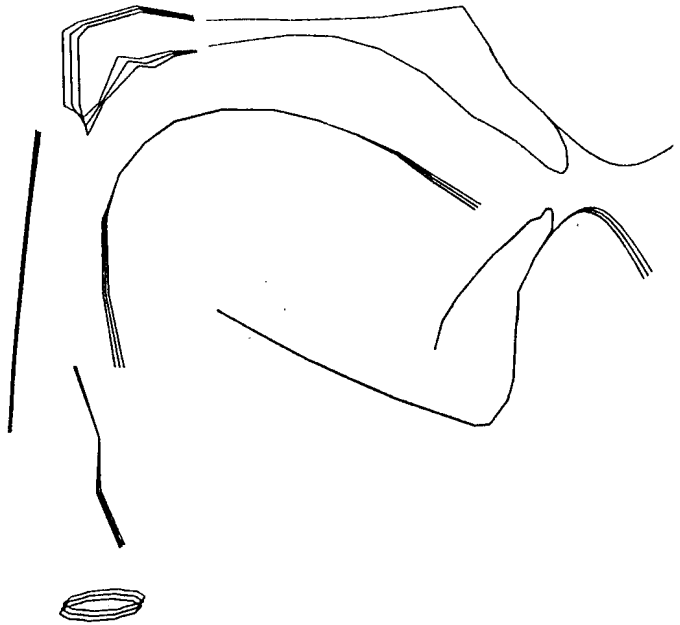


Figure 4.8c. Articulatory displacements generated by the Procrustean Nasal factor.

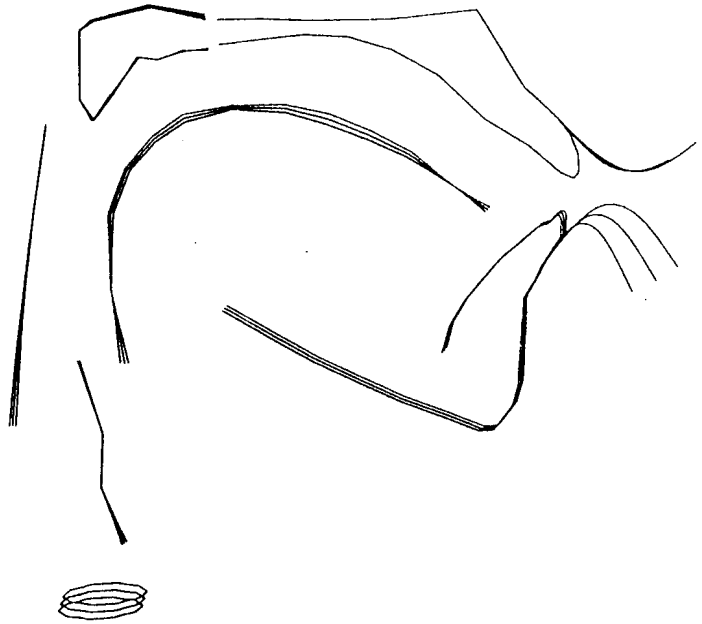


Figure 4.8d. Articulatory displacements generated by the Procrustean Round factor.

were initially binary rounded/unrounded and nasal/nonnasal dichotomies. After the PARAFAC program converged at a solution, the fitting errors were examined by vowel, and the loadings in the vowel mode were changed to improve the fit. A few iterations of this procedure - an afternoon at the computer - produced a solution with an R^2 of 0.6543 - comparable to the R^2 of the unconstrained PARAFAC 2-factor solution. The vowel loadings arrived at by this procedure are given in Table 4.4.

Table 4.4: Vowel loadings for the Procrustean fit to French

vowel	FR	BR	Nasal	Round
i	-1.537	0.9260	0.7483	-1.135
e	-1.024	0.5292	0.7483	-1.135
ɛ	0.0000	0.0000	0.7483	-1.135
y	-1.229	0.7937	0.7483	0.9461
ø	-0.7171	0.2646	0.7483	0.9461
œ	-0.5122	-0.6614	0.7483	0.9461
u	1.024	2.249	0.7483	0.9461
o	1.024	0.7937	0.7483	0.9461
ɔ	1.434	-0.3969	0.7483	0.9461
a	-0.5122	-1.984	0.0000	-1.135
ɛ̃	-0.7171	-0.1323	-1.497	-1.135
œ̃	-0.5122	-0.7937	-1.497	0.0000
ɔ̃	1.434	-0.6614	-1.497	0.9461
ɑ̃	1.024	-0.9260	-1.497	-1.135

One of the vowel loadings that had to be adjusted is worthy of note. The loading of /**a**/ of the Nasal factor was adjusted away from the value for non-nasal to a value intermediate between nasal and non-nasal because the mean squared error on that vowel was nearly twice as large as for any other vowel. Doing so reduced the error to be comparable with the error in other vowels. However, the vowel /**a**/ is not 'partially nasal' due to the adjacency of a nasal consonant - the token that I used in these analyses is the final "a" in the sentence "Il fume son tabac". This partial nasalization of /**a**/ is thus either a positional effect (since it is sentence final) or else intrinsic to the vowel.

The articulatory displacements associated with these factors are shown in Fig. 4.8. As we might expect, the French Front Raising and Back Raising factors nearly identically replicate the patterns of tongue displacement that occur in English. French Front Raising triggers very little jaw or lip movement; French Back Raising, on the other hand, goes along with quite a bit of both. The French Nasal factor produces, as we would expect, a large displacement of the velum. It also produces a small tongue displacement

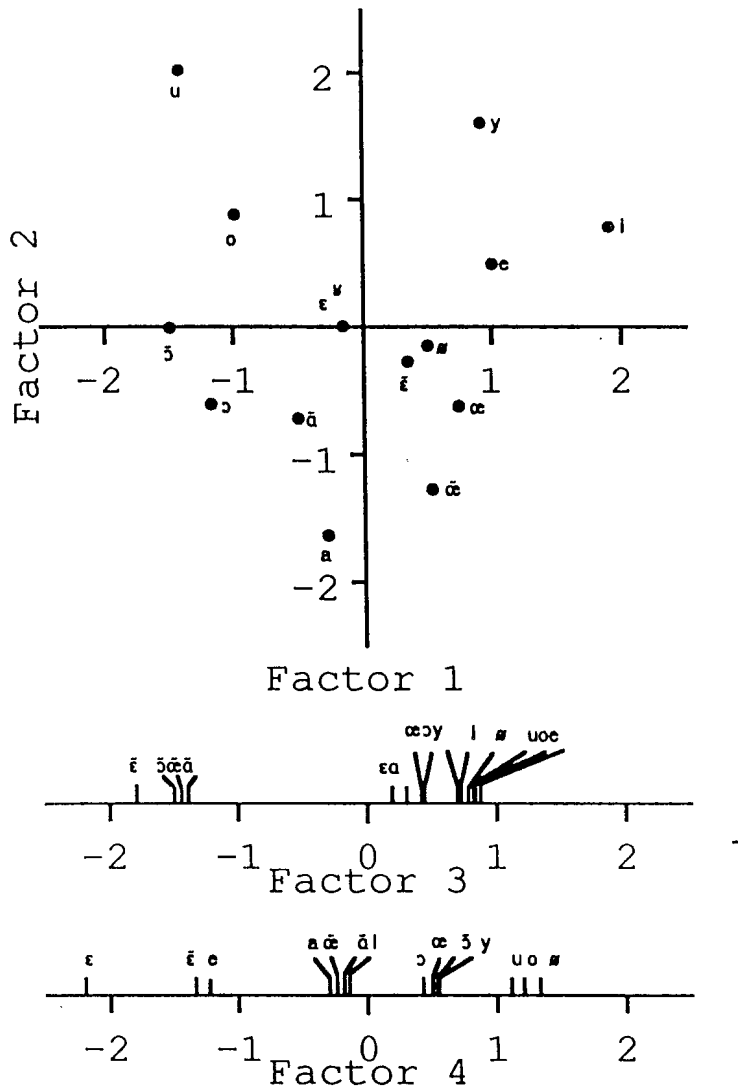


Figure 4.9. The relaxed Procrustean French vowel space.

similar to the Icelandic Tongue Arching factor and some larynx lowering. The Round factor produces even more larynx lowering to go along with lip protrusion and closure. Thus, none of these factors specified in the vowel mode has simple articulatory correlates unique to one articulator.

I then used these articulatory loadings to generated a 'relaxed Procrustean' fit to the data; holding the articulatory loadings constant, the PARAFAC program estimated the best vowel and speaker loadings. The solution thus obtained has an R^2 of 0.7349 - a fit value quite comparable to that of the unconstrained (and unconverged) four-factor solution. The vowel space shown in Fig. 4.9 results - it is essentially the same as that of Table 4.4, except that the dichotomously specified categories of Table 4.4 now have some internal structure. In particular, the vowel loadings on the Nasal factor now ranks the non-nasal vowels in roughly the same order as the Back Raising factor ($r \cong 0.7$) Similarly, the Round factor orders the vowels (within the rounded-unrounded categories) roughly according to their loading on the Front Raising factor. These correlations are of course the reason why unconstrained PARAFAC finds degenerate solutions at the higher dimensions.

The validity of this description of the French vowel space may be assessed in part by comparing it with impressionistic phonetic descriptions, such as that of Armstrong (1967). ('Impressionistic' in the technical sense of not relying on instrumental means). Armstrong (1967) describes the vowels [**i**, **e**, **ɛ**] as front vowels (pp. 35, 39, 42), but the front rounded vowels [**y**, **ø**, **œ**] as "retracted from the true front position" (pp. 59, 62, 64). He also describes the tongue position in the front rounded vowels as being uniformly slightly lower than the tongue position in the corresponding front unrounded vowels. In Fig. 4.9 the vowels [**y**] and [**ø**] are less peripheral than their unrounded counterparts, and thus perhaps could be spoken of as retracted, but it is not clear whether the differences between the front rounded vowels are quite as systematically related to the front unrounded vowels as Armstrong would have them be.

Another point of comparision is Armstrong's (1967) description of [**u**], [**o**], [**ɔ**], and [**ɔ̃**]. [**ɔ**] is "not a true back vowel. It has a distinct central (**ə**) quality in it" (p.50). Similarly, he notes that "The modern tendency in French seems to be to use an **o** which has a tongue position slightly advanced" (p. 52). [**u**] is also described as being "slightly advanced from the true back position" (p.56). On the other hand, he describes [**ɔ̃**] as having a fully back tongue position, intermediate between the Cardinal vowels **ɔ** and **o**. A similar pattern shows up in Fig. 4.9 with respect to the position of the vowels along the Front Raising axis (horizontal in the figure): [**ɔ̃**] has the greatest negative contribution

from Front Raising; [**u**, **o**, **ɔ**] all have smaller negative contributions and plot somewhat closer to zero on the Front Raising scale.

Armstrong (1967) notes that "a vowel intermediate between **ɑ** and **a** ... represented by the symbol **a** ... is used by many French speakers" (p. 49). Since he describes both [**ɑ**] and [**a**] as fully open vowels (p. 48 and 44, respectively), I take it that the intermediate [**a**] is supposed to be fully open (or low) as well. [**ɑ̃**], on the other hand, is described as being higher, i.e., almost half-open (p. 71). Once again, Fig. 4.9 agrees with Armstrong's description, with [**ɑ̃**] plotting substantially closer to the origin of the vowel space than [**a**]. Altogether, Armstrong's (1967) description and the description in Fig. 4.9 seem to agree on many points.

The results of these within-language analyses should emphasize the degree to which within-language analyses reflect language-specific articulatory organization, i.e., coordinative structures. Traditionally, lip-rounding is considered a kind of articulatory gesture that is independent of tongue positioning. Similarly, the movement of the velum required to produce nasal vowels is traditionally regarded as independent of tongue positioning. However, this potential independence is not recoverable from the within-language data analyzed above. Even when we specify some loadings on the basis of phonological expectations (nasal/nonnasal, rounded/unrounded), the articulatory correlates are complex.

To summarize, it appears that Akan allows a three-factor description; Chinese, two (and perhaps three); and French, four. Combined, these results suggest that in a cross-language analysis there ought to be at least four factors (if the *same* factors have shown up in every language), but no more than nine (if *different* factors have been found in each language).

Cross-language covariance analyses

The next step is to find a cross-language model based on the 4559 measurements of articulator position in the vowels of the three languages described above. As discussed in Chapter 3, I first split the speakers of each language randomly into two groups. The first group, split-half A, contains Akan speakers 1&4, Chinese speakers 1&3, and French speakers 3&4. Split-half B contains Akan speakers 2&3, Chinese speaker 2, and French speakers 1&2.

The $N_{measures}$ by $N_{measures}$ covariance matrix for each speaker in split-half A was calculated (eliminating the 'vowels' mode from the data). Values missing from the original data were replaced with estimates based on the within-language solutions discussed above. Since the missing values accounted for only about 3.5% of the data points, these estimates should not unduly bias our results. For the Chinese data, in which

there are no measures of larynx position, all the covariances associated with larynx position were marked as missing values. The resulting covariance matrices were then submitted to the PARAFAC program. As above, three runs were made at each dimensionality, with a limit of 400 iterations for convergence. Following the recommendations in Lundy & Harshman (1985), equal-average diagonalization (EAD) was used to normalize the covariance matrices. EAD brings the average of the covariance matrices to a correlation matrix without destroying the within-speaker proportionalities. This procedure effectively weights all the measurements - both high-variance and low-variance ones - equally. It is hoped that this procedure will facilitate recognition of factors that produce only small ranges of articulator displacement. The loadings on the articulatory measurements from the factors derived from split-half A were then used to fit the covariance matrices from split-half B.

It is worth noting that estimating each factor from split-half A involves quite a few free parameters. For each factor, there are 44 loadings of the factor on the measurements of articulator position, and a parameter estimating the factor's contribution to each speaker's covariance matrix. There are therefore 50 free parameters associated with each factor in split-half A. The number of free parameters per factor is thus about 1/45, or slightly more than 2%, of the number of data points in split-half A.

However, fitting the covariance matrices of split-half B with factor loadings fixed by split-half A requires the estimation of far fewer loadings. The factors' articulatory loadings are fixed, and only the speaker loadings are estimated. There are therefore only five free parameters associated with each factor in the analysis of split-half B. Each factor has less than 0.5% as many free parameters as there data points in split-half B. The fit values for split-half B are therefore much lower; however, due to the extremely small number of free parameters in the model, they are still quite significant.

The results of these PARAFAC analyses at one through nine dimensions are presented in Table 4.5. Table 4.5 does not present any information about the possible degeneracy of the solution because PARAFAC fits to covariance matrices are orthogonal in the mode across which the covariances are calculated (in this case, the vowel mode). Degeneracy is therefore not a concern.

Table 4.5: PARAFAC fits to split halves using normalized covariance matrices

	Split-half A factors			Fit to split-half B
r	R^2	converged	unique	R^2
1	0.2806	yes	yes	0.1321

2	0.4460	yes	yes	0.3290
3	0.6016	yes	yes	0.4004
4	0.6884	yes	yes	0.4664
5	0.7554	no	yes	0.4737
6	0.8013	2/3	yes	0.4717
7	0.8351	2/3	2/3	0.4690
8	0.870+	no	yes	0.4765
9	0.8919	2/3	2/3	0.4844

Table 4.5 shows that the fit to split-half B by factors from split-half A increases a little from the four-factor solution to the five-factor solution. The fit does not increase from the five-factor solution to the six-factor solution. Although none of the five-factor solutions converged, they were all clearly approaching the same optimum, with no loadings differing by more than 0.01. The goodness of fit to split-half B using factors from split-half A then declines slightly at six and seven dimensions, suggesting that these factors modeled sample-specific noise in split-half A. At eight and nine dimensions, the fit values climb a little, but not enough to warrant including extra factors.

The unnormalized covariance matrices for all eleven speakers were then analyzed using the PARAFAC program. Using unnormalized covariance matrices preserves the original scale of the data, thus allowing us to recover realistic patterns of articulator displacement. The fit results for the analyses are summarized in Table 4.6. The fact that the fit values in Table 4.6 are much higher than those in Table 4.5 indicates that the noisiest articulatory measures are also the ones with low variances. If all the measures had equivalent proportions of non-systematic (noise) variance, the fit values in Tables 4.5 and 4.6 would be comparable.

Table 4.6: PARAFAC fits to unnormalized covariance matrices

r	R^2	converged	unique
1	0.7624	yes	yes
2	0.8355	yes	yes
3	0.8806	yes	yes
4	0.9079	yes	yes
5	0.9258	yes	yes
6	0.9394	yes	yes

Examining Table 4.6 reveals the advantage of selecting the number of factors by the split-half cross-validation procedure of Table 4.5. In Table 4.5 the fit to the other split-half stops improving after five factors, whereas in Table 4.6 the fit continues to

improve, albeit slowly.

The cross-validation procedure summarized in Table 4.5 points to there being four or five systematic and recoverable factors in this data set. Although the fit value peaks at five factors, the fifth factor itself adds little to the fit. The five-factor solutions for split A also did not converge within 400 iterations.

Examining the results of fitting the PARAFAC model to all of the normalized covariance matrices suggests that this fifth factor contributes substantially to the overall data set. It may be the case that the particular split of the data used above created subsets in which the variation corresponding to fifth factor was not well instantiated. Table 4.7 shows that the fifth factor contributes some 5% to the overall fit when the EAD-normalized covariance matrices from all eleven speakers are analyzed simultaneously. The five-factor solution is also unique.

Table 4.7: PARAFAC fits to normalized covariance matrices

r	R^2	converged	unique
1	0.3047	yes	yes
2	0.4729	yes	yes
3	0.6116	yes	yes
4	0.7079	yes	yes
5	0.7520	yes	yes
6	0.7800	yes	no
7	0.8046	yes	no
8	0.8293	yes	no

The articulator loadings of the five-factor solution, based on data from Akan, Chinese, and French, are given in the first section of Table 4.8, below. The left-most column of the table lists the abbreviations for the measures of articulator position described in Chapter 3. The next five columns of the table contain the loadings on the measures of articulator position. The rightmost column of the table indicates the proportion of the variance in the data that is accounted for by the factors (variance accounted for: VAF). It can be seen that the measures of tongue position are best described by the model: generally 80% - 90% of the variance in the raw measures is accounted for. The measures of the velum, dorsal wall of the pharynx, and upper epiglottic position are moderately well described by the model. However, the lip, larynx, and jaw position measurements seem to be quite noisy.

The second section of Table 4.8 contains the mean within-language contribution of the factors to the normalized data. These quantities have no simple interpretation, since

EAD normalization removes the data scale; they can only be interpreted relatively.

The patterns of articulator displacement generated by these factors are presented in Figs. 4.10 to 4.15. In these figures, the articulatory displacements are scaled by the factors that were removed during EAD normalization, so as to recover realistic articulatory displacements. Two of the factors have very clear and unique articulatory correlates, the others bear more discussion.

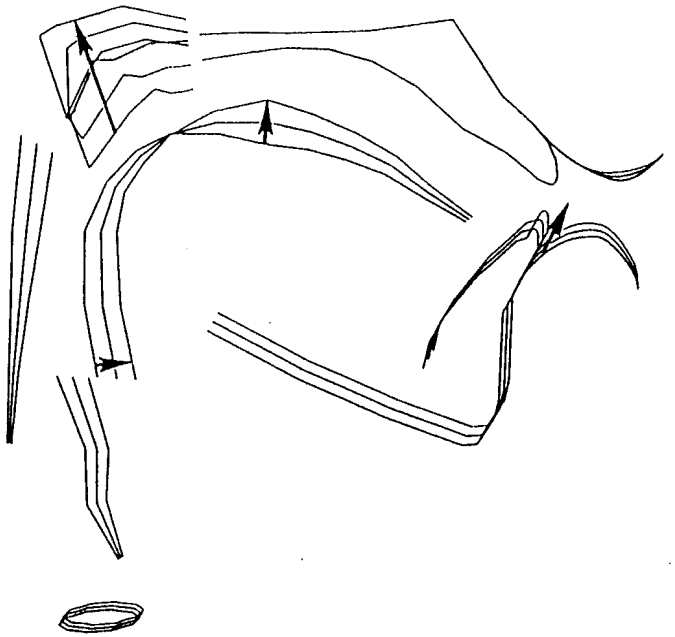


Figure 4.10. Articulatory displacements generated by the first factor from the cross-language covariance analysis.

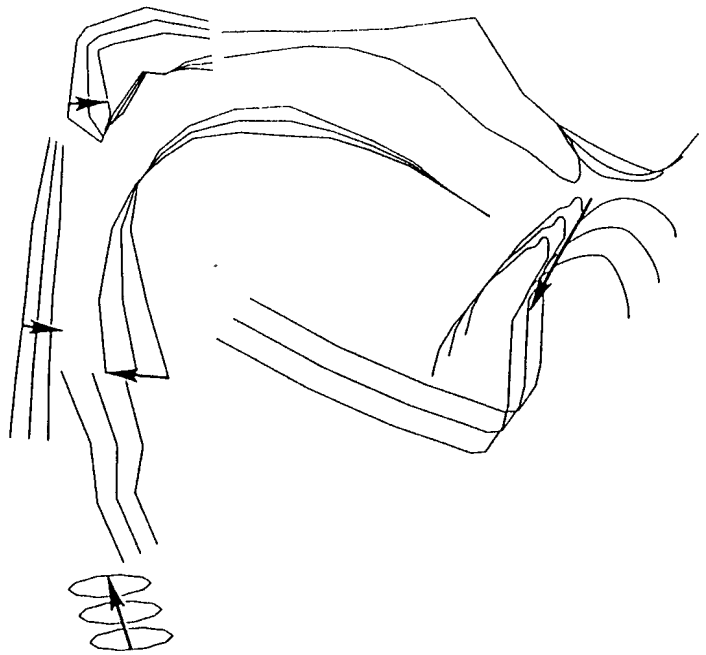


Figure 4.11. Articulatory displacements generated by the second factor from the cross-language covariance analysis.

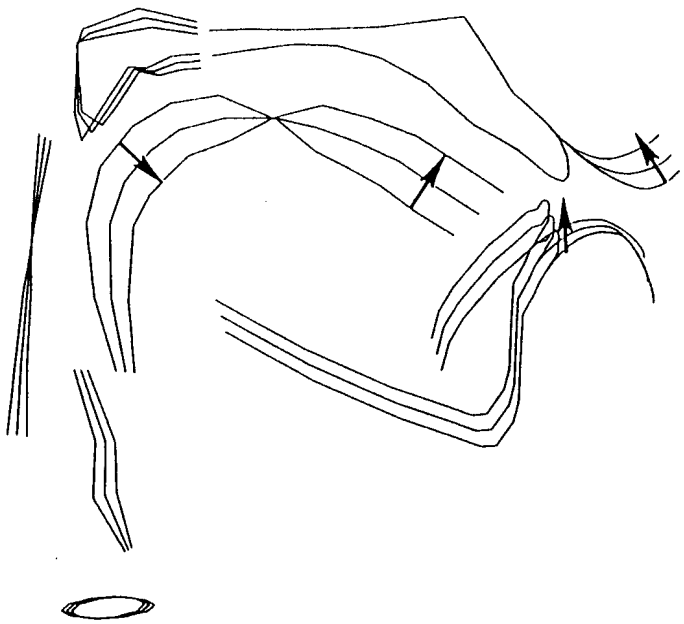


Figure 4.12. Articulatory displacements generated by the third factor from the cross-language covariance analysis.

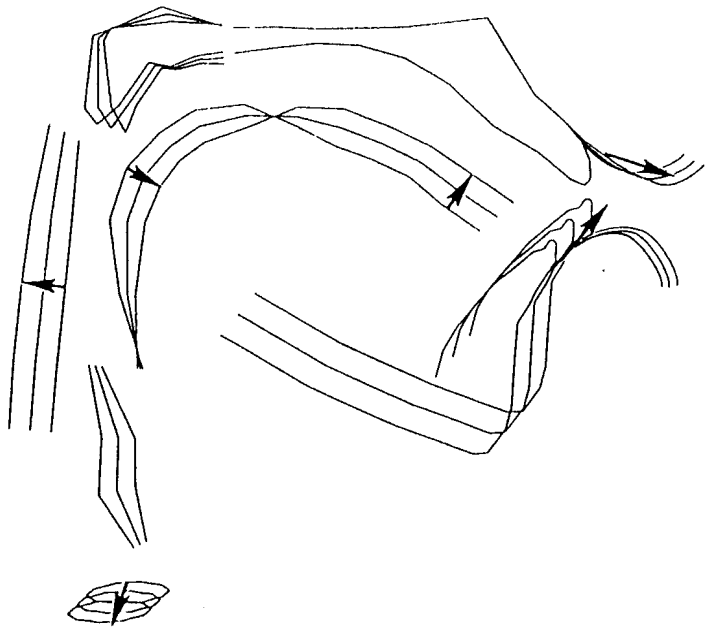


Figure 4.13. Articulatory displacements generated by the fourth factor from the cross-language covariance analysis.

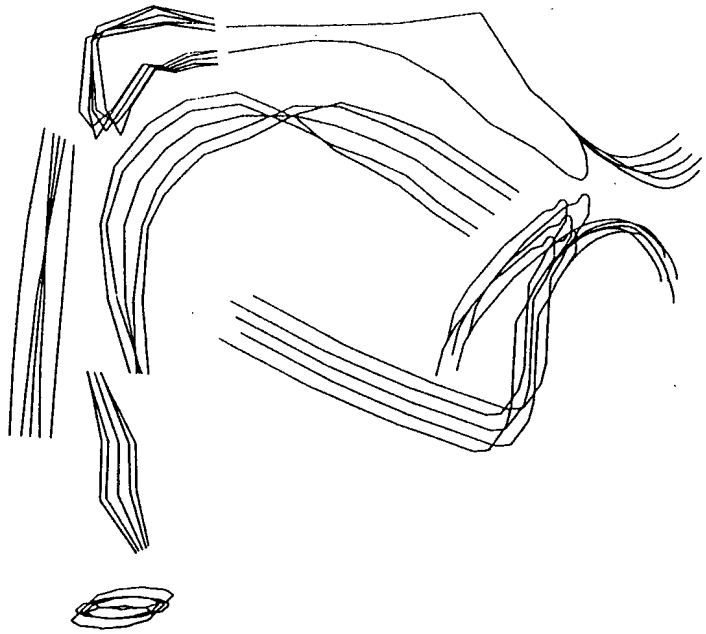


Figure 4.14. A comparision of the families of articulatory positions generated by the third and fourth factors.

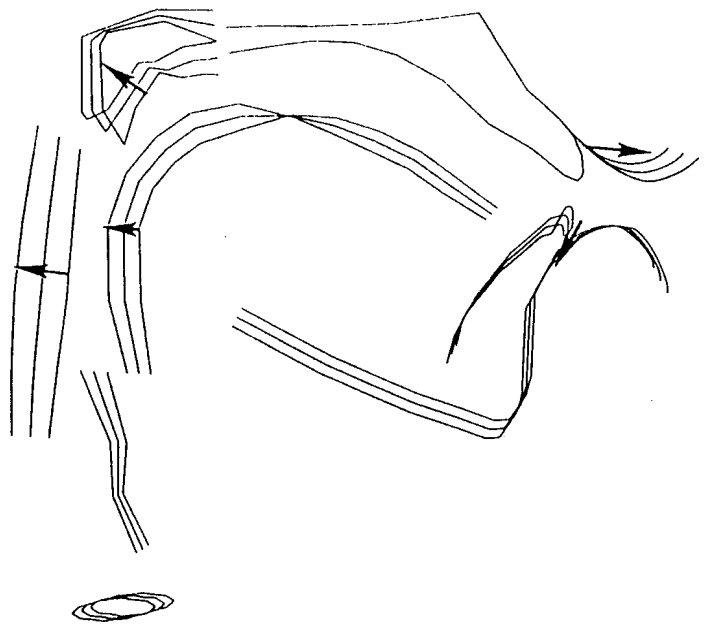


Figure 4.15. Articulatory displacements generated by the fifth factor from the cross-language covariance analysis.

Table 4.8: Scaled loadings from the five-factor PARAFAC simultaneous analysis of Akan, Chinese, and French

Measure	Factor					VAF
	1	2	3	4	5	
e1	-.1740	1.283	-.2595	.4571	.4680	.26
e2	-.8244	1.602	-1.009	1.545	.4115	.54
e3	-.9693	2.177	-.8445	1.402	.6472	.68
e4	-1.632	3.237	-.6851	.6266	1.380	.85
p3	-.1291	-1.063	-.5489	1.216	1.065	.65
p4	-.2198	-.7489	-.2789	1.052	1.087	.67
p5	-.3672	-.5699	-.1496	.9368	1.105	.76
p6	-.5138	-.5011	.0631	.7495	.9036	.78
p7	-.7132	-.2819	.2591	.6433	.9403	.70
t4	-3.647	5.815	-1.752	.4662	2.133	.94
t5	-3.550	4.611	-3.300	-1.522	3.091	.87
t6	-3.064	1.710	-4.457	-3.312	2.939	.83
t7	-1.996	.2302	-5.380	-3.639	2.805	.85
t8	-1.268	-.7436	-5.435	-3.143	2.745	.87
t9	-.1367	-1.055	-4.344	-2.606	2.436	.79
t10	1.323	-1.388	-2.645	-1.298	1.456	.64
t11	2.561	-1.710	.1300	1.119	-.1135	.46
t12	3.173	-1.254	4.179	3.456	-1.461	.78
t13	3.140	-1.032	5.896	4.030	-2.285	.84
t14	2.483	-.6388	6.997	4.214	-2.377	.89
t15	1.704	-.1120	7.647	5.012	-2.312	.95
t16	.9079	.2551	7.702	5.051	-1.977	.92
vi1	2.134	-.3323	-.4777	.5935	.9034	.75
vs1	0.101	-1.189	-.1970	.7490	.5753	.40
vi2	2.415	-.1533	-.4038	.4825	.7945	.79
vs2	2.636	-.9968	.0415	.5211	.8565	.64
vi3	1.839	-.0581	-.3894	.1768	.8587	.79
vs3	2.000	-1.225	-.5930	.0152	.2666	.55
vi4	1.325	.1596	-.3714	.2094	.7526	.68
vs4	1.006	-.7864	-.5606	-.4102	.3498	.52
vi5	.8285	.2486	-.2531	.2053	.5579	.53
vs5	.3172	-.2157	-.1504	.0149	.3704	.28
tvx	-.1956	.1029	.5711	-.7254	-.9062	.33
tvy	2.583	-.5124	.6063	.4417	1.206	.65

ljx	.2710	-.7731	-.1366	.8057	-.2367	.36
ljy	.8255	-1.817	1.555	1.916	-.8167	.66
llx	.1889	-1.691	-.6969	-.7087	.6324	.38
lly	.6330	-2.932	.9179	-.1026	-.1962	.44
ulx	.2320	-1.213	-.7930	-.7770	.5237	.51
uly	-.2124	-.0760	.8604	.3683	-.5251	.23
ild	-.9428	2.868	-.6712	.6850	1.103	.38
lid	-.3257	1.490	-1.727	-.3470	1.201	.52
uid	-.2005	-1.534	-.3337	-.5967	.9093	.36
lyy	.4758	3.422	-.1083	-1.100	-.2466	.32
Akn	.7604	1.265	1.455	1.114	.7974	
Chn	.8072	1.029	.8095	.8472	1.026	
Frn	1.368	.7832	.4840	.7401	.9150	

Fig. 4.10 shows the family of articulatory positions generated by the first factor. This family of articulatory positions includes large downward movements of the velum, together with lowering of the tongue, jaw, and lips. This factor contributes 50% more to the data from the French speakers than it does to the data from the other speakers. Since the data from the French speakers includes phonemically nasal vowels, and the other data does not, this is hardly surprising. Again, the (in principle accidental) correlation between low tongue positions and nasalization produced by lowering the velum shows up, since all the nasal vowels of French are low.

This result is in fact somewhat disappointing, though it has an important lesson. All factor-analytic procedures tend to capitalize on chance correlations; in this case, PARAFAC has capitalized on a non-chance but nonetheless in principle language-specific and non-universal correlation to produce a somewhat more compact description of the data. Without a larger sample of nasalized vowels, it is not possible to tell whether or not velum lowering *necessarily* produces tongue lowering.

Fig. 4.11 shows the family generated by the second factor. This factor generates large movements of the epiglottis and root of the tongue together with anterior movements of the dorsal wall of the pharynx and a considerable amount of larynx raising. There is also some flattening of the blade of the tongue and jaw and lip lowering. However, the tip of the tongue does not move at all. This factor contributes most to the data from Akan. The tongue-root displacments generated by this factor are similar to the expected articulatory correlates of the [+/-ATR] contrast in Akan.

The third factor, pictured in Fig. 4.12, generates a range of articulatory positions that ranges from an [**o**]-like posture with a protruded upper lip to an [**i**]-like posture with retracted and compressed lips. It thus would appear to represent a range of articulatory

postures found when going from a front unrounded vowel to a back rounded one. Factor 3 contributes most to Akan and Chinese.

The fourth factor's articulatory correlates are digrammed in Fig. 4.13. It also generates a range of tongue positions that goes from an [o]-like posture to an [i]-like posture, but unlike the previous factor, it also involves some pharyngeal expansion, and larynx lowering. This is the kind of factor we would expect to contribute to the production of front rounded vowels. There is less lip movement associated with this factor than with the third factor, and it appears to be mostly lip protrusion/retraction without compression. This factor makes substantial contributions to all the speaker's articulatory positions.

The similarities between these factors merit some more comment. The two factors generate very similar families of tongue positions, as can be seen in Fig. 4.14, where their families of articulatory positions are superimposed.

However, these two factors generate different patterns of movement. Factor 3 involves nearly vertical jaw motion and no larynx motion; Factor 4's jaw motion contains a large horizontal component that is probably part of the mechanism for producing the movements of the dorsal wall of the pharynx and substantial larynx motion. These results emphasize that a front rounded vowel is not the same as a front unrounded vowel with lip-rounding superposed. Front rounded and front unrounded vowels do have tongue positions that come from the same or very similar families of tongue positions. But front rounded vowels appear to require coordinated gestures involving the larynx, pharynx, and jaw - as seen in Factor 4 - that are rather different from those required for front unrounded vowels.

Another observation about these factors is that neither of them shows the pattern of pharyngeal expansion seen in the Akan Front Raising-like factor. The three-factor Akan solution has large tongue-root, laryngeal, and epiglottal movements associated with both the Front Raising-like factor and one of the Back Raising-like factors. However, neither Factor 3 nor Factor 4 of this solution has large tongue root movements. In this cross-language analysis, the large tongue-root and laryngeal displacements are confined to Factor 2. This is a desirable result: tongue-root advancement and retraction are represented independently of other articulatory gestures.

Factor 5 produces shapes ranging from an [ɛ]- or [æ]-like posture to an [u]-like one with slightly protruded lips (Fig. 4.15). There is some dorsal movement of the dorsal wall of the pharynx and arching of the soft palate associated with the tongue backing and raising. We might expect these articulatory gestures to be involved in the production of back rounded vowels, and indeed it accounts for a largish proportion of the data in all

three languages.

Exploring some hypotheses about articulatory primes

Above, we noted the similarity of the families of tongue positions generated by factors 3 and 4 (illustrated in Figs. 4.12 and 4.13) of the five-dimensional cross-languages PARAFAC solution. Furthermore, the tongue positions generated by these factors appear to be very similar to the ones generated by the English Front Raising factor (Fig. 1.13). Indeed, these factors appear to be a straightforward confirmation of the existence of a core articulatory prime - Front Raising - occuring in combination with various constellations of helping primes to form coordinative structures that generate the different patterns of lip, jaw, and epiglottal movement seen in these factors.

Some quantitative substantiation of the similarity of the factors comes from examining the correlation between factor loadings on individual articulators. For instance, the correlation between factor 3's and factor 4's loadings on the tongue is 0.987 (calculated over the thirteen tongue position measures t4-t16). Similarly, the correlation between factor 1's and factor 5's loadings on the measurements representing the position of the epiglottis and larynx is -0.937, suggesting that these gestures are very similar, though they are found in very different coordinative contexts (factor 1 generates large amounts of velum movement; factor 5 generates very little; the overall correlation across the articulatory mode between the two factors is only 0.644). Again, we may interpret this observation as evidence that one particular articulatory prime associated with a particular mode of larynx and epiglottis displacement has been recruited into two different coordinative structures.

Such comparisons can be carried to the absurd point of comparing single or just a few loadings. The correlations between the various factors' loadings on the measurements of jaw position are all 1.0. But this correlation - even as just a measure of the similarity between the loadings, and without any statistical interpretation - does not mean anything, since there are only two measures of jaw position. In general, since we do have multiple measures of each articulator's position (e.g., we have seven measures of lip position: ulx, uly, llx, lly, ild, uid, lid), comparisions like these should be made over all of the relevant measures.

In some cases, it seems as though the traditional notion of an 'independent articulator' (e.g., the velum) dictates an appropriate set of measures and loadings to use as the basis for comparision. But in others, the traditional notion seems to be difficult to use or less appropriate. For instance, many phoneticians and phonologists have claimed that the tongue can be regarded as several quasi-independent articulators - the blade, and the dorsum or body of the tongue are recognized, for instance, as separate

articulators by Chomsky & Halle (1968), Ladefoged & Halle (1988), and others. This division is somewhat problematic, although the results from Jackson (1988) on Icelandic (shown in Fig. 2.6) suggest that tongue blade raising can be identified as an articulatory prime. There does not seem to be widespread consensus on where exactly the blade of the tongue ends and the dorsum begins - indeed, the Blade Raising factor produces movement over most of the anterior portion of the tongue, although the tip and blade clearly move the most.

Similarly, it might be thought that the jaw and the lower lip are quasi-independent articulators. But in several of the factors in Table 4.8, the lower lip and the jaw move essentially in parallel, which suggests that they are not very independent. Apparently, in normal speech conditions, the lower lip and jaw are coupled and their potential independence is not exploited. The jaw is not a linguistic-phonetically independent articulator, even if experimental manipulations (e.g., bite blocks) can force speakers to use the lips independently of it.

These considerations have led me to compare the patterns of loadings from the five factors obtained from Akan, Chinese, and French in four blocks - the larynx and epiglottis, the tongue, the velum, and the lips and jaw. A distinct mode of variation within one of these blocks represents an articulatory prime. We may then eliminate some of the redundancy that is due to one articulatory prime's ocurrence in several coordinative structures by constructing a block-structured matrix from the loadings that represent articulator displacements due to each of the articulatory primes.

Repartitioning the factors in this way allows us to eliminate a great deal of the redundant information in the raw factors. This method of examining correlation matrices corresponds to a crude kind of 'simple structure' fitting. We are assuming that ideally, when uncontaminated by measurement error, misplacement of gridlines, speaker-specific anatomical peculiarities, etc., there ought to be perfect correlations across certain subsets of articulatory loadings due to a particular articulatory prime's having occured in more than one coordinative context.

The correlations between all the articulatory loadings of the factors are given in Table 4.9. The correlation matrices generated by partitioning the articulatory loadings into particular subsets are presented in Tables 4.10 through 4.13. Table 4.10 presents the correlations between the loadings from the various factors on the measures of tongue position. It can be seen that the correlations between the tongue loadings for factors 3, 4, and 5 are high. One the other hand, factors 1 and 2 do not have any correlations as high as the correlations between factor 3, 4 and 5. This suggests that there may be as many three articulatory primes associated with tongue displacement, one of which occurs in

three different coordinative structures (factors 3, 4, and 5). For the moment, I will suggest that none of these three patterns of tongue movement should be eliminated; all will be retained below. The pattern of tongue displacement seen in factors 3, 4, and 5, which is similar to the Front Raising factor of Harshman et al. (1977), will be called Ti; the tongue displacement pattern in factor 2 will be called Tii; and the pattern in factor 2 - the one that includes large amounts of tongue root displacement - will be called Tiii.

Table 4.9: Correlation matrix for all 44 loadings

	1	2	3	4	5
1	1.000				
2	-0.629	1.000			
3	0.582	-0.221	1.000		
4	0.508	-0.118	0.928	1.000	
5	-0.644	0.335	-0.924	-0.824	1.000

Table 4.10: Correlation matrix for 13 tongue loadings

	1	2	3	4	5
1	1.000				
2	-0.801	1.000			
3	0.710	-0.261	1.000		
4	0.698	-0.200	0.987	1.000	
5	-0.839	0.441	-0.975	-0.960	1.000

Table 4.11 presents the correlation matrix associated with larynx and epiglottis displacement. Factors 1 and 5 appear to have similar patterns: we will call the pattern of displacements they generate Li. Factors 3 and 4 also appear to have similar patterns; this pattern will be called Lii. The pattern of larynx and epiglottis displacement generated by factor 2 appears to be unique; we will call it Liii.

Table 4.11: Correlation matrix for 5 larynx and epiglottis loadings

	1	2	3	4	5
1	1.000				
2	0.004	1.000			
3	0.766	0.297	1.000		
4	-0.716	-0.617	-0.906	1.000	
5	-0.937	0.026	-0.509	0.543	1.000

The measures of the position of the dorsal wall of the pharynx seem to be quite noisy (typically, only about half of the variance in these measures is accounted for by the

within-language factor solutions). I have chosen to include the loadings of these measures in the L(arynx) primes, on the grounds that some movement of the dorsal wall of the pharynx and epiglottis is likely to be due to the action of the pharyngeal constrictors, and thus intrinsically correlated.

The correlation matrix associated with velum displacement (Table 4.12) does not show clear structure. It is worth noting, however, that only factors 1 and 5 have substantial loadings on the measures of velum position. Thus, it seems likely that if any of the factors have substantial non-noise components, they do. Observing that the correlation between them is neither very high nor very low, I have chosen to provisionally retain both, keeping in mind the fact that one may well turn out to be redundant. The pattern of velum displacement due to factor 1 will be called Vi below; the pattern seen in factor 5 will be called Vii.

Table 4.12: Correlation matrix for 12 velum loadings

	1	2	3	4	5
1	1.000				
2	-0.186	1.000			
3	-0.131	0.156	1.000		
4	0.485	-0.272	-0.169	1.000	
5	0.706	-0.138	-0.249	0.814	1.000

The correlation matrix for the lip and jaw displacement measures is similarly problematic (Table 4.13). As above, we should note that some factors - 2, 3 and 4 - have loadings on the lip and jaw displacement measures that are fairly large compared to the other factor's loadings. I have provisionally chosen to retain the loadings from factors 2, 3 & 4. The lip and jaw displacement pattern from factor 2 will be called LJii; from factor 3, LJi; and from factor 4, LJiii.

Table 4.13: Correlation matrix for 10 lip and jaw loadings

	1	2	3	4	5
1	1.000				
2	-0.870	1.000			
3	0.596	-0.534	1.000		
4	0.226	0.113	0.635	1.000	
5	-0.706	0.536	-0.899	-0.630	1.000

Thus, we have hypothesized that - distributed among the five factors of the cross-language analysis - there are some eleven articulatory primes combined in various

Table 4.14: Loadings of the hypothesized articulatory primes

Measure	Articulatory prime									
	Li	Lii	Liii	Ti	Tii	Tiii	Vi	Vii	LJi	LJii	LJiii
e1	0.468	-0.260	1.283								
e2	0.412	-1.009	1.602								
e3	0.647	-0.845	2.177								
e4	1.380	-0.685	3.237								
p3	1.065	-0.549	-1.063								
p4	1.087	-0.279	-0.749								
p5	1.105	-0.150	-0.570								
p6	0.904	0.063	-0.501								
p7	0.940	0.259	-0.282								
t4				-4.298	-3.647	5.815					
t5				-5.442	-3.550	4.611					
t6				-5.586	-3.064	1.710					
t7				-4.486	-1.996	0.230					
t8				-4.697	-1.268	-0.744					
t9				-3.384	-0.137	-1.055					
t10				-1.037	1.323	-1.388					
t11				2.081	2.561	-1.710					
t12				4.833	3.173	-1.254					
t13				6.109	3.140	-1.032					
t14				6.338	2.483	-0.639					
t15				6.059	1.704	-0.112					
t16				4.995	0.908	0.255					

vi1				2.134	0.903				
vs1				0.010	0.575				
vi2				2.415	0.794				
vs2				2.636	0.857				
vi3				1.839	0.859				
vs3				2.000	0.267				
vi4				1.325	0.753				
vs4				1.006	0.350				
vi5				0.829	0.558				
vs5				0.317	0.370				
tvx				-0.196	-0.906				
tvy				2.583	1.206				
ljx						-0.137	-0.773	0.806	
ljy						1.555	-1.817	1.916	
llx						-0.697	-1.691	-0.709	
lly						0.918	-2.932	-0.103	
ulx						-0.793	-1.213	-0.777	
uly						0.860	-0.076	0.368	
ild						-0.671	2.868	0.685	
lid						-1.727	1.490	-0.347	
uid						-0.334	-1.534	-0.597	
lyy	-0.247	-0.108	3.422						

ways. Factor 1, for instance is here viewed as some linear combination of the Li, Tii, Vi, (and LJii) primes. Factor 2 is a combination of the Liii, Tiii, and LJii primes. Factor 3 is a combination of the Lii, Ti, and LJi primes; Factor 4, Lii, Ti, and LJiii; and Factor 5, Li, Ti, Vii, (and LJi). Some of the apparent variation between factors is thus explained as a result of the combinatorial freedom of the underlying articulatory primes. There are eleven hypothesized articulatory primes underlying these factors: three having to do with larynx, epiglottis, and pharynx displacement; three for the tongue; two for the velum; and three for the lips.

With these observations in mind, we may now *construct* a matrix containing factors that correspond to the articulatory displacements due to these hypothesized articulatory primes. Each column of the matrix contains only loadings on one articulator; the precise values of the loadings are copied from one member of each of the 'similarity sets' discussed above, except for the Ti loadings. Because of the distortion in the shape of the factor noted in the preliminary study cited in Chapter 2 (c.f. Figs. 1.13, 2.11, and 2.12), the loadings from the within-language solutions' Front Raising-like factors were averaged to produce the loadings in Table 4.14. The eleven sets of articulatory loadings corresponding to our hypothesized articulatory primes are presented in Table 4.14.

Table 4.14 (our hypothesized articulatory primes), despite containing more factors than Table 4.8 (the original five-factor solution), contains *less* information. Many of the factor loadings in Table 4.14 are zero; there are only 114 non-zero loadings. On the other hand, Table 4.8 contains five factors, each of which has 44 non-zero loadings, for a total of 220. This reduction of the number of loadings specified in our model of vowel articulation is allowed by our elimination of redundant sets of loadings that we hypothesize to be due to the use of articulatory primes in several different coordinative contexts.

Now comes the interesting part: testing the hypotheses about the organization of vowel production quantitatively embodied in Table 4.14. First, the articulatory primes themselves: - How well do the eleven articulatory primes fit the original data? Is this provisional set satisfactory, too large, or too small? Second, some of the hypotheses under which the model was constructed may be tested: - Will these articulatory primes show the kind of functional dependency that we have postulated to be at the heart of their organization into coordinative structures? Do the different languages all require the same set of articulatory primes to describe the articulation of vowels, or do the different languages exploit different sets of articulatory primes? I will adopt a procedure related to the backward-stepping algorithm used in some forms of exploratory multiple regression to try to answer these and related questions. This procedure was inspired by

the discussion of Wald tests in Bentler (1986).

The backward-stepping algorithm I will use begins by modeling the measures of articulator position in each language with the eleven candidate articulatory primes in Table 4.14 using a PARAFAC representation. We can then drop each prime in turn to see whether or not the fit of this reduced model to the articulatory data decreases significantly. We then permanently drop the least-contributing factors, effectively setting its contribution to zero for that language.

This backward-stepping process can be repeated until none of the remaining primes can be dropped without a significant decrement in the fit of the model to the raw data. Any of the hypothesized primes in Table 4.14 that is dropped from all the languages is clearly suspect, and may only be due to correlated error in the data. On the other hand, if a language uses some but not all of the *valid* articulatory primes (as Chapter 2 suggested), we may conclude that its vowels lie in a subspace of the universally available vowel space.

A final observation is that the procedure used for constructing the articulatory primes has partially orthogonalized them (in the articulatory mode). Thus, the contributions to the model by, e.g., factors representing primes of tongue (T) displacement are independent of the contributions from, e.g., the velum (V) factors. This allows considerable acceleration of the early stages of the backward-stepping process, since joint effects of the sort that occur in multiple regression with correlated regressors are limited. Several independent, uncorrelated factors may be excluded at one step.

Backward-stepping analysis: Akan

Table 4.15 presents the results of the first step for Akan. When examining the converged, unique solution using the articulatory primes specified in Table 4.14, it was clear that three particular primes, Li, Vi, and LJiii, contributed very little to the Akan data. Therefore, they were excluded, reducing the model to eight primes.

Table 4.15: Backward-stepping analyses of Akan articulatory data: I

N_p	R^2	Primes remaining
11	0.8488	all
8	0.8345	Lii, Liii, Ti, Tii, Tiii, Vii, LJi, LJii

The fit value of the model decreases less than 1.5% when the three primes are eliminated. The eight articulatory prime model has an R^2 of 0.8345. For comparision, the PARAFAC three-factor model of the Akan data had an R^2 of 0.8376.

The next prime to eliminate is determined by dropping each of the primes in turn from the model and determining the fit value. The prime that contributes least is

dropped, if its removal does not significantly decrease the fit. Since the model has eight primes, there are eight potential seven-prime models to evaluate. Table 4.16 presents the fit figures for these reduced models.

Table 4.16: Backward-stepping analyses of Akan articulatory data: II

N_p	R^2	Primes remaining
7	0.8230	Liii, Ti, Tii, Tiii, Vii, LJi, LJii
7	0.8039	Lii, Ti, Tii, Tiii, Vii, LJi, LJii
7	0.6966	Lii, Liii, Tii, Tiii, Vii, LJi, LJii
7	0.8232	Lii, Liii, Ti, Tiii, Vii, LJi, LJii
7	0.8115	Lii, Liii, Ti, Tii, Vii, LJi, LJii
7	0.8252	Lii, Liii, Ti, Tii, Tiii, LJi, LJii
7	0.8156	Lii, Liii, Ti, Tii, Tiii, Vii, LJii
7	0.8108	Lii, Liii, Ti, Tii, Tiii, Vii, LJi

The sixth row of this table, for the model which has Vii excluded, shows the smallest decrement in R^2. This prime, like the first two eliminated, contributed less than 1% to the variance accounted for by the model. This is not surprising, since the Akan data set has no nasal vowels in it.

Table 4.17: Backward-stepping analyses of Akan articulatory data: III

N_p	R^2	Primes remaining
6	0.8135	Liii, Ti, Tii, Tiii, LJi, LJii
6	0.7944	Lii, Ti, Tii, Tiii, LJi, LJii
6	0.6806	Lii, Liii, Tii, Tiii, LJi, LJii
6	0.8108	Lii, Liii, Ti, Tiii, LJi, LJii
6	0.8006	Lii, Liii, Ti, Tii, LJi, LJii
6	0.8066	Lii, Liii, Ti, Tii, Tiii, LJii
6	0.8013	Lii, Liii, Ti, Tii, Tiii, LJi,

Table 4.17 shows the R^2 values for the next step. Lii is clearly the least important remaining prime; its omission decreases the fit value by some 1.2%.

Table 4.18: Backward-stepping analyses of Akan articulatory data: IV

N_p	R^2	Primes remaining
5	0.7779	Ti, Tii, Tiii, LJi, LJii
5	0.6688	Liii, Tii, Tiii, LJi, LJii
5	0.7991	Liii, Ti, Tiii, LJi, LJii

5	0.7889	Liii, Ti, Tii, LJi, LJii
5	0.7947	Liii, Ti, Tii, Tiii, LJii
5	0.7897	Liii, Ti, Tii, Tiii, LJi

Table 4.18 shows the results for models with 5 articulatory primes. When Tii is removed from the model, the R^2 drops only about 1.4%; removing any of the other articulatory primes decreases it of the model by about 2%. Is the removal of Tii justifiable? It is difficult to design statistical tests of significance for this kind of data because the errors in fitting different measurements are not independent of each other. Thus, the number of degrees of freedom in the error model are difficult to evaluate. It seems prudent to use two estimates of the degrees of freedom: first, the number of measurements actually being fit by the model (equivalent to the unlikely assumption that all the measurement errors are independent); and second, half that number (equivalent to assuming that half of the errors are actually predictable). I will take a crude stab at testing the significance of a particular prime's contribution to the model as follows. We know from earlier studies (in particular, Harshman et al. 1977) that we may reasonably expect a mean squared error of some 3 mm^2 over all the points of the tongue in a good PARAFAC model of tongue positions. This independent assessment of the probable measurement error σ allows us to calculate a χ^2 for the model with, and without Tii, as below.

$$\chi^2 = \sum_{i}\sum_{j}\sum_{k}(y_{ijk} - \sum_{r} a_{ir} b_{jr} c_{kr})^2 / \sigma$$

In this equation, $(y_{ijk} - \Sigma a_{ir} b_{jr} c_{kr})$ is the error in fitting the *i*th measure on the *j*th vowel as produced by the *k*th speaker. The sum of squared errors (SSE) is given by $\Sigma\Sigma\Sigma(y_{ijk} - \Sigma a_{ir} b_{jr} c_{kr})^2$; the PARAFAC program does not report this value, but instead reports the mean squared error, which is simply $SSE/(N_{measures} * N_{vowels} * N_{speakers})$. It is therefore fairly easy to at least estimate a χ^2 value for a PARAFAC model.

After calculating a χ^2 for each of the models (the one with Tii, and the one without Tii), we can apply a χ^2-difference test. In this case, the mean squared error (MSE) of the best 6-prime model for Akan is 6.673174 mm^2; the MSE of the model with Tii dropped is 7.145765 mm^2. The difference between the two models - entirely due to the exclusion of Tii - is 0.472591 mm^2. For a dataset with 1467 measurements (excluding missing values),

and the estimate given above for a plausible mean square error, this gives a χ^2 of (0.472591)(1467)/3, or about 230. The Tii prime is actually only fitting the 13 measures of tongue position, so the actual number of points that this χ^2-difference is attributable to is 13 x 9 x 4 = 468. However, since the errors in fitting these points are not independent, a crude, but more realistic estimate of the error d.f. is 468/2 = 234. Adding Tii adds 9 vowel loadings and 4 speaker loadings to the model. The probability of the best model excluding this prime under the assumption of 468 d.f. is p(468-13,230) $\cong$ 1.0. In other words, if there are really 468 independent measurements with independent errors, this prime does not add enough to the overall fit of the model to make it worthwhile. The probability of the model excluding this prime with 234 d.f. is p(234-13,230) $\cong$ 0.3 (evaluated by Press et al.'s (1986) GAMMQ function). If there are more like 234 independent errors, then this prime does add enough to the overall fit to make it worthwhile. Thus, it seeems likely that this prime should be included in the model.

For comparision, the probability of the Vii prime's being excluded from the model is $\cong$ 1.0 under the larger estimate of error d.f., and $\cong$ 0.77 under the half-as-large estimate. On the other hand, the probabilities of the correct model's excluding any of the five primes that remain after Tii is removed range from $< 10^{-7}$ to 0.1 under the larger error d.f. estimate, except for LJi, which has a probability of 0.88 of not being included. Under the smaller d.f. estimate, all of the primes have a probability $< 10^{-7}$ of being excluded. We thus stay with the best model in Table 4.17, namely the one with Liii, Ti, Tii, Tiii, LJi, and LJii.

Is there any structure to this six-prime model? In particular, is it the case that - as we hypothesized in Chapter 2 - there are articulatory primes which are functionally correlated? I assume (rather restrictively) that functional correlation of the sort that interests us may be expressed as a linear dependance of one articulatory prime on another. As I observed in Chapter 2, if the rth prime is functionally correlated with the sth prime in this way, then their contributions to the vowels of the language w_{jr} and w_{js} will be correlated. Table 4.19 presents the matrix of these correlations for all pairs of articulatory primes in the Akan solution.

Table 4.19: Correlation matrix for Akan vowel loadings on the articulatory primes

	Liii	Ti	Tii	Tiii	LJi	LJii
Liii	1.000					
Ti	-0.1457	1.000				
Tii	0.0903	-0.3914	1.000			
Tiii	0.8949	0.2249	-0.1760	1.000		
LJi	-0.0603	-0.8731	0.6627	-0.2850	1.000	
LJii	-0.8013	-0.1599	-0.2854	-0.8019	0.0592	1.000

There are four correlations significant at the 0.01 level or better in this matrix - the correlations between Liii & LJii ($p<0.01$); Liii & Tiii ($p<0.005$); Ti & LJi ($p<0.005$); and Tiii & LJii ($p<0.01$). Liii, Tiii, and LJii thus apparently consistently act together in producing the vowels of Akan. The contributions of Liii, Tiii, and LJii to the observed articulatory configurations of Akan vowels are strongly related and can probably be approximated by linear functions of one underlying variable. Ti and LJi are similarly related. These, then, are candidates for being coordinative structures, i.e., ensembles of articulatory primes that act together to produce the vocal tract positions characteristic of the vowels of a language. *However, we cannot assume that they are coordinative structures until it is demonstrated that these correlations are language-specific.* It is always possible that we divided up the primes above into articulatory primes incorrectly, and that loadings typical of some unitary articulatory gesture got distributed across several of our hypothesized articulatory primes. We will return to these candidate coordinative structures in Chapter 7, after gathering more evidence as to the ways in which different languages use these articulatory primes.

Backward-stepping analysis: Chinese

Various vowel tokens that were not used in the data set used to construct this model were available for Ohnesorg & Svarný's (1955) speaker B ([o]) and Zhou & Wu's (1963) speaker ([e] and [o]). I have used these extra vowels along with the five phonemic vowels originally used to construct the model.

Table 4.20: Backward-stepping analyses of Chinese articulatory data

N_p	R^2	Primes remaining
11	0.803+	all
5	0.7909	Ti, Tii, Tiii, Vii, LJii
4	0.7660	Ti, Tii, Tiii, LJii
3	0.7174	Ti, Tii, LJii

2	0.5395	Tii, LJii
2	0.5756	Ti, LJii
2	0.6371	Ti, Tii

These results suggest that the four-prime model for Chinese is best. At the first step, all of the primes having to do with laryngeal, pharyngeal, and epiglottal movement were discarded along with one of the velar primes, with only about a 1% decrease in the overall fit value. Dropping the Vii prime reduces the fit value about 2%; dropping any of the others reduces the fit value by at least 5%.

There are no significant correlations between any of the primes in the vowel mode.

Backward-stepping analysis: French

Table 4.21 presents the results of the backward-stepping procedure for French. As the table shows, it is possible to prune the model for the French articulatory data down to four articulatory primes with little loss in the goodness of fit. The first five articulatory primes removed each account for under 1% of the variance in the data set. The fifth accounts for just over 2%; the four remaining ones account for about 67%. The R^2 using four primes compares favorably with the R^2 achieved by the unconstrained PARAFAC 2-factor solution (0.6469). But it is not as good as the R^2 reached by the unconstrained PARAFAC procedure at four dimensions (0.762+); on the other hand, PARAFAC was unable to converge to a unique solution for several hundered iterations. It is also not as good as the R^2 of the relaxed Procrustean solution. The fit values for the four different possible three-prime models derived from the good four-prime model demonstrate that none of them are very good.

Table 4.21: Backward-stepping analyses of French articulatory data

N_p	R^2	Primes remaining
11	0.7516	all
6	0.7151	Li, Liii, Ti, Tii, Vi, LJii
5	0.7076	Liii, Ti, Tii, Vi, LJii
4	0.6855	Ti, Tii, Vi, LJii
3	0.5657	Tii, Vi, LJii
3	0.5236	Ti, Vi, LJii
3	0.6261	Ti, Tii, Vi
3	0.6129	Ti, Tii, LJii

Returning for a moment to the last prime dropped before the four-prime model (Liii), we might ask whether its exclusion is justified - 2% seems like a substantial

decrement in fit, particularly since the initial model does not fit the data very well. Using the same crude procedure as above, we may calculate the χ^2 improvement due to adding Liii and 18 (14 vowel loadings + 4 speaker loadings) free parameters to the model. The MSE for the five-prime model including Liii is 5.561186 mm^2; the MSE for the four-prime model is 5.967750. The improvement in MSE is 0.4065638 mm^2; since the dataset has 2449 points, the overall SSE has improved by 995.67 mm^2. Normalized by the rough (but independent!) estimate of the model error for a good model, this gives a χ^2 improvement of about 332. for the Liii prime. The Liii prime fits 560 measurements (10 measures of pharyngeal wall, epiglottis, and larynx position x 14 vowels x 4 speakers). As usual, the larger estimate of d.f. gives p(560-18,332.) $\cong$ 1.0: i.e., under the asssumption of completely independent errors, Liii does not fit enough variance to be worth including. The smaller estimate gives p(280-18,332.) $\cong 2.x10^{-3}$: under a more realistic assumption, Liii seems worthwhile. We should thus retain the five-prime model including Liii, Ti, Tii, Vi, and LJii for French.

The correlations between the various primes' contributions across vowels are presented in Table 4.22.

Table 4.22: Correlation matrix for French vowel loadings on the articulatory primes

	Liii	Ti	Tii	Vi	LJii
Liii	1.000				
Ti	.081	1.000			
Tii	.800	.046	1.000		
Vi	-.485	-.142	-.618	1.000	
LJii	-.618	-.367	-.392	.249	1.000

There is only one correlation that is significant at the 1% level or better in Table 4.22; the correlation between Tii & Liii ($p < 0.001$). There are two other pairs of primes approaching significant correlation - Tii & Vi and LJii & Liii (both $p < 0.02$). The relation between Tii & Vi is not surprising: as we have remarked several times, all the phonemically nasal vowels of French are low, and Tii is the articulatory prime that most closely corresponds to the traditional notion of a high/low articulatory parameter. Nor is the possibility of a relation between LJii & Liii surprising: it has been shown before that rounded vowels often involve larynx lowering (Riordan 1977).

Articulatory primes, coordinative structures, and things that go bump in the night

In these three languages, we have some evidence that at least seven (Liii, Ti, Tii, Tiii, Vi, LJi and LJii) of the eleven primes originally hypothesized are required to provide an accurate description of our articulatory data. It is clear that Ti, Tii, and LJii are the most widely used of the primes; Vi and LJi are only used in French and Akan, respectively.

What about the four primes that do not appear to contribute significantly to the original data (Li, Lii, Vii, LJiii)? There are several possibilities. First, as we noted in Chapter 2, the PARAFAC covariance-analysis approach does not represent the constraint that all speakers of a given language be modeled as having produced the 'same' vowels. The lack of this constraint introduces the possibility that PARAFAC could find individual-difference factors rather than truly systematic factors. If this has occurred, it is possible that some speaker or speakers have idiosyncratic patterns of - e.g. - pharynx displacement that are represented in the covariance model but are not represented in models of the raw data. This might occur, for example, if some speakers' heads shifted during the making of the x-rays.

Second, there is the possibility that some of these primes are not reliably detectable in the small within-language samples of speakers, even though they were detectable when all the speakers were pooled. For instance the Vii prime might represent a real but small in magnitude displacement of the velum associated with tongue lowering. In any one language's worth of speakers, it might not be a large enough effect to contribute significantly to the raw data. This is a typical problem of small-sample statistics.

Third, there is the possibility that some of these primes represent correlated errors in x-ray tracing or measurement. An error that occurred during the measurement of a particular token would create errors correlated across gridlines but not across vowels or speakers. Thus, it is likely that such errors would not be fit when PARAFAC is applied to raw data, where the data is simultaneously modeled across gridlines, vowel and speakers. But in the covariance analysis, the classification of the data by vowel no longer exists. Correlated errors are thus seen across gridlines but not across speakers. PARAFAC can model this situation by allowing a factor corresponding to correlated errors to vary in size across speakers.

Furthermore, in any of these cases, the apparent 'extra' factor might actually show up in linear combination with other factors. Then, the data from various speakers would be represented by various combinations of the 'pure' and 'corrupted' factors, depending on how much correlated error there was in the data from that speaker as a whole. But the same is also true of small effects that are not reliably detectable in small samples and of indivdual-difference factors. Thus, the four primes that do not appear to contribute

significantly to the data cannnot be immediately excluded from the inventory of articulatory primes.

Nonetheless, we have analyses consistent with the hypotheses advanced in Chapter 2. We have been able to obtain a small set of putatively universal articulatory primes (although some primes are not exploited in particular languages). These primes are capable of reproducing the observed patterns of articulatory positioning in three languages, with a degree of fit generally comparable to that achieved by unconstrained, within-language PARAFAC models. The patterns of tongue displacement due to Ti and Tii occur in all three languages. The tongue root displacement characteristic of Tiii occurs in both Akan and Chinese. The lip protrusion / jaw raising gesture represented by LJii occurs in all three languages.

There appear to be language-specific functional dependencies (that I have assumed to be linear) that link these articulatory primes into coordinative structures. For instance, Ti & LJi appeared to be functionally linked in Akan, but not in any of the other languages. Liii, Tiii & LJii are functionally linked in Akan, but not in the other languages. The combinatorial freedom of articulatory primes is suggested by the fact that Liii - a larynx-displacement gesture - can apparently be functionally linked with different kinds of tongue displacement. Liii is linked with Tiii in Akan, but with Tii in French.

However, these conclusions are based only on the languages used to construct the model - the set of articulatory primes - in the first place. There is thus some circularity in the exercise just completed. For non-circular confirmation of the generalizability of this model, we turn to extension: testing the model against comparable articulatory data from Icelandic, Spanish, and Swedish.

Chapter 5: Extending the articulatory model

Up to now, we have been concerned with the validity of the articulatory primes for describing the articulation of vowels from which the primes were - however indirectly - derived. But to have a truly useful model, we need to know that it is able to describe the articulation of vowels from other languages as well. Since our account of articulatory primes assumes that they are universal - part of the phonetic capabilities of mankind - cross-linguistic generalizability of the primes is a theoretical desideratum as well. We will therefore turn to using the primes from Chapter 4 as models of vowel articulation in Icelandic, Spanish, and Swedish.

Testing the articulatory primes against data from other languages also allows further theoretical conclusions to be drawn. Fitting the primes to data from other languages gives further evidence about the language-specificity or universality of functional correlations. In the case that a particular functional linkage is demonstrated to be language-specific, we may conclude that the articulatory primes are really primes. But if a particular functional correlation turns out to be uniform in our sample, then we must accept the conclusion that a unitary articulatory prime has had its effects (wrongly) partitioned between two (or more) of the hypothesized primes.

Some of the proposed primes of Chapter 4 also were not shown to contribute substantially to the articulatory data from Akan, Chinese, and French. Testing the articulatory primes against data from other languages provides a chance for 'real' primes that simply do not contribute much to Akan, Chinese, and French to show up. On the other hand, if some hypothetical articulatory prime does not make substantial contributions to either the data from the original three languages or to data from any additional language then it seems unlikely to be a systematically linguistically available prime.

Backward-stepping analysis: Icelandic

The Icelandic data consist of 1216 measurements (44 measures x 16 vowels x 2 speakers - 192 missing values) of articulator position. There are no measures of larynx or epiglottis position in the data set. The original publications by Pétursson (1974ab) contain tracings of tongue, lip, jaw, and velum position at several instants in each vowel. However, in each case, Pétursson has selected one tracing as typical of the articulatory extreme ('culminating phase', in his terms) of the vowel. I have used this tracing in all cases. The measurement grids for these speakers were constructed using the tracing for the vowel [ɛ].

Unconstrained PARAFAC analyses of the Icelandic data were obtained at one through five dimensions for comparision with the eventual fit using our hypothesized

articulatory primes. The fit values for these models are presented in Table 5.1.

Table 5.1: PARAFAC fits for the Icelandic dataset

r	R^2	converged	unique	degenerate
1	0.6206	yes	yes	-
2	0.7978	yes	yes	no
3	0.8598	yes	yes	maybe
4	0.8952	yes	yes	yes
5	0.919+	yes	no	yes

As has often been the case, it is not clear which dimensionality is correct for the Icelandic solution. In Icelandic, the front rounded vowels /yː, y, øː, ø/ are not as front and high as the front unrounded vowels /iː, i, ɪː, ɪ, eː, e/, so (as we saw in French) there exists a partial correlation between Front Raising and rounding. The three-factor solution's degeneracy appears to reflect this correlation. Nonetheless, it is highly interpretable. Thus it seems that this solution is the one that should be compared to any model fit using the articulatory primes. The vowel space produced by this solution is shown in Fig. 5.1.

The backward-stepping analysis of this dataset using the articulatory primes suggested in Chapter 4 is presented in Table 5.2.

Table 5.2: Backward-stepping analyses of Icelandic articulatory data

N_p	R^2	Primes remaining
11	0.795+	all
4	0.8029	Ti, Tii, Vi, LJii
3	0.6714	Tii, Vi, LJii
3	0.6546	Ti, Vi, LJii
3	0.7774	Ti, Tii, LJii
3	0.6768	Ti, Tii, Vi

Due to the small number of speakers, most of the models with large numbers of primes did not converge when fit to the Icelandic data. Nonetheless, it is clear that none of the three-prime models fit the data well. The best three-prime model uses Ti, Tii, and LJii. The four-prime model that adds Vi fits about 2.5% more of the overall variance.

The correlations between the various primes' contributions across vowels are presented in Table 5.3.

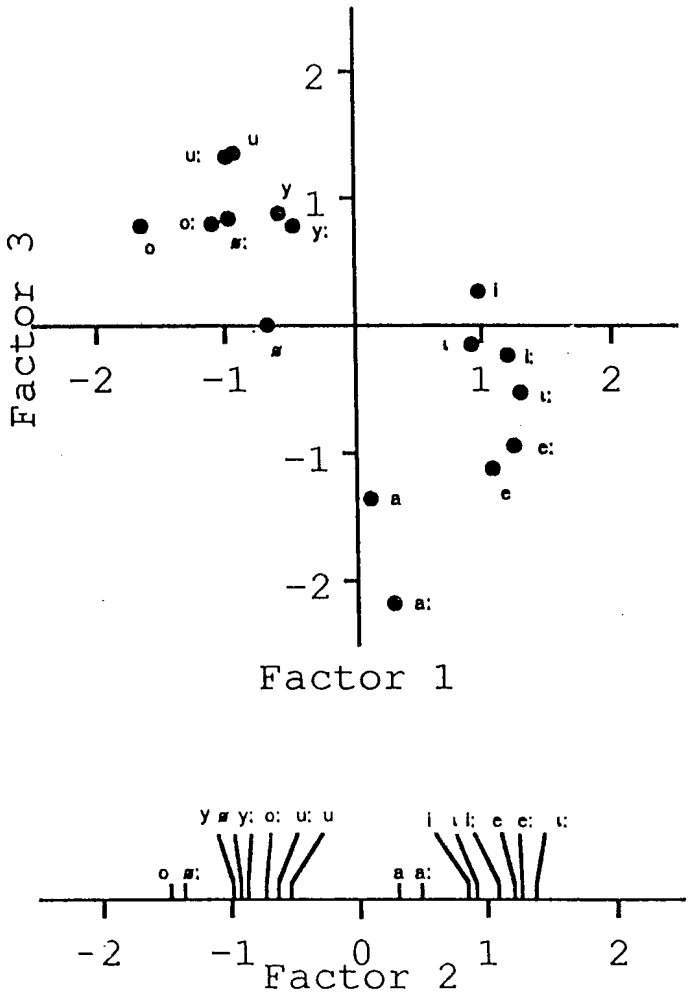


Figure 5.1. Icelandic 3-factor vowel space.

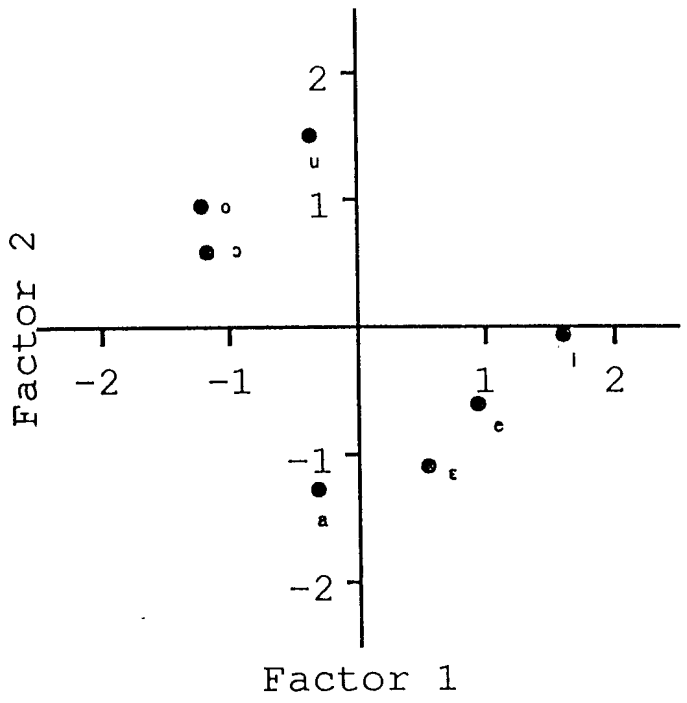


Figure 5.2. Spanish 2-factor vowel space.

Table 5.3: Correlation matrix for Icelandic vowel loadings on the articulatory primes

	Ti	Tii	Vi	LJii
Ti	1.000			
Tii	-.225	1.000		
Vi	-.209	-.243	1.000	
LJii	-.414	-.090	.836	1.000

Vi and LJii both make large negative contributions to /ɑː/ and /ɑ/, corresponding to lowered (though probably not fully lowered) velum positions and fully unrounded and open lip positions. They make near-zero contributions to front unrounded vowels. Vi makes moderate positive contributions to the high back vowels /uː/ and /u/; LJii makes positive contributions to all the rounded vowels, with the largest positive contributions being to the high rounded /u/- and /y/-like vowels. Velum position is thus strongly correlated with lip and jaw position ($r = 0.836$, $p < 0.001$).

Backward-stepping analysis: Spanish

The Spanish data set consists of 1215 measurements (44 measures x 7 vowels x 5 speakers - 325 missing values). Most of the missing values are due to the fact that the velum and larynx were not traced by many of the original authors. Others are due to the fact that x-ray tracings of the vowels [ɛ]and [ɔ] are not available for the speakers from Russell (1929-30) and Quilis (1981). The measurement grids for all speakers were constructed using the tracings of the vowel [e].

Unconstrained PARAFAC analyses of the Spanish data were obtained at one through six dimensions for comparision with the eventual fit to articulatory primes.

Table 5.4: PARAFAC fits for the Spanish dataset

r	R^2	converged	unique	degenerate
1	0.4886	yes	yes	-
2	0.7072	yes	yes	no
3	0.8388	yes	yes	yes
4	0.887+	no	no	yes
5	0.919+	no	no	yes
6	0.931+	no	no	yes

As the table shows, the 2-factor solution for Spanish appears to be the best. The solutions with greater numbers of factors are either degenerate, failed to converge within 500 iterations, or both. The vowel space of this solution is plotted in Fig. 5.2. It can be

seen that this solution is very similar to within-language solutions that we have seen before, in English, Akan, and Chinese.

The backward-stepping fits of the Spanish data to the articulatory primes are summarized in Table 5.5. As before, only the best model arrived at by eliminating primes from the previous model is presented until the last step.

Table 5.5: Backward-stepping analyses of Spanish articulatory data

N_p	R^2	Primes remaining
11	0.806+	all
5	0.7243	Liii, Ti, Tii, Vii, LJii
4	0.6993	Ti, Tii, Vii, LJii
3	0.6120	Ti, Tii, LJii

The first six primes dropped from the model for the Spanish vowels decreased the degree of fit by about 1.3% apiece. The five-prime model thus arrived at, however, showed a problem unusual in a PARAFAC model with one mode (the articulator loadings of the primes) fixed. The Liii prime's estimated contribution to the data from one speaker fluctuated by a factor of about two from one solution to the next. This nonuniqueness is apparently due to the large number of missing measurements for that speaker. The few measurements available for that speaker apparently were not sufficient to uniquely determine speaker scaling for that prime. For this reason, Liii was dropped from the analysis to yield a four-prime model.

Furthermore, it appeared very much as though the Vii prime's contribution was largely determined by the data from only one speaker. This speaker (from Russell 1929-30) had loadings on Vii that were up to two orders of magnitude greater than the other four speakers' loadings. Since Vii appeared to be speaker-specific, rather than systematic, it too was dropped, giving the final, three-prime model.

The correlations between the contributions of the primes to the seven vowels of Spanish are summarized in Table 5.6. There are no correlations significant at the 1% level in Table 5.6. It thus appears to be the case that there are no detectable functional linkings in this dataset. Given that nearly 25% of the measurements in this dataset are missing, these problems are to be expected.

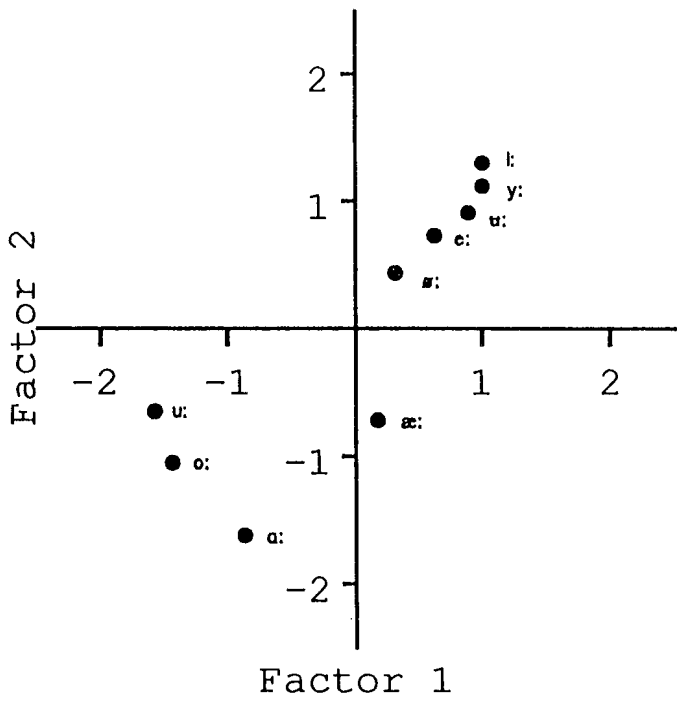


Figure 5.3. Swedish 2-factor vowel space.

Table 5.6: Correlation matrix for Spanish vowel loadings on the articulatory primes

	Ti	Tii	LJii
Ti	1.000		
Tii	-.034	1.000	
LJii	-.727	-.428	1.000

Backward-stepping analysis: Swedish

The Swedish data set consists of 1491 measurements (44 measures x 9 vowels x 4 speaker - 93 missing values). Most of the missing values are due to the fact that tracings for one speaker's productions of [æː, oː] are not available. The measurement grid for this speaker were constructed using the tracing of the vowel [eː]; the vowel [æː] was used for all the other speakers.

Unconstrained PARAFAC analyses of the Swedish data were obtained at one through six dimensions as before, for comparision with the eventual fit to articulatory primes.

Table 5.7: PARAFAC fits for the Swedish dataset

r	R^2	converged	unique	degenerate
1	0.6018	yes	yes	-
2	0.7224	yes	yes	no
3	0.7800	yes	yes	maybe
4	0.8336	yes	yes	yes
5	0.873+	1/3	no	yes
6	0.905+	no	yes	yes

As the table shows, the two-factor solution for Swedish appears to be the best. The solutions with greater numbers of factors are either degenerate, failed to converge within 500 iterations, or both. On the other hand, some of the later factors appear to add significantly to the R^2 value of the PARAFAC model, and the R^2 for the two-factor solution is not terribly good. This pattern is similar to the one that we saw in the PARAFAC analysis of the French and Icelandic data, and so we should bear in mind the fact that the Swedish data may contain correlations of the same sort that made the French data difficult to analyze. The vowel space of two-factor solution is plotted in Fig. 5.3.

The backward-stepping fits of the Swedish data to the articulatory primes are summarized in Table 5.3. As before, only the best model arrived at by eliminating primes

from the previous model is presented until the last step. The five possible four-prime models are all shown, to show that all of them produce substantial decrements in the R^2 of the model. The R^2 of the five-prime model is about the same as that of the unconstrained two-factor PARAFAC model; and about 7% lower than the R^2 of the unconstrained three-factor PARAFAC model.

Table 5.8: Backward-stepping analyses of Swedish articulatory data

N_p	R^2	Primes remaining
10	0.7552	all
7	0.7345	Lii, Liii, Ti, Tii, Tiii, Vii, LJii
6	0.7250	Lii, Liii, Ti, Tii, Tiii, LJii
5	0.7158	Lii, Liii, Ti, Tii, LJii
4	0.6700	Liii, Ti, Tii, LJii
4	0.6909	Lii, Ti, Tii, LJii
4	0.5864	Lii, Liii, Tii, LJii
4	0.5973	Lii, Liii, Ti, LJii
4	0.6550	Lii, Liii, Ti, Tii

The correlations between the loadings of the nine Swedish vowels in this sample on the various articulatory primes are presented in Table 5.9. As we can see, there are probably significant correlations between the Tii & Lii primes ($p < 0.01$).

Table 5.9: Correlation matrix for Swedish vowel loadings on the articulatory primes

	Lii	Liii	Ti	Tii	LJii
Lii	1.000				
Liii	0.324	1.000			
Ti	-.681	-.255	1.000		
Tii	.844	.578	-.352	1.000	
LJii	-.411	-.482	.066	-.547	1.000

No other correlations in this matrix are significant. This is the only language to require the Lii prime. The magnitude of Lii's contribution to the data can be seen from the fact that dropping Lii from the five-prime model produces a decrement in fit that is almost as large as that produced by dropping LJii. Articulations with the larynx are apparently almost as important as those with the lips and jaw for these speakers of Swedish.

On the other hand, it may be the case that only one out of the four is really using an articulatory gesture corresponding to the Liii prime. All of the speaker loadings for Ti,

Tii, and LJii are within a factor of about two of each other; on the other hand, the speaker loadings on Lii are 0.3223, 1.013, 1.417, and 0.6234; on Liii, 1.412, 0.1180, 0.3779, and 0.0835. Thus, the first speaker's loading on Liii is more than three times as large as the other speakers': the mode of articulatory displacement represented by Liii does not seem to be used consistently by all of the speakers.

Success of the extension

We are now in a position to assess the success of this set of articulatory primes as a mechanism for describing the articulatory configurations of vowels and thus the space of possible vowel specifications that underlies the phonetic targets of any particular language. A comparision of the fits acheived by unconstrained PARAFAC analyses with those attainable by PARAFAC with the articulatory loadings fixed - i.e., using the articulatory primes - is presented in Table 5.10.

Table 5.10: A comparision between unconstrained PARAFAC models and models using articulatory primes

	Unconstrained PARAFAC				Articulatory primes				
Language	N_m	N_{fp}	SSE	χ^2	N_m	N_{fp}	SSE	χ^2	
Akan	1467	171	9213	3071	1089	78	6451	2150	
Chinese	643	104	1641	1234	772	30	672	505	
French	2449	248	12457	4152	2449	90	13703	4568	
Icelandic	1216	168	4627	1542	1236	72	7530	2510	
Spanish	1215	112	14542	4847	682	36	7544	2515	
Swedish	1491	114	9158	3053	1120	65	7554	2518	

Language	R^2	χ^2/d.f.	R^2	χ^2/d.f	
Akan	0.8376	2.37	0.8135	2.13	Liii,Ti,Tii,Tiii,LJi,LJii
Chinese	0.7651	2.29	0.7660	0.68	Ti, Tii, Tiii, LJii
French	0.7349	1.89	0.7076	1.94	Liii, Ti, Tii, Vi, LJii
Icelandic	0.8598	1.47	0.8029	2.19	Ti, Tii, Vi, LJii
Spanish	0.7072	4.39	0.6120	3.91	Ti, Ti, LJii
Swedish	0.7224	2.22	0.7158	2.39	Lii, Liii, Ti, Ti, LJii

Table 5.10 presents

- the number of measurements in the dataset ($N_m = N_{vowels} * N_{speakers} * 44$ for a dataset with no missing values);
- the number of free parameters in the unconstrained PARAFAC model ($N_{fp} = (N_{vowels} + N_{speakers} + 44) * N_{factors}$)
- the number of measurements modeled in the model with articulatory primes. The unmodeled measurements are excluded from this number - i.e., if there are no larynx primes in the final model, the measurements of larynx position are not included.
- the number of free parameters in the model using articulatory primes ($N_{fp} = (N_{vowels} + N_{speakers}) * N_{primes}$);
- the sum of squared errors (SSE) for each model. These quantities are given for each

language. The statistics pertaining to the best unconstrained PARAFAC model of the vowel articulation in that language are presented in the left-hand half of the table, and the statistics for the finally selected model of the vowels using the eight articulatory primes that appear to make a substantial contribution to the language are in the right-hand portion of the table.

From these quantities, the χ^2 of each model is estimated by dividing the SSE by an estimate of the measurement error in the data. I am using 3 mm^2 as an estimate of the mean squared measurement error, based on the results from Harshman et al. (1977) for most of the data. However, since the Chinese x-ray tracings are approximately 2/3 the scale of the other tracings, I am using $(2/3)^2(3) = 1.33$ mm^2 as the estimate of the expected mean squared measurement error for the Chinese models.

Some indication of the sensibility of 3 mm^2 as an estimate of the expected mean squared error can be gained by looking at model errors cited in Liljencrants (1971, 1985) and in Sekimoto et al. (1978). Liljencrants presents rms errors for his model on a per-token basis; overall, the mean squared error in his model is about 3.4 mm^2. Sekimoto et al. report an rms error of 1.7 mm, i.e., a mean squared error of about 2.9 mm^2 over all their speakers and tokens. A good compromise between Harshman et al. (1977), Liljencrants (1971, 1985), and Sekimoto et al. (1978) seems to be in the neighborhood of 3 mm^2.

The bottom columns of Table 5.10 show the χ^2/d.f. ratio and the R^2 of the models. The d.f. of a model is the difference between the number of data points fit and the number of free parameters in the model: $N_m - N_{fp}$. The χ^2/d.f. ratio is used to get a rough indication of the fit per degree of freedom of the models. This quantity is not suitable as a primary indicator of fit: rather, it states something about the efficiency with which a model achieves whatever fit it achieves (Wheaton 1987, p.128). Low values of χ^2/d.f. indicate relatively 'efficient' models. R^2, on the other hand, is a relatively good indicator of overall fit.

When fitting raw measurements, the χ^2 statistic converges to a normal distribution with its mean equal to the d.f. of the model. Thus, the χ^2/d.f. ratio should have a lower bound of 1.0 when considering models fit to raw data (the situation is different when fitting covariances). A value lower than that should make us suspect overfitting.

The summary information in Table 5.10 suggests that our procedure has been moderately successful in constructing a cross-linguistically valid model of vowel articulation. The R^2s for models using articulatory primes are 2-5% worse than R^2s of

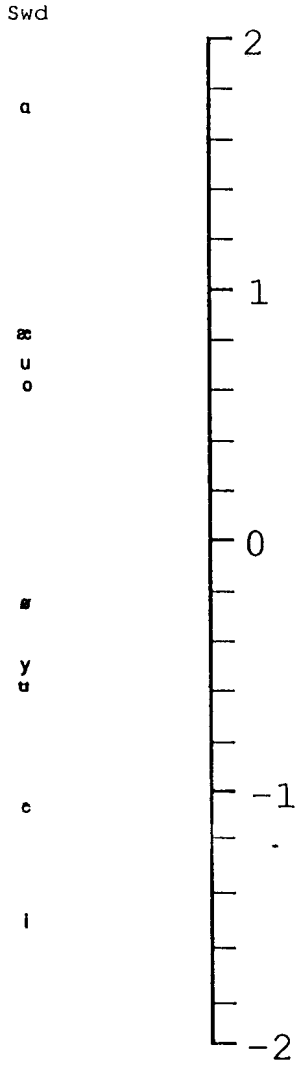


Figure 5.4. Swedish vowels on the Lii axis.

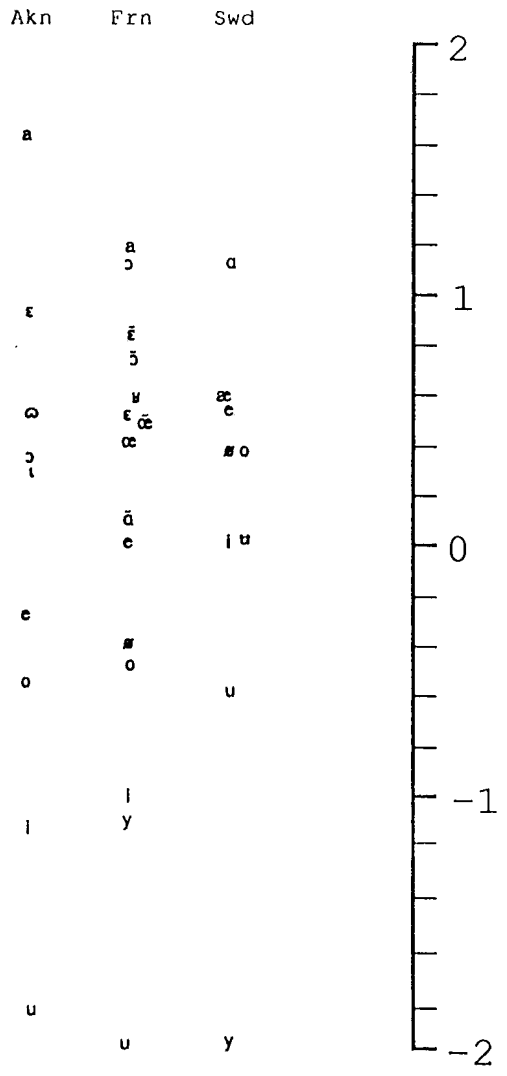


Figure 5.5. Akan, French, and Swedish vowels on the Liii axis.

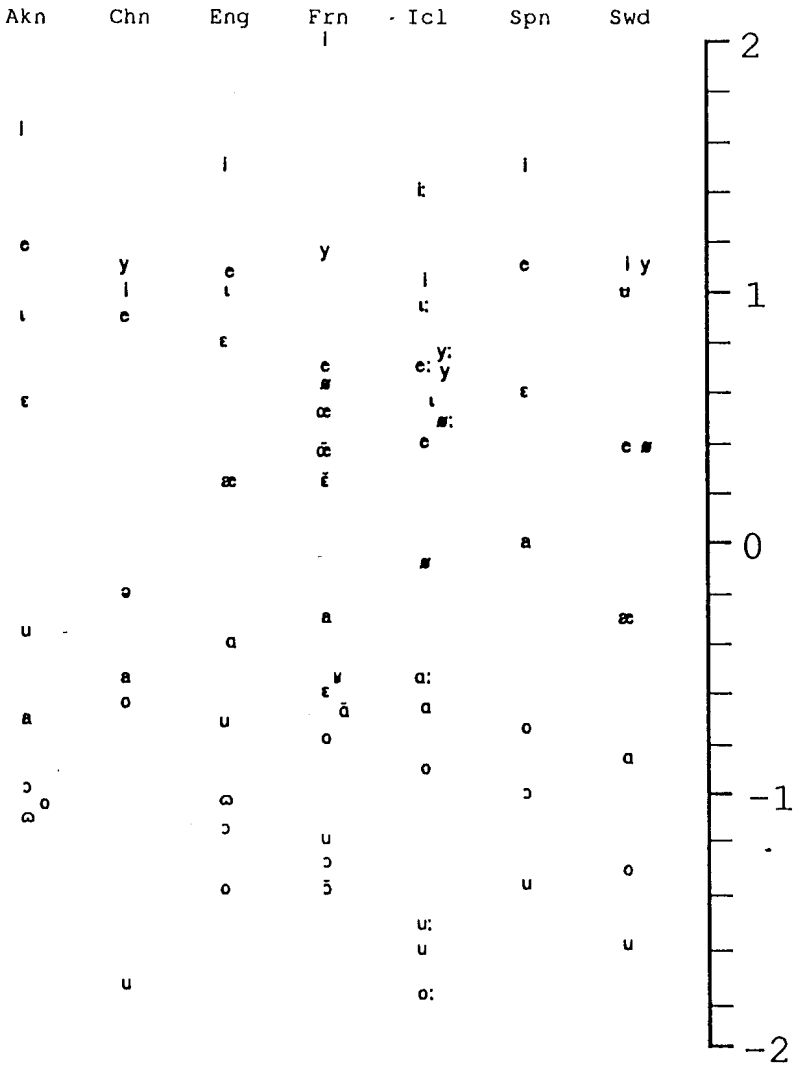


Figure 5.6. Akan, Chinese, English, French, Icelandic, Spanish, and Swedish vowels on the Ti axis.

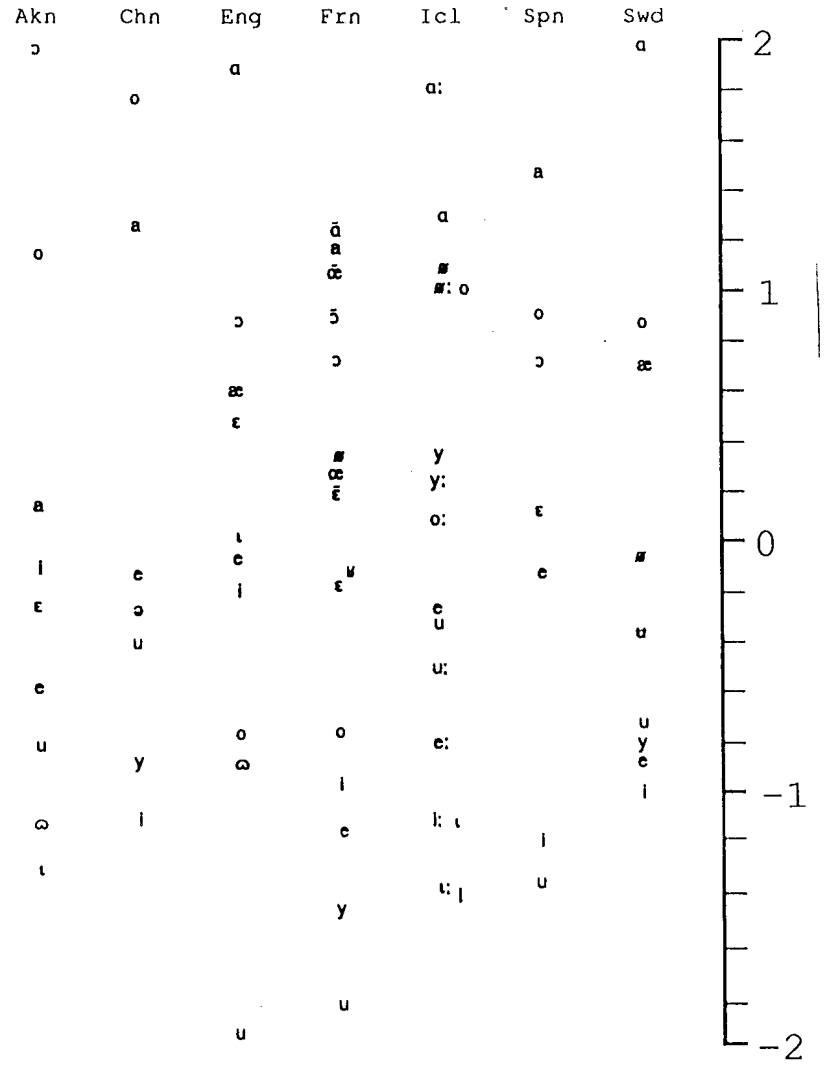


Figure 5.7. Akan, Chinese, English, French, Icelandic, Spanish, and Swedish vowels on the Tii axis.

the unconstrained PARAFAC models, except for Spanish. Given that most of these models do not fit all of the measures of articulator position (e.g., Akan, with no V(elum position) prime, and Spanish, with no L(arynx) prime, in their final models), these decreases in overall fit do not seem too large to take.

The χ^2/d.f. ratios in Table 5.10 are also encouraging. At least half of the models using articulatory primes in the table are 'more efficient' at fitting the proportion of the variance that they fit than the corresponding unconstrained PARAFAC models. However, the articulatory prime model for Chinese does look as though it might be overfit.

The estimated contributions of these articulatory primes to the vowels of the various languages are plotted in Figs. 5.4 through 5.11. Fig. 5.4 plots the vowels of Swedish on the Lii axis. The negative end of this scale represents low larynx positions combined with dorsal movement of the epiglottis; the positive end represents high larynx positions combined with an anterior position of the epiglottis. This prime seems to make rather different contributions to the rounded and unrounded (non-low) front vowels of Swedish. The larynx is apparently consistently raised for all of the rounded vowels and /ɑː/.

Fig. 5.5 plots the vowels of Akan, French, and Swedish on the Liii axis. The negative end of this axis indicates low larynx position, posterior movement of the epiglottis, and anterior movement of the dorsal wall of the pharynx. In Akan, this prime clearly separates the [+ATR] set /**i, e, o, u**/ - low larynxes - from the [-ATR] set /**ɪ, ɛ, a, ɔ, ʊ**/ - high larynxes. In French, this prime puts high vowels with the Akan [+ATR] vowls and low vowels with the [-ATR] vowels. The one speaker of Swedish who makes substantial use of this articulatory prime appears to follow a similar pattern.

Fig. 5.6 shows the Ti axis for all the vowels of all the languages that we have discussed here, together with the vowels of English. (The measurements of tongue reported in Harshman et al. 1977 were fit to Ti and Tii in the same manner as the tongue position measurements from the other languages). As we have remarked over and over again, /**i**/-like vowels pattern at one extreme of the Ti axis. At the other extreme lie /**ɔ**/- or /**o**/-like vowels and the /**u**/-like vowels of Chinese, Spanish, and Swedish. The /**u**/-like vowels of Akan, English, French, and possibly Icelandic, on the other hand, seem to be less extreme on the Ti scale than the /**ɔ**/- and /**o**/-like vowels of those languages. /**a**/- and /**æ**/-like vowels generally fall near the middle of this axis of the vowel space.

The vowels of all the above languages are plotted on the Tii axis in Fig. 5.7. One extreme of this axis has /**a**/- or /**ɑ**/-like vowels in most of the languages but /**o**/-like vowels Akan and Chinese. The other has /**i**/-like vowels in Chinese, Icelandic, and Swedish;

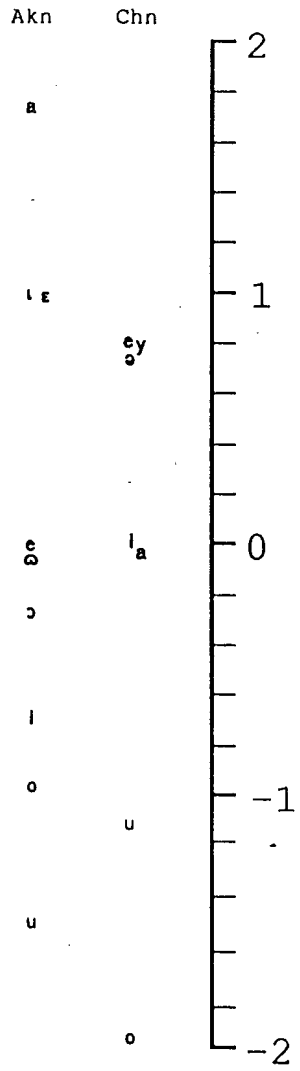


Figure 5.8. Akan and Chinese vowels on the Tiii axis.

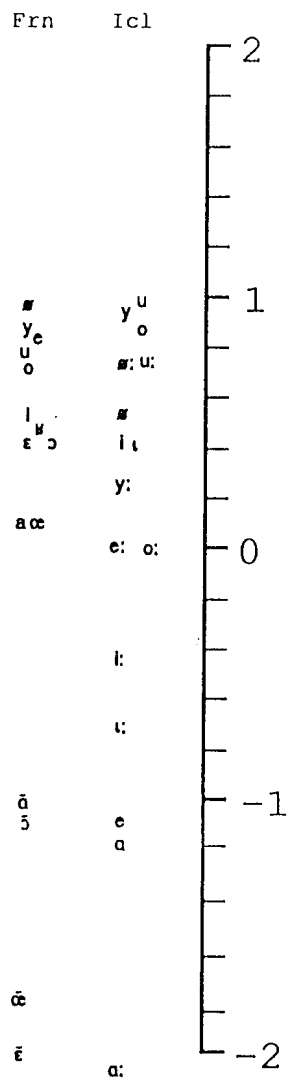


Figure 5.9. French and Icelandic vowels on the Vi axis.

/u/-like vowels in English, and Spanish; and both /i/- and /u/-like vowels in Akan and French.

Quilis (1981) observes that "... el /e/ español es mucho más abierto que el /e/ cardinal, sin llegar al /ɛ/ cardinal ... el /o/ español se aproxima, sin llegar al /ɔ/ cardinal, siendo algo más cerrado que este último ..." (p. 175). Taking these remarks one at a time, we see that Spanish [e] and [ɛ] are both further towards the /ɑ/-like end of the Tii scale that the French [e] which Armstrong (1967, pp. 39-40) describes as nearly cardinal. Even though Kenyon (1956, p. 61) and Jones (1956, p. 357) describe American English [ɔ] as lower than cardinal; and Armstrong (1967, pp. 50-51) describes French [ɔ] as centralized, our analysis agrees with Quilis in placing the Spanish [o] and [ɔ] close to the American English and French [ɔ]s, and not close to the [o]s.

The rankings of the French front rounded vowels /y, ø, œ/ is also generally consistent with their traditional phonetic description. They generally rank closer to the origin of the vowel space (on the Ti and Tii scales) than the corresponding front unrounded vowels /i, e, ɛ/. Armstrong (1967, pp. 59-65) also describes the front rounded vowels as "retracted" and lower than their corresponding unrounded vowels.

These plots also suggest to me that there are two kinds of /u./ One kind of /u/ has relatively small contributions from Ti but conversely fairly large ones from Tii - this is the vowel in Akan, English, and possibly French. The other kind of /u/ has fairly large contributions from Ti but tends to have smaller ones from Tii - this is the vowel in Chinese, Icelandic, Spanish, and Swedish. It is perhaps more like a very close English /o/.

It has been noted before that American English /u/ is not a particularly back vowel - Jones (1956) remarks that "Many Americans ... use a more advanced variety ... resembling the 'crooner's u" (p. 358). Similarly, Armstrong (1967) notes that French /u/ is "slightly advanced" (p. 56). Disner's (1983, pp. 64, 106) comparative acoustic study shows that /u/ in Swedish is acoustically backer than the /u/s produced by the American English speakers studied by Peterson & Barney (1952), supporting the articulatory hypothesis advanced above. The assignment of Spanish /u/ to the same category as Swedish /u/ is supported by Quilis' (1981) observation that "... el /u/ español es más posterior y algo más abierto que el /u/ cardinal" (p. 175.)

Fig. 5.8 shows the Tiii axis. Since only Akan and Chinese appeared to use this articulatory prime, only these two languages' vowels are plotted. In Akan, the contributions of this prime to particular vowels again reflect the [+/-ATR] distinction, though they do so less clearly then Liii. The Chinese /u/ and /o/ group with the Akan [+ATR] vowels, and the other Chinese vowels appear to group more or less with the

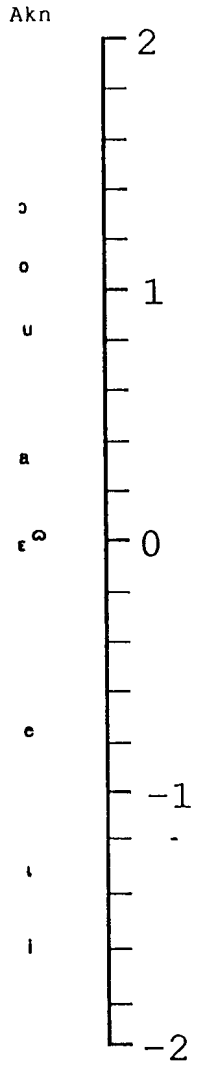


Figure 5.10. Akan vowels on the LJi axis.

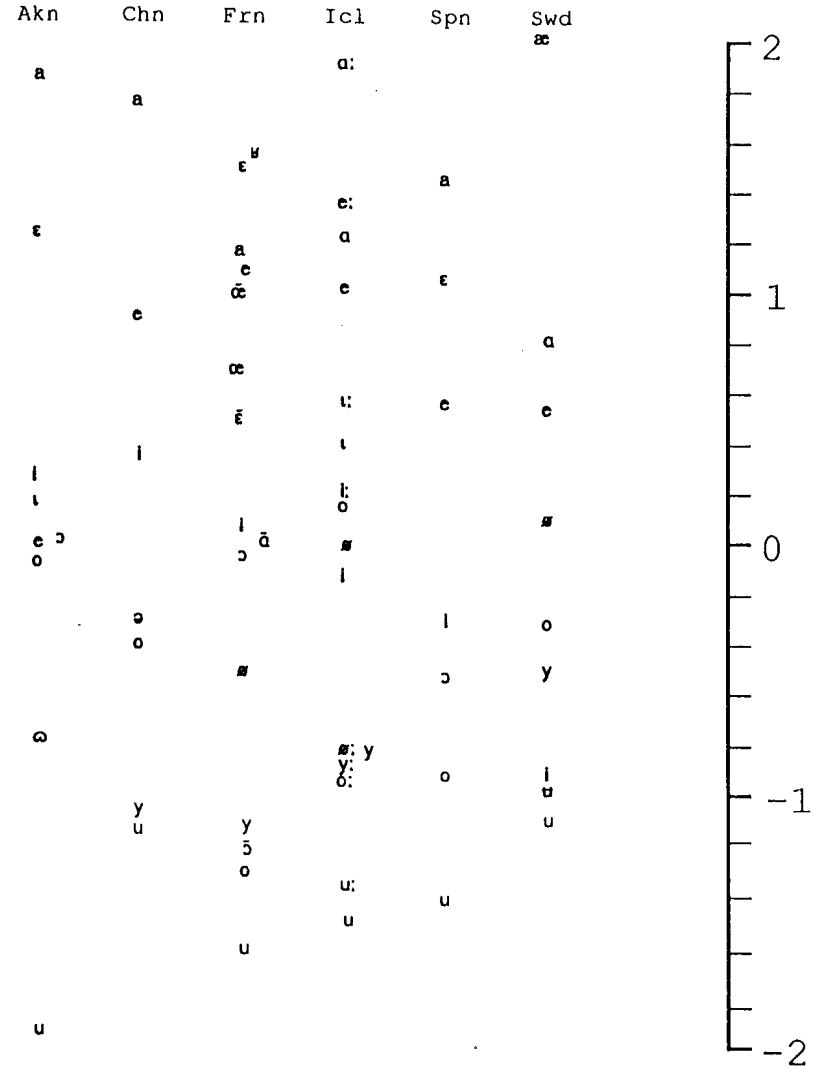


Figure 5.11. Akan, Chinese, French, Icelandic, Spanish, and Swedish on the LJii axis.

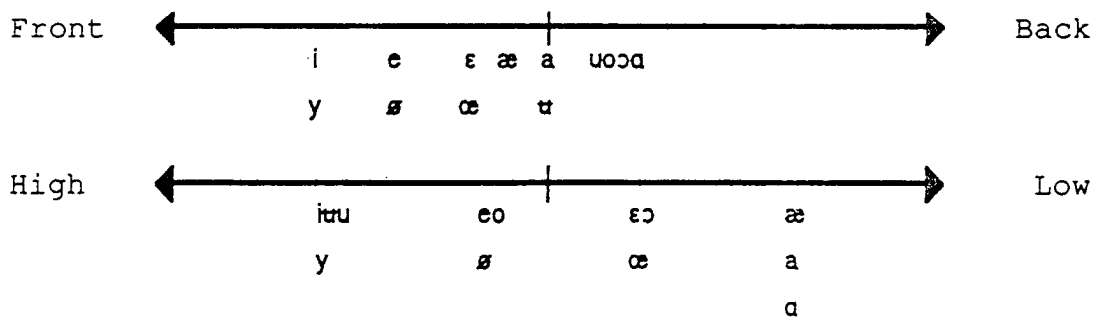


Figure 5.12. The IPA axes.

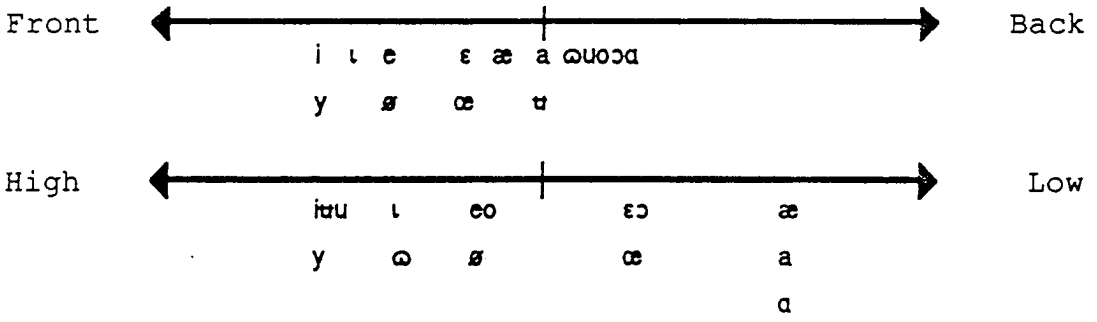


Figure 5.13. Additional vowels on the IPA axes.

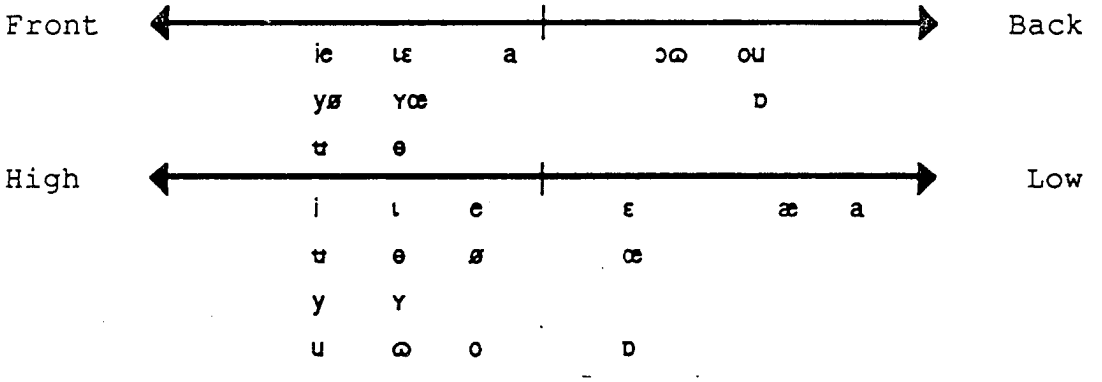


Figure 5.14. Lindau's (1978) classification of vowels.

[-ATR] vowels.

Fig. 5.9 plots the vowels of Icelandic and French on the Vi axis. In French, this prime contributes the velum lowering to nasal vowels. There is also apparently some tendency for there to be velum lowering in low nonnasal vowels such as /**a**/ and /**œ**/. In Icelandic, the vowels /**ɑ:**/ and /**ɑ**/ evidently have lowered velums as well (indeed, Pétursson 1974b describes them as having velar opening but no perceptible nasality). In both French and Icelandic, the velum position for /**u**/-like and /**y**/-like vowels appears to be higher than it is in general for /**i**/-like vowels.

Fig. 5.10 plots the vowels of Akan on the LJi axis, and Fig. 5.11 plots the vowels of all six languages on the LJii axis. The LJi prime contributes jaw raising and lip retraction to the unrounded Akan vowels /**i, ɪ, e, ɛ**/, the converse to the lower rounded vowels, and little to the higher rounded vowels and /**a**/. The LJii prime contributes jaw raising and lip protrusion to most languages' /**u**/-like and /**y**/-like vowels, and jaw lowering and lip retraction to /**a**/- and /**æ**/-like vowels. This prime tends to contribute little to /**i**/- and /**ɔ**/-like vowels.

Traditional descriptions vs. articulatory primes

Some of these observations roughly agree with other studies. For instance, Linker's (1982) study of lip rounding in vowels produced scales quite similar to our LJii scale in several languages. Similarly, the ranking of the vowels of Akan on the Tiii and Liii axes is similar to that proposed Lindau's (1978) discussion of expanded/constricted pharynx vowels.

Where our account differs the most from other accounts is in the description of tongue positions. There are a number of competing proposals for the articulatory description of the space of possible vowels.

Ignoring for the moment the arguments against the front-back / high-low framework as an articulatory description of vowels, let us consider two versions of this framework. The most traditional version of this framework, presented as two scales similar to the ones we have seen above, is depicted in Fig. 5.12 (c.f. Fig. 1.1). It is based on my estimate of where some of the vowel species that we have discussed above would be placed on a chart like that in the *Principles of the International Phonetic Association* (IPA 1949). The first axis is the front-back dimension of tongue position: the front vowels are strung out over a fairly long region, with /**i**/ being the frontest and /**ɐ**/ the least front. (Note that modern practice would generally use /**æ**/ for the vowel quality denoted by IPA /**a**/). The back vowels are all located in a fairly compact bunch, with /**u**/ the least back and /**ɑ**/ the most. The second axis is the high-low dimension: the vowels are grouped into four major

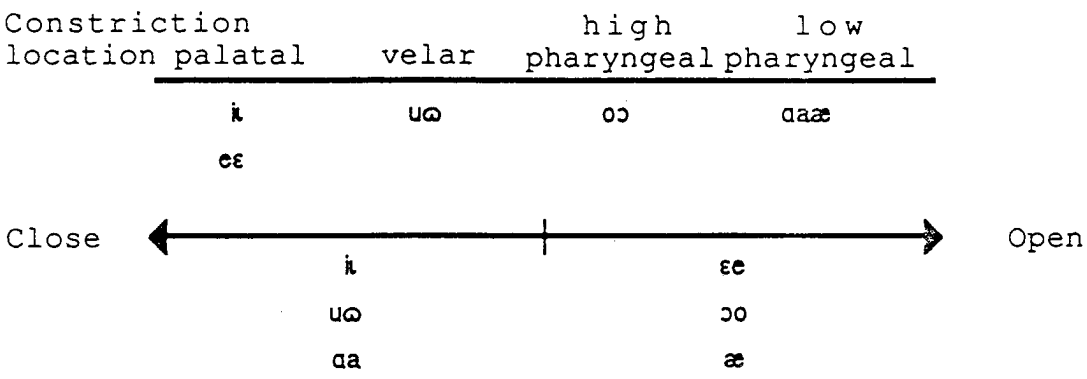


Figure 5.15. Wood's (1982) classification of vowels.

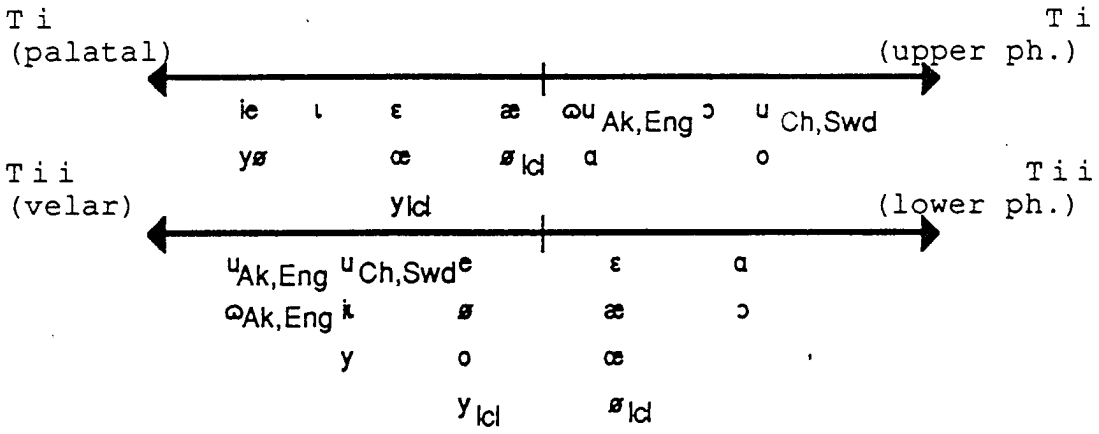


Figure 5.16. Vowels classified according to their loadings on articulatory primes.

height categories, and inside each category the front vowels tend to be higher than the back ones.

A few additional vowels - /ɪ, ʊ/, the lax counterparts of /i, u/ - are shown in Fig. 5.13. As we noted in Chapter 1, Jones (1956) was suspicious of tense/lax as an articulatory parameter, and believed that tension was only applicable to high vowels (p. 40). Jones' own practice (e.g., FIgure 1.1b) was to describe the 'lax' vowels as centralized versions of the 'tense' ones (also see Lindau 1978 on 'peripheral/central').

Lindau's (1978) system greatly resembles this one in doing away with tension as an articulatory dimension and recognizing additional gradation along the front-back and height dimensions. The peripheral/central contrast puts vowels like /ɪ/ and /ʊ/ a 'half-step' closer to the origin of the chart than /i/ and /u/. Lindau uses phonological patterns to argue that the vowel chart should be organized roughly as in Fig. 5.14. There are three major categories along the front-back dimension and four major categories along the high-low dimension. These extremes of these dimensions are in turn split by peripherality - thus /i/ and /u/ are more peripheral than /ɪ/ and /ʊ/. As far as I can tell, Lindau (1978) only considers the difference between /i, u/ and /ɪ, ʊ/ to be one of peripherality *per se*; the difference between /e, o/ and /ɛ, ɔ/ could also be a difference of peripherality, but Lindau discusses it directly in terms of height (p. 545). Lindau's system is also explicitly supplemented by expanded/constricted pharynx and other articulatory parameters, making it the most comprehensive proposal to date.

Running somewhat against the tide, Wood (1982) reintroduces tension as an articulatory parameter. He argues on combined phonological and articulatory grounds that the vowel space has three major classificatory parameters: constriction location, openness, and tension. Constriction location goes from the palatal end of the vocal tract - /i/ - to the lower pharynx - /a/. Thus, some of the vowels that are regarded as front in the proposals above, such as /æ, a/ (as produced by Wood's one speaker of SBE and one speaker of Cairo Arabic) wind up classified at the other end of the scale from the front vowels /i, e/. The back vowels are also redistributed, though somehow the effect of splitting /u/ and /o/ into different but adjacent categories is less disturbing.

Wood's proposal, schematized in Fig. 5.15 also differs from all of the above proposals in placing /i/-like and /u/-like vowels near each other on both the constriction location and the openness scales. The proposals above only put /i/ and /u/ together on the height scale. Openness is otherwise like a binary catergorization of height. Distinctions that aren't made on the openness scale - such as /i/ vs. /ɪ/ - are represented as differences in tension.

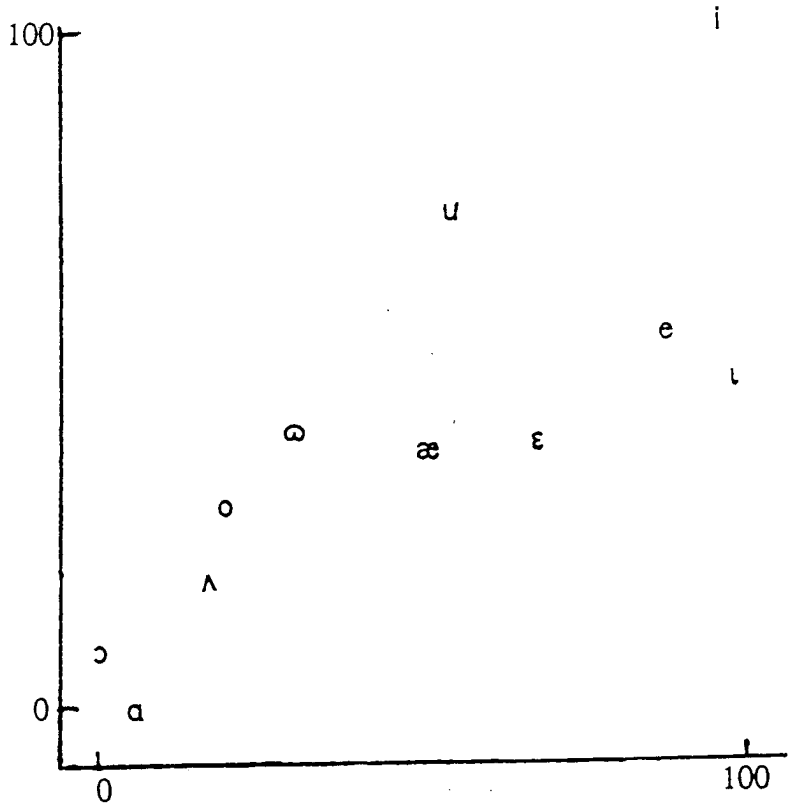


Figure 5.17. Tongue body positions in American English vowels (after Nearey 1978).

Fig. 5.16 presents a similar condensation of the system developed in this dissertation. The Ti scale is, perhaps, somewhat in between Wood's constriction location axis and the IPA chart's front-back axis in its ordering of vowels. By luck or design, we have also recognized two different kinds of vowel conventionally written /**u**/: they are labelled with language names as subscripts in the figure. The Icelandic vowels conventionally written /**y**/ and /**ø**/ are also rather different in quality from the French and Swedish vowels written the same way, and so they are also subscripted.

The Tii scale differs from all of the proposed height and openness scales proposed above. One kind of /**u**/ ranks highest on this scale - as opposed to the IPA representation, in which /**u**/ (and back vowels in general) ranked *lower* that its front counterpart /**i**/. Furthermore, it seems as though /**æ**/ and /**a**/ do not generally rank as low as /**ɑ**/ on the Tii scale. This constrasts with all of the above proposals, which put /**æ, a, ɑ**/ at the same height/openness.

Nearey (1978) quantitatively compared tongue body positions, as determined from the center of a circle osculating the midsagittal profile of the tongue, to the height-backness framework for describing vowels. Fig. 5.17 shows his plot of the height and backness of the tongue body in eleven vowels of American English as produced by three speakers. It can be seen that this plot is very skewed compared to the IPA chart (Fig. 1.1). The front-to-back ordering of the vowels is not unlike that proposed by Wood (1982) and in the current work. The high-to-low ordering is, however, more like that of the IPA system (Fig. 5.12) - /**i**/ is higher than /**u**/; /**ɪ**/ higher than /**ʊ**/; /**e, ɛ**/ higher than /**o, ɔ**/.

Rossi (1983) interprets data from Maeda (1979), Nearey (1978), and Zerling (1979) as showing that there is an 'axe vélaire' going from /**ɔ**/ through /**o, ʊ**/ to /**u**/; and an 'axe alvéolaire' going from /**ɑ**/ through /**æ, ɛ, e**/ to /**i**/. Rossi then reclassifies /**ɑ**/ and /**æ**/ as being "pharyngales ... des axes vélaire et alvéolaire (p. 100)", waffling towards a more traditional position. Zerling (1979) identifies the 'axe alvéolaire' as characterized by increased genioglossal and decreased hyoglossal activity (or vice versa), and the 'axe vélaire' as characterized by increased activity of the styloglossus (and palatoglossus). In any case, the effect is to classify the front 'alvéolaire' and back 'vélaire' vowels by either their degree of stricture or their height.

After a look at another way to produce maps of the vowel space (Chapter 6), we will return to the question of the explanatory value of these different parameterizations of the vowel space in Chapter 7.

Chapter 6: An alternate methodology: investigating the vowel space directly

The aim of this chapter is to propose a method for investigating the vowel space directly. In the previous two chapters, we have been concerned primarily with the articulatory space - the space of vocal tract configurations generated by various combinations of articulatory primes, and subspaces generated by particular combinations of articulatory primes used in specific languages. This method has only allowed an indirect investigation of the vowel space, our hypothesized underlying parameterization of this articulatory space. The goal of the alternative methodology proposed in this chapter is to investigate the parameterization of the vowel space directly.

Bentler, Poon, & Lee (1987) show how to use EQS to estimate the model proposed by Tucker (1966) for data such as ours, in which the various facets of classification are systematically varied. Harshman & Lundy (1984a, p. 173 ff.) show that PARAFAC may be regarded as a Tucker model with certain constraints. As a preliminary step towards the estimation of the model given in equation (2.6) (repeated below as 6.1), I have attempted to combine the content of these two works, in order to allow PARAFAC-like models to be estimated using the EQS program.

$$\mathbf{Y^{*T}Y^{*}} = \mathbf{V^{*}\,S^{*T}\,\Phi\,S^{*}\,V^{*T}} \qquad (6.1)$$

In (6.1), $\mathbf{V^{*}}$ is a matrix built up by placing as many copies of $\mathbf{V}_1$ (the matrix containing the coordinates in the vowel space of language 1's vowels) corner-to-corner as there are speakers of language 1, followed by as many copies of $\mathbf{V}_2$ as there are speakers of language 2, etc. $\mathbf{S^{*}}$ is built up by placing diagonal matrices $\mathbf{S}_{k(l)}$ of speaker-specific scaling coefficients side-to-side. Φ is a matrix that represents the correlations of the articulatory loadings of the various factors. Thus, in this model, the matrix product $\mathbf{Y^{*T}Y}$ is parameterized in terms of speaker-specific scaling coefficients (the $\mathbf{S}_{k(l)}$), a matrix that represents the similarity of the various articulatory primes to each other (Φ), and - most importantly - the coordinates in the vowel space of the vowels of the individual languages (the $\mathbf{V}_l$).

A pilot study

Since this method has not been used before, I conducted a pilot study using the data from Harshman et al. (1977) to investigate the degree to which PARAFAC, applied to raw data; and EQS, applied to covariance matrices, yield analyses consistent with each other. This pilot study requires a slight reformulation of (6.1). The required reformulation of the PARAFAC model is identical to equation (2.5), with only one level

of *l*, i.e., only one language. This special case of that model is given in (6.2), below.

$$\mathbf{Y^*} = (\mathbf{Y}_1\ \mathbf{Y}_2 \ldots \mathbf{Y}_k) = \mathbf{A}\,(\mathbf{S}_1\ \mathbf{S}_2 \ldots \mathbf{S}_k)\begin{pmatrix} \mathbf{V}^T & \phi \\ & \ddots \\ \phi & & \mathbf{V}^T \end{pmatrix} \tag{6.2}$$

The form of the within- and cross-speaker cross-product matrix (analogous to equation 2.6) is given in (6.3).

$$\mathbf{Y^{*T}Y^*} = \mathbf{V^*}\,\mathbf{S^{*T}}\,\Phi\,\mathbf{S^*}\,\mathbf{V^{*T}} \tag{6.3}$$

The elements of the cross-product matrix $\mathbf{Y^{*T}Y^*}$ are the dot products of the vectors of the measurements of articulatory position of each speaker-vowel combination. Each dot product represents a measure of the articulatory similarity between some vowel produced by some speaker and some other vowel produced by some other speaker. This cross-product form is transformed into a within- and cross- speaker covariance matrix by removing the mean measurement in each speaker-vowel combination ("centering within fibers", in Harshman & Lundy's terminology).

The data presented and analyzed in Harshman et al. (1977) consists of measurements of tongue position along thirteen gridlines for ten vowels of American English as produced by five speakers. This data array was first centered in the vowel mode (by subtracting the within-speaker and -gridline means) as in Harshman et al.'s original study. The array was then centered in the gridline mode (subtracting means calculated *across* gridlines and *within* speakers and vowels) to bring the cross-product form of (6.3) to a variance-covariance matrix. The centered data (a 13 gridline x 10 vowel x 5 speaker array) was analyzed using the PARAFAC program and the speaker-vowel covariance matrix (50 speaker-vowel combinations x 50 speaker-vowel combinations) was analyzed using EQS. Multivariate normal distribution of the variables was assumed in EQS, and least-squares estimation, rather than the statistically preferred maximum-likelihood method, was used because the small number of measurements relative to the size of the covariance matrix guaranteed the singularity of the covariance matrix.

Both programs converged to well-determined two-factor solutions within 40

iterations. The PARAFAC solution has an R^2 of 0.9464, and the EQS solution has a Bentler-Bonett normed fit index of 0.906. These measures of fit are not simply related: the R^2 is calculated from the mean squared error of the fit to the raw data, whereas the Bentler-Bonett normed fit index is calculated from the squared error of the fit to the covariances. However, these fit values are comparable.

The PARAFAC solutions under single- (as in the original study) and double-centering (as in the comparision with EQS) of the data are very similar, replicating the Front Raising and Back Raising factors. The PARAFAC utility program CMPARE was used to compare the solution obtained under the two conditions. The correlations between the loadings on the various factors ranged from 0.968 to 1.00 as shown in Table 6.1.

Table 6.1: Correlations between the two PARAFAC solutions

Mode	FR	BR
Gridlines	1.000	0.999
Vowels	0.997	0.996
Speakers	1.000	0.968

These correlations indicate that the overall shape, or profile, of the factors was nearly identical in the two solutions. The cross-products of the factors (after normalization to unit length) is equivalent to the correlation if there is no difference in factor elevation (i.e., a non-zero intercept to the regression line) between the two solutions. The cross-products are tabulated in Table 6.2.

Table 6.2: Cross-products between the two PARAFAC solutions

Mode	FR	BR
Gridlines	0.999	0.943
Vowels	0.997	0.996
Speakers	1.000	0.940

It can be seen that some of the cross-products for the Back Raising factor are rather different from the corresponding correlations in Table 6.1. There is therefore some difference in the mean loading of the Back Raising factor on the gridlines and speakers. The measurement gridlines are roughly radial over a large part of the tongue (as can be verified from the example of the grid presented in Harshman et al., or from the sample grid, constructed by a similar method, shown in Chapter 3), so such a difference would correspond to a roughly radial contraction/bunching or dilation/flattening of the tongue

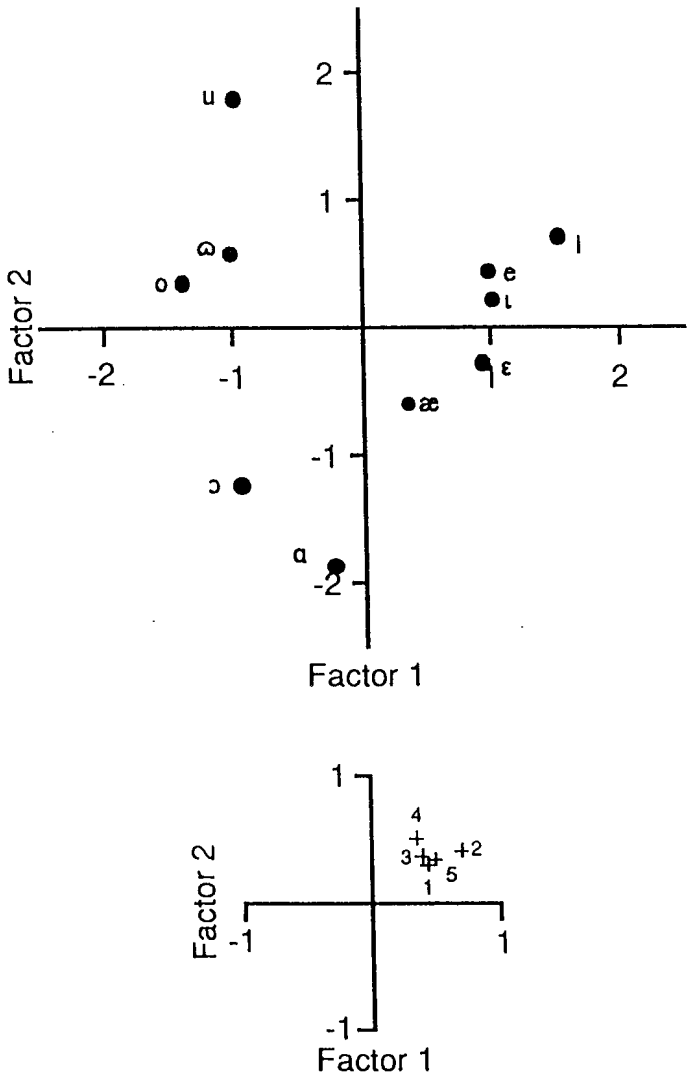


Figure 6.1a. Double-centered PARAFAC solution for English vowel tongue positions.

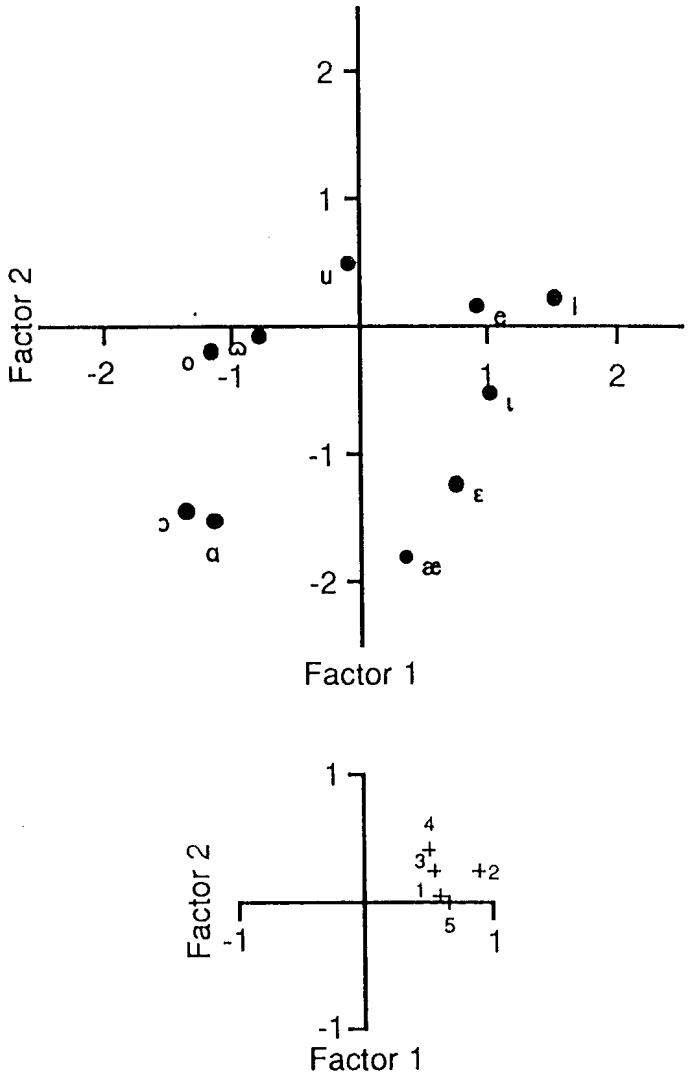


Figure 6.1b. EQS solution for English vowel tongue positions.

displacements analyzed by the PARAFAC program. This difference is not unexpected, given that the operation of centering the data across gridlines has removed exactly this kind of mean difference. We may conclude that the solutions found by the two different analytic procedures are essentially the same.

EQS, in fitting the model in (6.3), yields a correlation matrix that estimates the correlations across gridlines between the two factors (Φ). This correlation is -0.853 according to EQS; the same correlation, calculated from the PARAFAC analysis of the double-centered tongue position data, is -0.791 (for comparision, the correlation calculated from the single-centered Harshman et al. solution is -0.744). We can see that even if the two procedures yielded equivalent coordinate systems for the vowel space, EQS has estimated a pair of dimensions in the articulatory space that are rather more similar in their effect on tongue position that those found by PARAFAC.

The CMPARE utility was also used to compare the vowel and speaker loadings estimated by EQS to those estimated by PARAFAC from the double-centered data. The correlations, summarized in Table 6.3, are generally not as high as in the previous comparision.

Table 6.3: Correlations between the EQS and PARAFAC solutions

Mode	FR	BR
Gridlines	-	-
Vowels	0.904	0.882
Speakers	0.991	0.923

Although the factors found by the two procedures are clearly similar, they are not as similar as those examined in the previous comparision. Examining the cross-products (Table 6.4) shows that Front Raising in the two solutions is probably practically identical - the cross-product is about equal to the correlation. However, the cross-products of the Back Raising factors found by PARAFAC and EQS are rather different from the correlations, suggesting that there is some sort of shift in the origin (or mean position) of the factor space. A plot of the vowel and speaker loadings from the two solutions (Fig. 6.1a - the double centered PARAFAC solution; 6.1b - the EQS solution) confirms this impression. There is some rotational difference between the two solutions, and it appears that all the vowels, especially [ɔ] and [**u**] score lower on the Back Raising scale in the EQS solution.

Table 6.4: Cross-products between the EQS and PARAFAC solutions

Mode	FR	BR
Gridlines	-	-
Vowels	0.904	0.711
Speakers	0.999	0.865

As we noted in Chapter 1, the Front Raising factor of tongue position appears to be much more stable under different analytic procedures than subsequent factors.

To test whether the two solution found equivalent vowel spaces, but simply located the axes of the coordinate system in slightly different orientations, as Fig. 6.1 suggests, I constructed a multiple regression model relating the loadings in the PARAFAC solution to the vowel loadings in the EQS solution.

Table 6.5: R^2 for PARAFAC loadings as a function of the EQS loadings

Mode	FR	BR
Gridlines	-	-
Vowels	0.911	0.805
Speakers	0.985	0.866

As can be seen from the high proportion of the variance accounted for by these multiple regression models, the spaces described by these two solutions, while not identical, are quite similar.

I consider this high degree of correspondance encouraging. The EQS solution describes a space that is very similar to the one described by the PARAFAC solution, despite the rotational difference between the two solutions. It remains to be shown whether or not the solutions can be brought into greater correspondance by changing some of the default statistical assumptions of EQS.

A partial analysis of cross-language data

The results from this pilot study suggested that running the full set of data from Akan, Chinese, and French vowels was impractical. There would be 56 French vowel tokens (14 vowels x 4 speakers), 21 Chinese vowels (7 x 3), and 36 Akan vowels (9 x 4), producing a 113 x 113 covariance matrix. The pilot study, with a 50 x 50 covariance matrix, used an average of 1.02 megabytes of memory and ran for some 115 seconds on the UCLA IBM 3090. The Akan-Chinese-French dataset would require over four times as much memory (since a 113 x 113 matrix has more than four times as many elements as a 50 x 50 matrix). The storage required to represent the special structure of the **V***

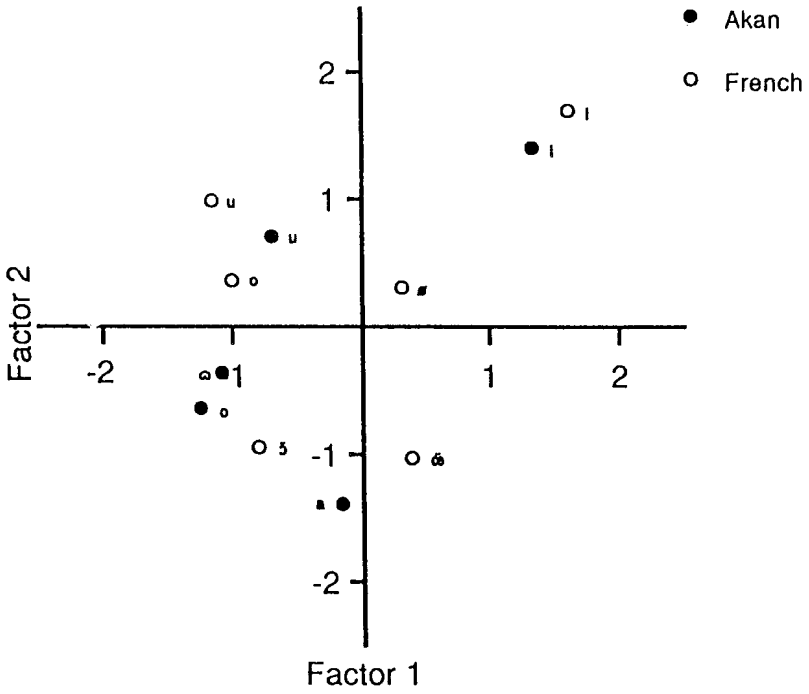


Figure 6.2. 2-factor EQS solution for selected Akan and French vowels.

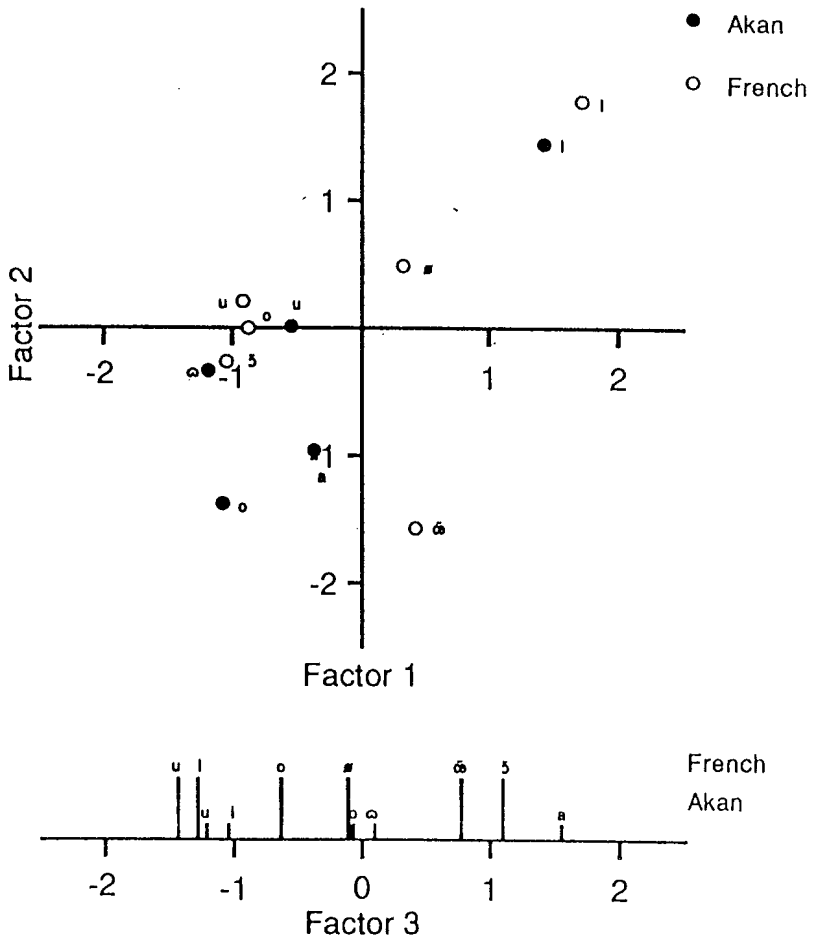


Figure 6.3. 3-factor EQS solution for selected Akan and French vowels.

matrix (see equation 2.5) would also be about four times as large as that required for the **V** matrix of the pilot study. Given the fact that the time required for matrix operations such as inversion and multiplication typically increases as the cube of the number of rows or columns of the matrix, the solution would have run at least eight times as long!

In order to prune the covariance matrix down to a reasonable size, I selected several vowels that loaded highly on the various factors of the within-language solutions, since such vowels should contain most of the variance that the factors account for. I used the Akan vowels /**i, a, o, ɔ, u**/ and the French vowels /**i, ø, o, u, œ̃, ɔ̃**/. The Akan set thus includes three [+ATR] vowels /**i, o, u**/ and two [-ATR] vowels /**a, ɔ**/. The French set includes three front vowels /**i, ø, œ̃**/, three back vowels /**o, u, ɔ̃**/, four non-nasal vowels /**i, ø, o, u**/, and two nasal vowels /**œ̃, ɔ̃**/. The selection as a whole has three unrounded vowels and seven rounded vowels. All tokens of these vowels, as produced by the four Akan and four French speakers, were used, giving a 44 x 44 covariance matrix (5 vowels x 4 Akan speakers + 6 vowels x 4 French speakers). However, this reduction in sample size also probably reduces the number of significant factors that can be reliably estimated.

EQS requires initial estimates to be supplied for the free parameters to be estimated. For the runs described below, I used the vowel and speaker loadings from the within-language PARAFAC solutions for the initial estimates of the **V*** and **S*** matrices in equation (2.6). The initial estimates of the off-diagonal elements of the Φ matrix were taken from the correlations between the first, second, etc., factors of cross-language solution.

Trial runs using the model in equation (2.6) with from two through four factors did not converge in less than 60 iterations (twice the usual number of iterations the the EQS program runs for). Examination of the intermediate outputs showed that all the solutions had high correlations (0.9 and higher) in the off-diagonal elements of the Φ matrix. This suggested that the solutions had drifted into much the same kind of degeneracy that we have noted occurs in PARAFAC models. Another fact that suggested this conclusion is that the vowel loadings diverged quite substantially from the original estimates, with the smallest loadings nearly 1/60th the magnitude of the largest loadings.

In order to improve the estimability of the model, I also imposed the constraint that the factors be orthonormal in the articulatory mode (i.e., the Φ matrix of equation (2.6) was set to **I**). Solutions were obtained at two and three dimensions and normalized in a manner parallel to the usual normalization for PARAFAC solutions.

The vowel spaces yielded by the two solutions are plotted in Figs. 6.2 and 6.3. The two-dimensional EQS solution (Fig. 6.2) resembles many of the vowel spaces we have

seen before. The first factor (the horizontal axis of Fig. 6.2) groups /i/ vowels at one extreme and /o/- or /u/-like vowels at the other, just like the within-language solution from Akan (Figs. 4.1 and 4.2) or the solution from the pilot EQS study reported above (Fig. 6.1b). The second factor in this solution, however, is somewhat different from that seen in the other solutions. In Figs. 4.2 and 6.1a, the /u/s definitely score higher on the second factor than the /i/s. But in the two-dimensional EQS solution, the /i/s score higher than the /u/s. Indeed, this plot overall resembles the IPA vowel chart (Fig. 1.1) much more than most of the other plots we have seen!

Two dimensions of the three-dimensional EQS solution (Fig. 6.3) are basically the same as those found in the two-dimensional solution. The third factor groups the [+ATR] vowels of Akan against the [-ATR] vowels. All the [+ATR] vowels in this subspace have negative loadings on this factor, and all the [-ATR] vowels have positive loadings, although the tense /o/ and lax /u/ are both quite close to a loading of zero. In French, this factor groups the non-nasal vowels against the nasal vowels /œ̃/ and /ɔ̃/. However, these factor interpretations should be taken with the caution appropriate to the small number of vowels used in the analysis. For instance, the second factor could be interpreted as being related to jaw position, and the third factor could be interpreted as some sort of general tongue height parameter.

If the interpretation of the third factor as representing [ATR] in Akan but [nasal] in French is correct, it is probably an 'accidental' consequence of the fact that [+/-ATR] and [+/-nasal] are both articulatorily quasi-orthogonal to most other articulatory gestures. [+/-ATR], as we saw in the PARAFAC analysis of the Akan data (Fig. 4.3), involves mostly movement of the epiglottis and root of the tongue. Nasalization is produced by movements of the velum. Both of these movements combine fairly freely with other articulatory gestures. Furthermore, [+/-ATR] is not systematically exploited in French, and [+/-nasal] is not systematically present in this sample of Akan vowels. The EQS procedure has therefore been able to interpret the third factor in a manner 'nested' within the languages, picking different major correlates of the factor in each language. This result emphasizes the similarity of the model in (2.6) to a nested ANOVA design.

In conclusion, this exploratory application of EQS to the task of articulatory modeling seems to be only mildly successful. The orthogonalized model fit by EQS yields about the same vowel space that the analysis of covariances by PARAFAC does. However, further research into the reasons for the poor convergence of the model in (2.6) is clearly needed.

Chapter 7: Phonetic theory and phonological processes

In the previous chapters, we developed a theory and methodology for investigating the articulation of vowels. Assuming that the position of various articulators during a vowel's production may be described as a weighted combination of articulatory primes, we have investigated the vowel inventories of a number of languages. Subsidiary hypotheses investigated in this context include conjectures about the language-specificity of the organization of articulatory primes into coordinative structures, and the degree to which this functional language-specific organization may be modeled with simple linear dependencies.

We have constructed a model which may be summarized as in Fig. 7.1. At the top of Fig. 7.1 is a vowel space, a language-specific representation of vowel articulation in terms of the contribution of various coordinated gestures to the vowel. Representations in this vowel space are mapped through a measurement space to specific articulatory configurations which are represented as linear combinations of various articulatory primes. The articulatory primes are the independent articulatory gestures which we investigated in Chapters 4 and 5. When one parameter of the vowel space controls several articulatory primes, we say that those primes form a coordinative structure.

We have investigated the vowel spaces of several individual languages. As we have had occasion to remark several times, the vowel spaces of individual languages differ systematically from traditional notions about the organization of the vowel space in terms of (features or) scales of height and backness. The typical result is that the axes of the phonetic vowel space are tilted with respect to the traditional and phonological classification. Thus, the most reliably replicated axis of the phonetic vowel space generates tongue shapes similar to the Front-Raising tongue factor of Harshman et al. (1977) and puts /**i**/- and /**o**/-like vowels at opposite extremes of the vowel space. This family of vocal tract positions is generated by an articulatory prime that I have called Ti.

The purpose of this chapter is to begin to address the question of whether or not, and why, this framework for describing vowels is preferable to other frameworks. This task is not trivial: phoneticians and phonologists have devoted a great deal of effort to their descriptive mechanisms, and they might well already be as good as possible. Thus, evidence favoring one framework over another tends to be difficult to find. Even worse, it is often difficult to assess, because it often depends on rare, or poorly described phenomena, or on tendentious and uncertain historical inferences about earlier stages of languages. The apportionment of the descriptive burden between phonetic and phonological theories is also a difficult area.

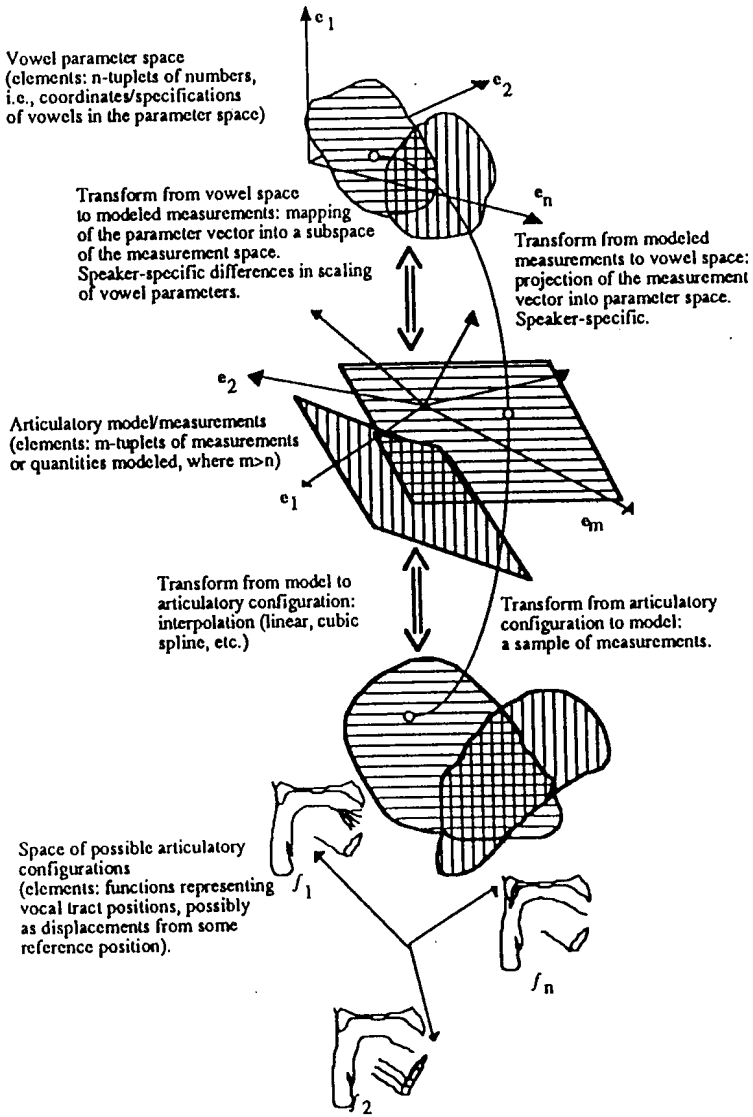


Figure 7.1. Summary of the articulatory model.

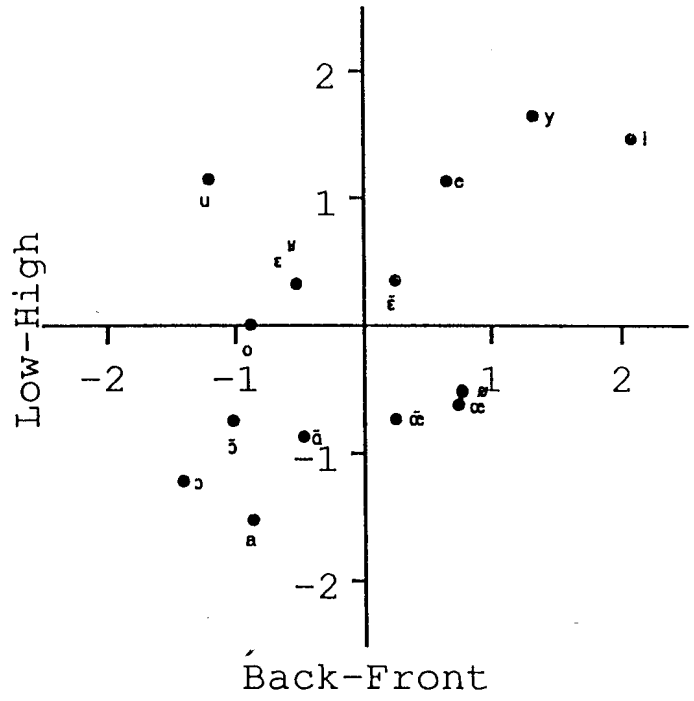


Figure 7.2. Height-backness relaxed Procrustean French vowel space.

Descriptive desiderata for a phonetic theory

A phonetic theory such as the one that we have constructed for vowel articulation should be able to accomplish several tasks. It is of course responsible for describing the phonetic facts of vowel articulation. It is also responsible for providing primitives for the description of phonological patterning, just as it provides primitives for the coordinative structures that implement phonological goals.

Some problems with the height-backness framework

Some of the defects of the traditional height-backness system for describing vowel articulation were detailed in Chapter 1. Height and backness have been criticized as articulatory descriptive parameters because they fail to accurately describe many of the details of vowel articulation. It is clear that the articulatory primes that we have developed do a good job of describing articulatory configurations in vowels (see, e.g., Table 4.8).

It can also be shown that the vowel loadings derived from articulatory primes model the articulatory configurations better than analogous loadings derived from the coordinates of the vowels in the height-backness system. For instance, the fit figure (R^2) of the relaxed Procrustean fit to Front Raising, Back Raising, Nasal and Round loadings is 0.7349. A similar relaxed Procrustean fit, starting with loadings derived by measuring the x- and y- coordinates of the French vowels on a height-backness diagram (derived from Armstrong 1967), produced the vowel space shown in Fig. 7.2. This model does not fit the data as well as the model derived from estimated Front Raising and Back Raising loadings ($R^2 = 0.7333$). As a comparision of Fig. 7.2 with Fig. 4.9 will show, the vowel loadings derived from height-backness estimates have relaxed to a configuration similar to that derived from Front Raising and Back Raising. In an unrelaxed height-backness model, the fit value would be even lower. The height-backness framework thus is not only incorrect in detail (as shown by the workers cited in Chapter 1) but also quantitatively poorer than a model based on articulatory primes.

Some problems with constriction location and degree

Having discussed some of the deficiencies of the height-backness system for describing vowel articulation, I would now like to discuss the framework proposed in Wood (1982) and point out several deficiencies in it. Wood (1982) argues that

- vowels exhibit only a limited number of places of constriction, due to the limited ability of the vocal tract musculature to position the tongue.
- these places of constriction are palatal, velar, upper pharyngeal, and lower pharyngeal.
- vowels may be either close or open at a particular place of constriction.

In addition, Wood recognizes tense-lax as a classificatory parameter.

In my opinion, this kind of view has several deficiencies. First, and perhaps least serious, articulatory configurations (or even just tongue shapes) associated with vowels may in fact show several constrictions. It has been known for quite a while that in Akan, for instance, the [-ATR] vowels have constrictions in the lower pharynx as well as in the palatal or velar region. Similar observations have been made with respect to rhotacized vowels (Lindau 1985). However, it might be possible for a proponent of Wood's position to simply allow vowels to be described in terms of their degree of constriction at each of the various locations that Wood lists.

A more serious deficiency that is true of both Wood's original proposal and the minor modification I have sketched is that not all of the possibilities for constriction at various positions of the vocal tract are independent. An [i]-like constriction in the palatal region always goes along with expansion in the upper pharyngeal region. The converse is also true: [o]-like constriction in the upper pharynx always induces lowering of the blade of the tongue away from the palate. Wood's original proposal - that vowels are organized into families, each associated with a single location of vocal tract stricture - of course predicts this result. For instance Wood (1971, p.86) considers the feature value [+pharyngeal] to be incompatible with the feature value [+palatal] in his description of West Greenlandic vowel allophony adjacent to uvulars. In his account, vowels assimilating to an adjacent uvular stop or continuant are assigned [+pharyngeal] and by convention changed to [-palatal].

But the contention that pharyngeal constriction is incompatible with palatal constriction is clearly incorrect. [i]-like tongue shapes *can* be combined with lower pharyngeal constriction, as in Cairo Arabic pharyngealized front vowels (c.f. Norlin's (1987) review of the x-ray literature and modeling study on Arabic); or tongue root and epiglottal constriction as in Akan [-ATR] vowels (c.f. Lindau 1979). The position that vowels should be described by one, and only one, constriction location apiece cannot be maintained in general.

One could object that the tongue root gesture in [+/-ATR] vowels was in some sense 'secondary', and thus not subject to the same exclusivity that the 'primary' constrictions (palatal, velar, upper pharyngeal, lower pharyngeal) are subject to. But there are also vowels that apparently fall in between these major constriction families of Wood's: French [y, ø] and Chinese [ə], to take two examples. This defect is shared by other frameworks that presuppose a strict division of articulatory configurations into 'palatal' and 'velar '(eg., Rossi 1983).

It is thus necessary to allow the independent specification of *some* combinations of vocal tract constrictions - but not all! The particular constraints needed to make Wood's

system work appear rather arbitrary - why allow free combination of palatal and lower pharyngeal constrictions, but not palatal and upper pharyngeal constrictions? In the model I have developed, however, these constraints arise naturally out of the articulatory primes. It is simply not possible to raise the front of the tongue, creating a constriction in the palatal region, without also pulling the tongue forward and creating an expansion in the upper pharynx by using the Ti prime. The other primes do not generate much movement of the blade and tip of the tongue. But Ti generates little displacement of the tongue in the lower pharynx, allowing a language or a speaker to do whatever they want down there. Rather than having stipulated constraints on the combinatorial possibilities of constriction locations, we have primitives - articulatory primes - that simply cannot generate certain possibilities.

Phonological primitives from phonetic descriptions

These phonetic theories differ in the primitives that they make available for the purpose of describing phonological patterning. The traditional height-backness continua have long been used for phonological description. Lindau (1978) argues, for instance, that phonological theories need to recognize four categories of height. These categories are 'discretized' versions of the continuous phonetic parameter that underlies the vowel space. Lindau (1978) also argues that phonological theory requires three discrete backness categories based on the phonetic backness continuum.

Wood (1971, 1982) argues that the constriction locations - palatal, velar, upper pharyngeal, and lower pharyngeal - are precisely the primitives available to the phonological theory. Since, in his theory, the constriction locations are discrete categories, there is no 'discretization' of the phonetic categories in the translation from phonetic representation to phonological representation.

In a description based on articulatory primes, we must recognize the same kind of discretization of a continuous phonetic parameter into a small number of phonological categories. The primitives that the current proposed phonetic parameterization makes available to phonological theories should thus be some number of categories along the Ti, Tii, Tiii, etc., axes.

Phonological evidence for phonetic categorization

There are several common kinds of phonological rule in the languages of the world that - in part because of their commonness - are generally considered to be phonetically motivated. We might list among such rules nasal assimilations to adjacent stops, assimilations in consonant clusters, voicing assimilation, and so forth. Here, I will discuss one class of such putatively phonetically motivated rules: umlaut rules. Umlaut rules generally considered to be the result of some *assimilatory* process, i.e., a process in which

a *target* vowel acquires some of the articulatory characteristics of some nearby *trigger*.

Umlaut rules have been considered to be prototypically phonetically motivated rules since at least the neogrammarian theories of language change of the 19th century. Modern theories of phonological segment structure and rule formulation often lump umlaut rules together with other vowel harmony processes, admitting at most that umlauting is *bounded* (applying over some restricted domain, such as one syllable), whereas harmony rules are unbounded (see, e.g., Anderson 1980).

However, despite their formal similarity, harmony rules are generally not phonetically motivated in the same way that umlaut rules are. For instance, harmony processes show exceptions, such as vowels 'neutral' to the harmony process, e.g., /**i**/ in Khalkha Mongolian rounding harmony; /**i**/ in Hungarian backing harmony; and /**i, y, e, œ**/ in Finnish palatal harmony (Anderson 1980) of a kind that umlaut processes do not. Harmony processes are also typically morphologized in a way that historical umlaut processes are not (although the modern descendants of historical umlauts are often morphologized). Such morphologized rules cannot be expected to reflect the phonetic conditions that gave rise to them; and indeed I am not aware of any generally accepted historical account of the origin of the harmony systems of Uralic and Altaic languages such as Finnish, Hungarian, and Mongolian. Nor is there any generally accepted account of the origin of [ATR] harmony systems. Therefore, our present knowledge of the origin of harmony rules is too limited to provide evidence about the phonetic motivations of phonological processes. Umlaut rules, on the other hand, are better known.

An inventory of umlaut rules

Table 7.1 presents an inventory of umlaut rules gleaned from the literature. First, a crude typology: the first five rules in the table are triggered by /**i**/ or the /**i**/-like glide /**j**/. I will refer to these rules, naturally, as i-umlaut rules. These rules typically front and raise some vowel near to this trigger. (The Old English rules are provided as specific examples of processes that ocurred in the protohistory of all the Germanic languages but Gothic).

The next group - from Montañes Spanish, Menomini, Early Germanic, and Rumanian - are rules that assimilate the height of the target vowel to that of the trigger. The first two *raise* the target vowel; the other two *lower* it.

Then there three processes that assimilate the backness (or frontness) of the target vowel to that of the trigger. The Chamorro rule is quite productive, whereas the historically attested Germanic rule appears to have been sporadic. We will have discuss the Old English u-o-a Umlaut rule more extensively below.

Finally, there are three rules that assimilate the target vowel's rounding to that of the trigger. This kind of rule may be triggered by either /**u**/ (Icelandic) or /**o**/ (Lowland

Murut).

Table 7.1: An inventory of phonological umlaut rules

Language	Process	Example
1. Modern Uighur (after Anderson 1974)	V -> [-low,-back] / [+low] __ C_0 **i** [-stress]	
2. Latvian (after Halle & Zeps 1966)	**æ** -> **e** / __ C_0 **{i, j}**	
3. Takelma (after Anderson's (1974) discussion of Sapir 1922)	**a** -> **i** / __ [-stem] C [+voice] **i**	
4a. proto-Germanic (all regions but but Gothic; after Hockett 1958)	V -> [-back(,-low)] / __ C **{i, j}**	Gothic **haf jan** vs. Old Icelandic **hef ja** 'to raise'
4b. Old English i-Umlaut (after Cassidy & Ringler 1971, p. 39 except as noted)		
ɑn, ɔn	**ɛn**	
æ	**ɛ**	
œ	**œ** (Barrack 1975, p. 61)	
ɛ	**ɪ** (Barrack 1975, p. 62)	
u ->	**y** / __ C **{i, j}**	
ɔ	**ɛ**	
ɑ	**æ**	
ɛə,ɛɔ,iɔ	**iɛ**	
5. Warlpiri (after Nash 1979)	u -> i / i C_0 __ [-labial]	
6. Montañes Spanish (after McCarthy 1984)	V -> [αhigh] % __ C_0 [-low,+stress,αhigh]	
7. Menomini (after Bloomfield 1939)	V: -> [+high] / [-low] __ C_0 **{i, u, j, w}**	

8. Early Germanic Lowering (Schane 1984, p. 134)

i -> **e** / __ C {**e, a, o**}
u -> **o**

9. Rumanian Breaking (*ibid.*)

e -> **ea** / __ C {**e, ə, a**}
o -> **oa**

10. Chamorro (after Topping 1968)

V -> [-back] / [-back] # C_0 __ **oksɔʔ** 'hill',
gi eksɔʔ 'at the hill'

11. Backing umlaut (sporadic) (after Hockett 1958)

V -> [+back] / __ C u pre-Old English /**sibun**/
-> */**sifon**/
-> OE /**səfon**/ 'seven'

12. Old English u-o-a Umlaut (after Cassidy & Ringler 1971, p. 40)

æ -> **æə**
ɛ -> **ɛə** / __ C {**u, ɔ, ɑ**}
i -> **iə** [+stress]

13. Javanese (after Steriade (to appear) analysis of Dudas 1975)

ɑ -> **ɔ** / __ (C) **ɔ** # **dɑnɔwɔ** 'giant'

14. Lowland Murut (after Steriade to appear)

/**a**/ -> /**o**/ / __ C_0 /**o**/
[+stress]

15. Old Icelandic (after Anderson 1972)

ɑ -> **ɔ** / __ C_0 u

Many of these rules are stated in terms of the traditional height-backness framework. In this framework, the height umlaut process of Montañes Spanish, Menomini, Early Germanic, and Rumanian are easily described. The rounding umlaut rules (Javanese, Lowland Murut, and Old Icelandic) are also easy to describe. But the numerous i-umlaut rules are difficult to describe in a unified manner. Some phonological processes (e.g., Uighur, Latvian, Takelma) involve apparent raising triggered by /i/ or /**j**/. Others (e.g., proto-Germanic, Warlpiri) apparently involve fronting. The Old English i-umlaut rule requires *both* raising and fronting, but of different target vowels: /**æ, ɛə**/ raise to /**ɛ, iɛ**/, respectively; and /**u, ɔ**/ front to /**y, ɛ**/.

Another, more theory-bound problem attested to in this inventory of rules is the frequency with which the segments /i/ and /j/ - the class [-back,+high] trigger rules, as compared with either [-back]- or [+high]-triggered rules. Under some proposals about the way in which phonological theory should represent 'natural' or 'common' (i.e., 'unmarked') rules this is a surprising fact. If we believe that classificatory features such as [high] and [back] represent natural classes that tend to act together in rules (e.g., as triggers), and if we believe that phonologies minimize the number of such features that must be specified in a rule (as was tentatively proposed in Chomsky & Halle 1968), then we would expect classes such as [-back] or [+high] to trigger rules more frequently than the intersection of those classes represented by [-back,+high].

Even under more recent feature-hierarchy proposals, such as those advanced in Clements (1985) *et seq.*, this asymmetrical distribution of rule triggers is surprising. In such proposals, the features [high] and [back], and others, are typically considered to form a constituent, and thus a natural object (target or trigger) for an assimilation rule of the sort summarized in Table 7.1. But even if the particular *pair* of features [high,back] is thereby expected to occur as commonly as each of the features individually, the particular selection of feature *values* [+high,-back] remains a mystery.

In the Wood's (1982) constriction-based framework, some i-umlauts are described rather simply by assignment of [+palatal]. In Latvian for instance, the vowel /æ/ [+pharyngeal,-close] is umlauted to /e/ [+palatal,(-pharyngeal,)-close]. Similarly, in Takelma, the vowel /a/ [+pharyngeal,+close] is umlauted to /i/ [+palatal,(-phayrngeal,),+close]. Other i-umlauts are described by assignment of [+close]: for instance, the Modern Uighur rule can be translated into the constriction-based framework as assignment of [+close] (and possibly [+palatal] as well). i-umlauts are thus described naturally are assimilations of the target vowel to the triggering palatal vowels.

The Old English i-umlaut rule requires both [+palatal] and [+close] to be assigned in particular cases; in most, only one of the two can be assigned. /æ/ [+pharyngeal,-close] is converted to /ɛ/ by assignment of [+palatal]. /ɛə/ is converted to /iɛ/ (in part) by assignment of [+close]. These changes are naturally represented as assimilations to the [+palatal,+close] trigger /i/.

However, in Wood's (1971, 1982) framework, changing /a/ to /æ/ must be represented by assignment of [-close], since /æ/ is the openest vowel of the pharyngeal constriction family. This putatively assimilatory umlaut to a [+close] vowel is represented counterintuitively by assignment of [-close]!

Some height umlauts (e.g., Menomimi) are easy to represent in Wood's framework

as changes in the value of the feature value [close]. However, it is very difficult to represent the class of segments that trigger the lowering rules of Early Germanic and Rumanian Breaking. For instance, the triggering environments for Early Germanic Lowering include both /e/ and /o/ - /e/ is [+palatal,-close], but /o/ is [+upper pharyngeal,+close] (or [+uvular,+close]). This conjunction of classes is difficult to represent. The change in Early Gemanic Lowering is also difficult to represent: /i/ to /e/ requires assignment of [+close]; /u/ to /o/ requires assignment of [+upper pharyngeal]!

The formulation of Old English u-o-a umlaut in Wood's constriction-based phonetic and phonological framework is also problematic. The inserted /ə/ offglide must presumably be represented as a constrictionless vocalic segment in this system. Why the [+velar] /u/, the [+upper pharyngeal] /ɔ/ and the [+lower pharyngeal] /ɑ/ should act together to *remove* constriction specifications from [+lower pharyngeal] /æ/ - a *dissimilation* from /ɑ/, not an assimilation! - is difficult to explain.

Part of this problem is resolved by observing (Barrack 1975, p. 24-25; following Campbell 1964) that the backing umlaut in question probably only occured before /o:, o/; /o:/ subsequently changed to /ɑ/. Thus, "it would seem more probable ... that back umlaut functioned before *u/*o* but failed when **o* became *a*" (Barrack 1975, p. 25). If this conjecture is correct, then the relevant triggers for this rule are only /u, ɔ/, and the rule - if not an outright assimilation - is at least intuitively something like 'remove a feature specification ([+palatal], or [+lower pharyngeal]) that conflicts with the features of the trigger'. One must still ask why [+lower pharyngeal] /ɑ/ was not affected in the same way that [+lower pharyngeal] /æ/ was.

This constriction-based framework is similar to the height-backness framework in that the commonest trigger for umlauts - /i/ - is represented as the intersection of two sets: [+palatal,+close]. This frequency with which this particular feature combination triggers assimilations is surprising for the same reasons that the frequency with which [+high,-back] triggers assimilation is surprising.

One the other hand, /i/ is fairly simple to decribe in terms of the articulatory primes proposed in this study. /i/ is always the vowel at the positive end of the Ti scale (except in languages like Chinese and Swedish, where front rounded vowels like /y/ are right next to /i/). Thus, the representation of /i/ is simpler in this framework than in the height-backness or constriction-based frameworks. If we believe that phonetic theory should provide simpler representations to commoner processes (by analogy to the 'ink-counting' metric tentatively proposed in Chomsky & Halle 1968), then the representational simplicity of /i/ as the extreme vowel on the Ti parameter of the vowel

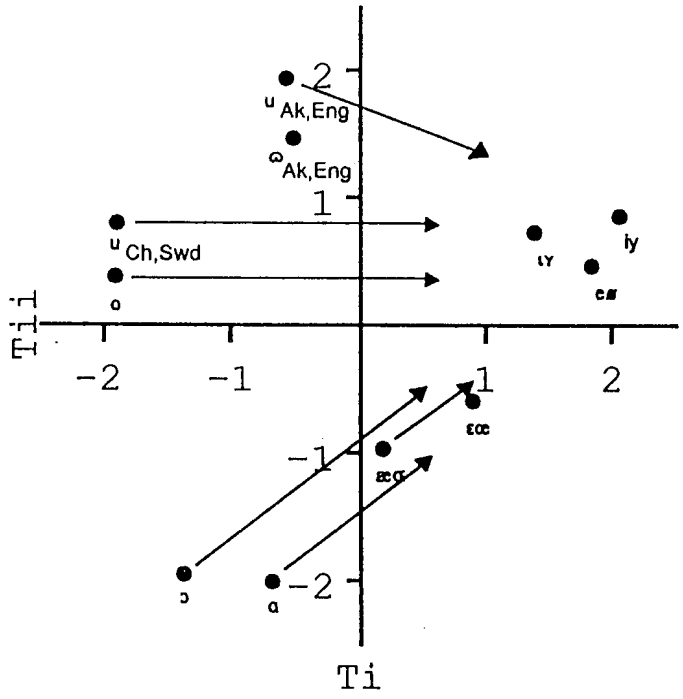


Figure 7.3. Ti-Tii plane of the model vowel space and i-umlaut.

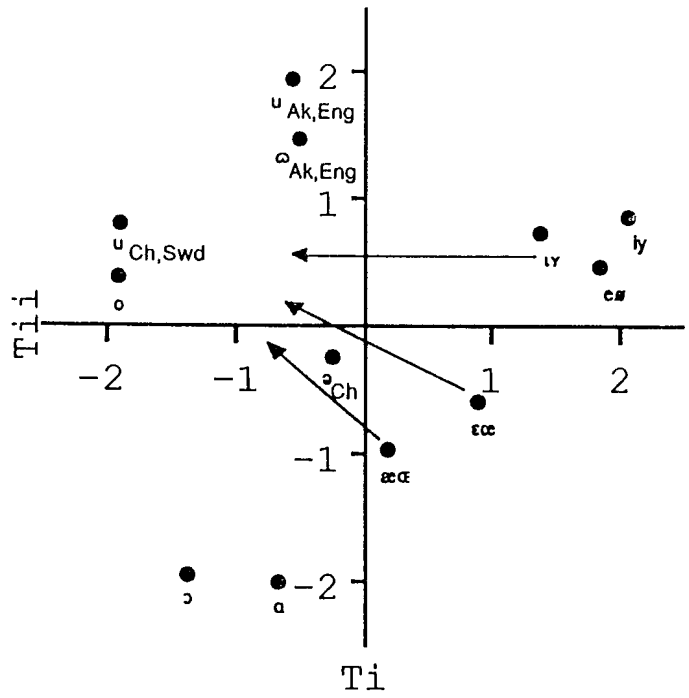


Figure 7.4. Ti-Tii plane of the model vowel space and u-umlaut.

space should be preferred over the representation of /i/ as the vowel that is simultaneously the most high and the most back. Similarly, the description of /i/ in terms of articulatory primes should be preferred to its description in terms of constriction location and degree.

We have discussed how i-umlaut must be represented by changes of both the height and backness, or by changes of the location and degree of constriction, of target vowels. Again, the changes in vowel quality associated with i-umlaut are simpler to describe in terms of articulatory primes than in the other frameworks. Fig. 7.3 plots the Ti-Tii plane of the schematized axes given in Fig. 5.16. In Fig. 7.3, the assimilatory changes due to i-umlaut are represented by the arrows. If we accept the idea that i-umlaut requires the quality of the target vowel to change until it is some moderate distance from the triggering /i/ (about 1.5 arbitrary units in this case), the assimilations in Figue 7.3 result. Notice that back vowels like /ɑ, ɔ/ move to be near to /æ, ɛ/ - backness changes; and that low front vowels like /æ, œ/ move towards /ɛ, œ/ - height changes. Thus, the changes in vowel quality which are representationally complex in the height-backness framework are represented here in a unified, simple manner.

Height-changing rules (e.g., Menomini, Early Germanic, Rumanian) are also represented fairly simply in terms of articulatory primes. The Tii axis is in fact very similar to a height parameter (c.f. Figs. 5.14, 5.16), and so these changes can be represented either by changes parallel to the Tii axis (vertical in Fig. 7.3; e.g., Early Germanic Lowering) or by changes towards the triggering vowel (e.g., Rumanian Breaking).

Old English u-o-a umlaut can also be captured in this framework. Again, following Barrack (1975) and Campbell (1964), we can assume that /u/ and /o(ː)/ triggered this rule. Assuming additionally that the relevant stage of Old English had an /u/ like that of Swedish and Icelandic, rather than an /u/ like that of modern (American) English, the recorded vowel changes would be similar to those depicted in Fig. 7.4. As we can see, the arrows all end near the origin of the vowel plane. This region is indeed where vowels like the Chinese /ə/ plot in the Ti-Tii plane, in agreement with the reconstructed results of this process.

Phonetic motivation

Why should it be that these umlaut processes occur at all? I can only offer some speculation as to the ultimate phonetic motivation for these processes. Above, we have discussed umlaut rules as assimilatory processes, in which a target vowel's quality drifts over towards that of the trigger vowel. But these processes could equally be seen as ones in which the trigger vowel's quality spreads into the target vowel, as in autosegmental

phonological treatments of umlaut. A similar viewpoint is represented in Schane's (1984) particle phonology: the vowels /**i, u, a**/ are seen as both instances of particular vowels and as operators that trigger the changes in other vowels. The prevalance of particular triggers - notably /**i**/ - suggests that certain articulatory gestures are more prone to spread that others.

Another insight into the nature of these assimilations comes from the phonetic literature on coproduction of segments (e.g., Ohman 1966). In particular, it has been suggested that slow articulatory gestures, e.g., those involving the tongue body, are coproduced with nearby segments. The slower gestures required for the production of some segment tend to overlap in time with the quicker gestures associated with the nearby segments. Another way of putting it, of course, is that the articulatory state that is the goal of the gesture spreads into the nearby consonants and vowels.

i-umlaut is probably the result of this phonetic tendency for the spreading of slow gestures. Modeling studies like that of Coker (1967) and Mermelstein (1973) have shown that the motions associated with the contraction of the genioglossus and lips are slow compared to other motions in the vocal tract. Coker (1967) cites a time constant of 150 ms for tongue body raising and 300 ms for lip protrusion; these time constants indicate gestures much slower than tongue tip raising (50 ms; Coker 1967) or jaw movements (jaw lowering - 100 ms, jaw closure - 125 ms: Kim & Fujisaki 1974; jaw lowering - 75 ms: Mermelstein 1973). This contraction of the genioglossus is essential for the attainment of /**i**/-like tongue positions (see, e.g. the modeling study Kakita et al. 1985 or the EMG study Alfonso et al. 1982). Since this contraction is relatively slow, /**i**/-like tongue positions will tend to spread into nearby consonants and vowels, possibly leading to phonetic assimilations, and eventually, phonological umlaut rules.

Conclusions

This work has been concerned with the development of a realistic model of vowel articulation. I have attempted to quantitatify some of the abstractions of phonetic theory: articulatory primes, and their functional grouping into coordinative structures. A model based on these considerations has been developed and tested on a moderately broad range of data: x-ray tracings of some 250 vowels as produced by about 25 speakers of six languages. The model is successful at describing vowel articulation; more interestingly, it embodies articulatory parameters that are rather different from the traditional ones.

The articulatory parameters developed each specify the contribution of a particular kind of family of articulator shapes and positions to the vocal tract shapes observed in particular vowels. These parameters - articulatory primes - are potentially independent articulatory gestures that are part of the linguistic phonetic capabilities of human beings.

They form the basis for systematic phonetic descriptions of vowels.

In individual languages, some articulatory primes are functionally linked into coordinative structures. I have adopted the hypothesis that functional linking can be expressed as a linear relationship between the contributions of two primes to all the vowels of a language. The data analyzed here do indeed show signs of such functional relationships. A point worthy of note is that these functional relationships may differ from language to language. For instance, in Akan, the contribution of an articulatory prime that controls larynx and epiglottis position is approximately linearly related to the contribution of an articulatory prime that affects tongue root position. But in French, the same articulatory prime affecting larynx position is functionally linked with a different tongue position prime.

These functional linkings induce a certain amount of language-specific redundancy in the description of vowels. Thus, the coordinative structures that give rise to these functional linkings, and not articulatory primes themselves, form the elements of the most parsimonious phonetic description of the vowels of a particular language.

The relation of articulatory primes and coordinative structures to phonological description is clearly a topic needing further research, but I have argued that some phonological processes are better described using this articulatory parameterization. In this view, phonological vowel umlaut rules are categorical, discretized representations of phonetic assimilations. We may expect the phonological representation of vowels at some level to be a discretized version of the underlying phonetic continuua.

References

Alfonso, P. E., K. Honda, T. Baer, & K. S. Harris. 1982. "Multichannel study of tongue EMG during vowel production", J. Acoust. Soc. Am. **71** S54(A).

Anderson, S. R. 1972. "Icelandic u-umlaut and breaking in a generative grammar", in E. Firchow, K. Grimstad, N. Hasselmo, & W. O'Niel, eds. *Studies for Einar Haugen*. The Hague: Mouton.

___. 1974. *The Organization of Phonology*. New York: Academic Press.

Armstrong, L. E. 1967. *The Phonetics of French: A Practical Handbook*. London: G. Bell & Sons, Ltd.

Bao, H. & L Yang. 1986. *The Articulatory Movements of Standard Chinese*. Beijing: Beijing Language Institute Press.

Barrack, C. M. 1975. *A Diachronic Phonology from Proto-Germanic to Old English stressing West-Saxon Conditions*. The Hague: Mouton.

Bentler, P. M. 1985. *Theory and Implementation of EQS, a Structural Equations Program*. Los Angeles: BMDP Statistical Software.

___. 1986. *Lagrange Multiplier and Wald Tests for EQS and EQS/PC*. Los Angeles: BMDP Statistical Software.

Bentler, P. M., W.-Y. Poon, & S.-Y. Lee. 1987. "Generalized multimode latent variable models: Implementation by standard programs", in press, *Computational Statistics and Data Analysis*.

Bloomfield, L. 1939. "Menomini morphophonemics", *Travaux du Cercle Linguistique de Prague* **8**, 105-115. Reprinted in V. Makkai, ed. *Phonological Theory*, 58-64. New York: Holt.

Bothorel, A., P. Simon, F. Wioland, & J.-P. Zerling. 1986. *Cinéradiographie des Voyelles et Consonnes du Français (Travaux de l'Institut de Phonétique de Strasbourg)*. Strasbourg: Université des Sciences Humaines de Strasbourg.

Brichler-Labaeye, C. 1970. *Les Voyelles Françaises, Movements et Positions Articulatoires à la lumière de la Radiocinématographie*. Paris: Klincksieck.

Browman, C. P. and L. M. Goldstein. 1986) "Towards an articulatory phonology". *Phonology Yearbook* **3**, 219-252.

Carroll, J. D. 1965. "Generalization of canonical correlation analysis to three or more sets of variables". *Proceedings of the 76th Annual Convention of the APA*, 227-228.

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

Carroll, J. D. & S. Pruzansky. 1984. "The CANDECOMP-CANDELINC family of models and methods for multidimensional data analysis", in Law, Snyder, Hattie, & MacDonald, eds. (1984), 372-402.

Cassidy, F. G. & R. N. Ringler. 1971. eds. *Bright's Old English Grammar & Reader, Third Edition*. New York: Holt, Rinehart and Winston, Inc.

Cattell, R. B. 1944. "'Parallel Proportional Profiles' and other principles for determining the choice of factors by rotation". *Psychometrika* **9**, 267-283.

Cattel, R. B. 1978. *The Scientific use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.

Cattell, R. B. and A. K. S. Cattell. 1955. "Factor rotation for proportional profiles: Analytical solution and an example". *British Journal of Statistical Psychology* **8**, 83-92.

Chomsky, N. & M. K. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clements, G. N. 1985. "The geometry of phonological features", *Phonology Yearbook* **2**, 225-252.

Coker, C. H. 1967. "Synthesis by rule from articulatory parameters", *Proceedings of the 1967 Conference on Speech Communication and Processing*, Paper A9, 52-53; reprinted in Flanagan & Rabiner, eds. (1973), 396-399.

___. 1968. "Speech synthesis with a parametric articulatory model", *Speech Symposium, Kyoto*, Paper A-4; reprinted in Flanagan & Rabiner, eds. (1973), 135-139.

Coker, C. & O. Fujimura. 1966. "Development of a quantitative description of vowel articulation", *J. Acoust. Soc. Am.* **40**, 1271.

Coker, C. H., N. Umeda, & C. P. Browman. 1973. "Automatic synthesis from ordinary English text", in Flanagan & Rabiner, eds. (1973), 400-411.

Colton, M. A. 1909. *La Phonétique Castillane: Traité de Phonétique Descriptive et Comparative*. Paris.

Dart, S. 1987. "A bibliography of x-ray studies of speech". *UCLA Working Papers in Phonetics* **66** 1-97.

Disner, S. F. 1983. *Vowel Quality: the Relation between Universal and Language-Specific Factors*. UCLA Ph.D. dissertation; also *UCLA Working Papers in Phonetics* **58**.

Edwards, J. & K. S. Harris. 1985. "A two-dimensional model of jaw movement during speech", presentation SS15, 109th Meeting of the Acoustical Society of America, *J. Acoust. Soc. Am.* **77** S1, S100.

Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

___. 1965. "Formants and cavities", in E. Zwirner & W. Bethge, eds. *Proceedings of the 5th International Congress of Phonetic Sciences*, 120-141.

___. 1974. ed. *Speech Communication Seminar, Stockholm, 1974: Speech Production and Synthesis by Rules. Speech Communication* **2**. Stockholm: Almqvist & Wiksell.

Fischer-Jørgensen, E. 1985. "Some basic vowel features, their articulatory correlates, and their explanatory power in phonology", in Fromkin, ed. (1985), 79-99.

Flanagan, J. L. & Rabiner, L. R. 1973. eds. *Speech Synthesis*. New York: John Wiley & Sons, Inc.

Fromkin, V. 1985. ed. *Phonetic Linguistics*. Orlando, FL: Academic Press, Inc.

Fujimura, O. & Y. Kakita. 1978. "Remarks on quantitative description of the lingual articulation", in Lindblom & Ohman, eds. (1978).

Gorsuch, R. L. 1983. *Factor Analysis (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Halle, M. K. 1983. "On distinctive features and their articulatory implementation". *Natural Language and Linguistic Theory* **1**, 91-106.

Halle, M. K. & V. J. Zeps. 1966. "A survey of Latvian morphophonemics", *MIT Quarterly Progress Reports* **83**, 105-113.

Harshman, R. 1984. "How can I know if it's 'real?' A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling", Appendix A in Law, Snyder, Hattie, & MacDonald, eds. (1984), 566-591.

Harshman, R., & W. De Sarbo. 1984. "An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques", Appendix C in Law, Snyder, Hattie, & MacDonald, eds.

(1984), 602-642.

Harshman, R., P. N. Ladefoged & L. Goldstein. 1977. "Factor analysis of tongue shapes", *J. Acoust. Soc. Am.* **62**, 693-707

Harshman, R. & M. E. Lundy. 1984a. "The PARAFAC model for three-way factor analysis and multidimensional scaling", in Law, Snyder, Hattie, & MacDonald, eds. (1984), 122-215.

___. 1984b. "Data preprocessing and the extended PARAFAC model", in Law, Snyder, Hattie, & MacDonald, eds. (1984), 216-284.

Hashimoto, K. & K. Sasaki. 1982. "On the relationship between the shape and position of the tongue for vowels", *Journal of Phonetics* **10**, 291-299.

Hashimoto, K. & S. Suga. 1986. "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue", *Journal of the Acoustical Society of Japan* (English) **7** 1, 39-46.

Haugen, E. 1935. *Beginning Norwegian*. New York: Appleton Century Crofts.

Henderson, E. J. A. 1971. *The Indispensible Foundation: A Selection from the Writings of Henry Sweet*. London: Oxford University Press.

Henke, W. L. 1966. *Dynamic Articulatory Model of Speech Production Using Computer Simulation.* Unpublished MIT Ph. D. Dissertation.

Hiki, S. & J. Oizumi. 1974. "Speech synthesis by rule from neurophysiological parameter". in G. Fant, ed. (1974) 219-225.

Hiki, S. & K. Niyata. 1973. "Articulatory model for vowel production", in *Speech Processing*, University of Tokyo Press.

Hockett, C. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Holbrook, R. T. & F. J. Carmody. 1937. *X-Ray Studies of Speech Articulations* (*University of California Press Publication in Modern Philology* **20**), 187-237.

International Phonetic Association. 1949. *The Principles of the International Phonetic Association*. Department of Phonetics, University College, London.

Ishizaki, S. 1980a. "Observation and normalization of vowel coarticulation", *Progress Report on Speech Research '79*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

___. 1980b. "On a dynamic processing method for coarticulation", *Progress Report on Speech Research '79*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

___. 1981a. "Dynamic processing effect in vowel coarticulation", *Progress Report on Speech Research '80*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

___. 1981b. "Consonant feature extraction using articulatory parameters obtained from area functions", *Progress Report on Speech Research '80*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

Ishizaki, S., K. Fuchi, & T. Nakajima. 1977. "Estimation of articulation points by use of the Kalman smoothing filter", *Progress Report on Speech Research '76*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

Ishizaki, S. & T. Nakajima. 1977. "Articulatory model separating the tongue front from the tongue body", *Progress Report on Speech Research '76*. Tokyo: Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

Jackson, M. T. T. 1986. "Vowels and diphthongs in Icelandic", presentation EE5, 112th Meeting of the Acoustical Society of America, *J. Acoust. Soc. Am.* **80** S1, S62.

___. 1988. "Analysis of tongue positions: language-specific and cross-linguistic models", accepted for publication in *J. Acoust. Soc. Am.* **84**.

Jacobsen, L. 1978. *DhoLuo Vowel Harmony: a Phonetic Investigation*. UCLA Ph. D. dissertation; also *UCLA Working Papers in Phonetics* **43**.

Jennrich, R. 1972. "A generalization of the multidimensional scaling model of Carrol and Chang", *UCLA Working Papers in Phonetics* **22**, 45-47.

Jones, D. 1950. *The Pronunciation of English, Third Edition*. Cambridge: Cambridge University Press.

___. 1956. *An Outline of English Phonetics, Eighth Edition*. Cambridge: W. Heffer & Sons Ltd.

___. 1963. *The Pronunciation of English, Fourth Edition*. Cambridge: Cambridge University Press.

Jospa, P. 1974. "On the use of critical formant frequencies in designing an articulatory model", in G. Fant, ed. (1974) 285-292.

Josselyn, F.-M. 1907. *Etudes de Phonétique Espagnole*. Paris: Welter.

Kakita, Y. & O. Fujimura. 1977. "Computational model of the tongue: a revised version", in *J. Acoust. Soc. Am.* **62** S1, S15.

Kakita, Y., O. Fujimura, & K. Honda. 1985. "Computation of Mapping From Muscular Contraction Patterns to Formant Patterns in Vowel Space", in Fromkin, ed. (1985), 133-144.

Keating, P. 1984. "Phonetic and Phonological Representation of Stop Consonant Voicing". *Language* **60** 2, 286-319.

Kelso, J. A. Scott, B. Tuller, & K. S. Harris. 1983. "A 'Dynamic Pattern' Perspective on the Control and Coordination of Movement", in MacNeilage, ed. (1983), 137-174.

Kelso, J. A. Scott, E. L. Saltzman, & B. Tuller. 1986. "The dynamical perspective on speech production: data and theory". *J. Phonet.* **14**, 29-60.

Kent, R. D. 1970. *A cinefluorographic-spectrographic investigation of the component gestures in lingual articulation.* Unpublished Ph.D. dissertation, University of Iowa.

___. 1972. "Tongue body articulation during vocal and diphthong gestures", *Folia Phoniatrica* **24**, 286-300.

Kenyon, J. S. 1956. *American Pronunciation, Tenth Edition*. Ann Arbor: George Wahr Publishing Co.

Kiritani, S. 1986. "X-ray microbeam method for the measurement of articulatory dynamics - techniques and results". *Speech Communication* **5** 2, 119-140.

Kiritani, S., K. Miyawaki, & O. Fujimura. 1976. "A computational model of the tongue", *Ann. Bull., Res. Inst. Logop. Phoniatr., Univ. Tokyo* **10**, 243-252.

Kreysig, E. 1978. *Introductory functional analysis with applications*. New York: John Wiley & Sons.

Koolhoven, H. 1968. *Dutch*. London: The English Universities Press.

Kroonenberg, P. M. 1981. *User's guide to TUCKALS2. A program for three-mode principal component analysis with extended core matrix (WEP-Reeks, WR 81-35-RP)*. Leiden, The Netherlands: University of Leiden.

Kroonenberg. P. M. & J. de Leeuw. 1980. "Principal components analysis of three-mode data by means of alternating least-squares algorithms", *Psychometrika* **5**, 69-97.

Kruskal, J. B. 1984. "Multilinear methods", in Law, Snyder, Hattie, & MacDonald, eds. (1984), 36-62.

Ladefoged, P. 1976. "The phonetic specification of the languages of the world", *UCLA Working Papers in Phonetics* **31** 3-21.

___. 1982. *A Course in Phonetics, 2nd ed.* (First ed. 1975) New York: Harcourt Brace Jovanovich.

Ladefoged, P. N., & M. K. Halle. 1988. to appear in *Language*.

Law, H. G., C. W. Snyder Jr., J. A. Hattie, & R. P. MacDonald. 1984. eds., *Research Methods for Multimode Data Analysis.* New York: Praeger.

Liljencrants, J. L. 1971. "A Fourier series description of the tongue profile", *Speech Transmission Lab (Stockholm) Quarterly Progress and Status Report* **4**, 9-18.

Liljencrants, J. L. 1985. *Speech Synthesis with a Reflection-Type Line Analog.* Ph. D. Dissertation. Royal Institute of Technology, Stockholm: Department of Speech Communication and Music Acoustics.

Lindau, M. 1975. "Features for vowels", *UCLA Working Papers in Phonetics* **30**.

___. 1978. "Vowel features", *Language* **54** 3, 541-563.

___. 1979. "The feature expanded", *Journal of Phonetics* **7**, 163-176.

___. 1986. "Vowel features in Akan and English", presentation EE6, 112th Meeting of the Acoustical Society of America, *J. Acoust. Soc. Am.* **80** S1, S62.

Lindblom, B. E. 1968. *On the Production and Recognition of Vowels.* Unpublished Ph. D. dissertation, Lund University.

Lindblom, B. E. & J. E. F. Sundberg. 1971. "Acoustical consequences of lip, tongue, jaw, and larynx movement", *J. Acoust. Soc. Am.* **50**, 1166-1179.

Lindblom B. E., S. Pauli & J. E. F. Sundberg. 1974. "Modeling coarticulation in apical stops", in G. Fant, ed. (1974). 87-94.

Lindblom, B. E. & S. Ohman, eds. 1978. *Frontiers of Speech Communication Research.* London: Academic Press.

Linker, W. 1982. *Articulatory and Acoustic Correlates of Labial Activity in Vowels: A Cross-linguistic Study.* UCLA Ph. D. dissertation; also *UCLA Working Papers in Phonetics* **56**.

Lundy, M. E. & R. A. Harshman. 1985. *Reference Manual for the PARAFAC Analysis Package*. London, Ontario: Scientific Software Associates.

McCarthy, J. J. 1984. "Theoretical consequences of Montañes vowel harmony", *Linguistic Inquiry* **15**, 291-318.

McClean, R. J. 1969. *Swedish, A Grammar of the Modern Language*. London: The English Universities Press.

MacNeilage, P. F. 1983. ed. *The Production of Speech.* New York: Springer-Verlag.

Maddieson, I. & P. Ladefoged. 1985. "'Tense' and 'Lax' in four minority languages of China". *Journal of Phonetics* **13** 4, 433.

Maeda, S. 1978. "Une analyse statistique sur les positions de la langue: Etude préliminaire sur les voyelles françaises", *Actes des 10è Journées d'Etudes sur la Parole*, 191-200. Lannion: Groupement d'Acousticiens de la Langue Française.

___. 1979. "Un modèle articulatoire basé sur une étude acoustique", *Bulletin de l'Institut de Phonétique de Grenoble* **8**, 35-55.

Malmberg, B. 1963. *Spansk Fonetik.* Lund: Gleerups.

Mermelstein, P. 1973. "Articulatory model for the study of speech production", *J. Acoust. Soc. Am.* **53**, 1070-1082.

Miller, G. A. and P. E. Nicely. 1955. "An analysis of perceptual confusion among some English consonants". *J. Acoust. Soc. Am.* **27**, 338-352.

Miyawaki, K., H. Hirose, T. Ushijima, & M. Sawashima. 1975. "A preliminary report on the electromyographic study of the activity of lingual muscles", *Ann. Bull., Res. Inst. Logop. Phoniatr., Univ. Tokyo* **9** 91-103.

Moll, K. L. 1962. "Velopharyngeal closure on vowels". *Journal of Speech and Hearing Research* **5** 1, 30-37.

Mosteller, F. and J. W. Tukey. 1977. *Data Analysis and Regression: A Second Course in Statistics*. Reading, MA: Addison-Wesley.

Nakajima, T. 1978. "Identification of a dynamic articulatory model by acoustic analysis", in Sawashima & Cooper, eds. (1978), 251-275.

Nash, D. 1979. "Warlpiri vowel assimilation", in K. Safir, ed. "Papers on Syllable Structure, Metrical Structure, and Harmony Processes", *MIT Working Papers in Linguistics* **1**.

Navarro-Tomás, T. 1916. "Siete Vocales Españolas". *Revista de Filología Española* **3**, 51-62.

___. 1968. *Manual de Pronunciación Española (6th ed.)*. New York: Hafner.

Nearey, T. M. 1978. *Phonetic Feature Systems for Vowels*. Bloomington, IN: Indiana University Linguistics Club.

Nielsen, N. & P. L. Hjorth. 1971. *Dansk sproglaere for svenskere*. Lund: Gleerups Forlag.

Norlin, K. 1987. *A Phonetic Study of Emphasis and Vowels in Egyptian Arabic*. Lund University Ph. D. dissertation, also *Working Papers, Lund University Department of Linguistics and Phonetics* **30**.

Ohnesorg, K. & O. Svarný. 1955. "Etudes expérimentales des articulations chinoises". Ceskoslovoské Akademie Ved: *Rozpravy* **65** 5. Prague: Czech Akademy.

Ohman, S. E. G. 1966. "Coarticulation in CVC utterances: spectrographic measurements", *J. Acoust. Soc. Am.* **39**, 151-168.

Ondrácková, J. 1973. *The Physiological Activity of the Speech Organs: An analysis of the Speech Organs During the Phonation of Sung, Spoken, and Whispered Czech Vowels on the Basis of X-Ray Method*. The Hague: Mouton.

Overall, J. E. 1962. "Orthogonal factors and uncorrelated factor scores", *Psychological Report* **10**, 651-662.

Parmenter, C. E. & S. Treviño. 1932. "An X-ray study of Spanish vowels". *Hispania* **15**, 483-496.

Perkell, J. 1969. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study (MIT Press Research Monograph #53)*. Cambridge, MA: MIT Press.

___. 1974. *A Physiologically Oriented Model of Tongue Activity During Speech Production.* Unpublished MIT Ph.D. Dissertation.

Perkell, J. & D. Klatt. 1986. eds. *Symposium on Invariance and Variability of Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Peterson, G. & H. L. Barney. 1952. "Control methods used in a study of vowels", *J. Acoust. Soc. Am.* **24** 175-184.

Pétursson, M. 1974a. "Peut-on interpréter les données de la radiocinématographie en fonction du tube acoustique à section uniforme?" *Phonetica* **29**, 22-79.

___. 1974b. *Les Articulations de l'Islandais à la lumière de la Radiocinématographie.* Paris: Librarie C. Klincksieck.

Popp, N. 1965. *Introduction to Altaic Linguistics.* Wiesbaden: Otto Harrassowitz.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. 1986. *Numerical Recipes: The Art of Scientific Computing.* New York: Cambridge University Press.

Quilis, A. 1981. *Fonética Acústica de la Lengua Española.* Madrid: Biblioteca Románica Hispánica.

Ramsay, J. O. 1982. "When the data are functions". *Psychometrika* **47**, 379-396.

Riordan, C. 1977. "Control of vocal tract length in speech". *J. Acoust. Soc. Am.* **62** 4, 998-1002.

Rossi, M. 1983. "Niveaux de l'analyse phonétique: nature et structuration des indices et des traits", *Speech Communication* **2**, 91-106.

Rubin, P. E., T. Baer, & P. Mermelstein. 1981. "An articulatory synthesizer for perceptual research", *J. Acoust. Soc. Am.* **70**, 321-328.

Russell, G. O. 1928. *The vowel.* Columbus: Ohio State University Press.

___. 1929-30. "The mechanism of speech", *J. Acoust. Soc. Am.* **1**, 83-109.

Saltzman, E. and J. A. S. Kelso. 1987. "Skilled actions: A task-dynamic approach", *Psychological Review* **94**, 84-106.

Sapir, E. 1922. "The Takelma language of Southern Oregon", in F. Boas, ed. *Handbook of American Indian Languages, Bulletin of the Bureau of American Ethnology* **40**, II. Washington DC: BAE.

Sawashima, M. & F. S. Cooper. 1978. eds. *Dynamic Aspects of Speech Production.* Tokyo: University of Tokyo Press.

Schane, S. A. 1984. "The fundamentals of particle phonology", *Phonology Yearbook* **1**, 129-155.

Sekimoto, S., H. Imagawa, & S. Kiritani. 1978. "Dynamic characteristics of tongue movement in the production of connected vowels". *Ann. Bull., Res. Inst. Logop. Phoniatr., Univ. Tokyo* **12**, 11-20.

Shirai, K. & M. Honda. 1978. "Estimation of articulatory motion", in Sawashima & Cooper, eds. (1978), 279-302.

Shirai, K. & T. Kobayashi. 1986. "Estimating articulatory motion from speech wave". *Speech Communication* **5** 2, 159-170.

Sonoda, Y. & K. Nakakido. 1986. "Effect of speaking rate on jaw movements in vowel sequences", *Journal of the Acoustical Society of Japan* (English) **7** 1, 5-12.

Steriade, D. to appear. "Vowel tiers and geminate blockage", *Natural Language and Linguistic Theory*.

Stevens, K. N. & A. S. House. 1955. "Development of a quantitative model of vowel articulation", *J. Acoust. Soc. Am.* **27**, 484-493.

Stevens, K. N., S. J. Keyser, & H. Kawasaki. 1986. "Toward a phonetic and phonological theory of redundant features". In Perkell & Klatt, eds. (1986).

Stevens, K. N. & J. S. Perkell. 1977. "Speech physiology and phonetic features", in Sawashima & Cooper, eds. (1978), 323-341.

Strang, G. 1976. *Linear Algebra and its Applications*. New York: Academic Press.

Sundberg, J. 1969. "Articulatory differences between spoken and sung vowels in singers", *Stockholm Royal Institute of Technology - Speech Transmission Laboratory/Quarterly Progress and Status Reports* **1**, 33-46.

ten Cate, A. P., P. Jordens, & W. van Lessen Kloeke. 1976. *Deutsche Phonetik: Laut- und Aussprachelehre für Niederländer*. Groningen: Wolters-Noordhoff.

Terbeek, D. 1977. *A cross-language multidimensional scaling study of vowel perception*. UCLA Ph. D. dissertation, also *UCLA Working Papers in Phonetics* **37**.

Thongkum, T. 1987. "Phonation types in Mon-Khmer languages", *UCLA Working Papers in Phonetics* **67** 29-48.

Topping, D. 1968. "Chamorro vowel harmony", *Oceanic Linguistics* **7**, 67-79.

___. 1973. *Chamorro Reference Grammer*. Honolulu: University of Hawaii Press.

Tucker, L. R. 1966. "Some mathematical notes on three-mode factor analysis", *Psychometrika* **31**, 279-311.

Wells, J. C. 1975. "The Association's alphabet", *Journal of the International Phonetic Association* **5** 52-58.

Wheaton, B. 1987. "Assessment of Fit in Overidentified Models with Latent Variables", *Sociological Methods & Research* **16**, 118-154.

Wood, S. 1971. "A spectrographic study of allophonic variation and vowel reduction in West Greenlandic Eskimo", *Lund University Phonetics Laboratory Working Papers* **4**, 58-94.

___. 1982. *X-Ray and Model Studies of Vowel Articulation.* Lund University Ph. D. dissertation, also *Working Papers, Lund University Department of Linguistics and Phonetics* **23**.

Zerling, J.-P. 1979. *Articulation et Coarticulation dans les Groupes Occlusive-Voyelle en Français*. Thèse de Troisième Cycle, Université de Nancy, France.

Zhou, D.-F. & Z.-J. Wu. 1963. eds., *Articulation Album of Putonghua*. Beijing: Commercial Press.