

UCLA

Working Papers in Phonetics

Title

WPP, No. 37: A Cross-language Multidimensional Scaling Study of Vowel Perception

Permalink

<https://escholarship.org/uc/item/3nx4g138>

Author

Terbeek, Dale

Publication Date

1977-08-01

a cross-language
multidimensional scaling
study of
vowel perception

Dale Terbeek

A cross-language multidimensional scaling
study of vowel perception

Dale Terbeek

Working Papers in Phonetics 37

August 1977

University of California, Los Angeles

for Kathy

errata

page 2-1, paragraph 2, line 6: 'Singh (1974) has'
should read
'Wiener and Singh (1974) have'

page 2-2, paragraph 2: insert the following between
lines 7 and 8:

'might organize vowels. This would
overcome objection 1. Still, there
might be considerable personal vari-
ation in the way listeners'

page 4-1, paragraph 1: replace lines 9 and 10 with the
following:

'valid. By reliability is meant re-
peatability under identical condi-
tions, and by validity is meant con-
formance with a priori expectations
of the data.'

Table of Contents

Acknowledgments	
Abstract	
1. Introduction	1
1.1 The abstract nature of the mental representation of speech sounds	1
1.2 Some evidence for the reality of perceptual features	4
1.3 The notion of perceptual distance	9
2. Methods and results: a review	11
2.1 The collection of perceptual data	11
2.2 The use of 2-mode MD-scaling	15
2.3 Four 2-mode studies of vowel perception in light of the problem of rotational indeterminacy	17
2.4 Three-mode MD-scaling (PARAFAC and INDSCAL) as a tool for finding "true" perceptual dimensions and evaluating hypotheses about linguistic uni- versals	38
2.5 Testing PARAFAC on real and synthetic data	41
3. The experimental design	58
3.1 Languages used in the experiment	58
3.2 Selection of the stimulus vowels	63
3.3 Design of the stimulus tokens	67
3.4 Preparation of the stimuli	69
3.5 Selection of listeners	78
3.6 The experimental task	79
3.7 Notes on the listeners' behavior	83
4. Preliminary survey of the collected data	85
4.1 Reliability of the data	85
4.11 Observations of discrepancies for repeated triads	86
4.12 Direction of deviation from chance values	90
4.13 The effect of rate of judgment on relia- bility	91

4.2	Validity of the data	91
4.21	The distribution of values within each individual vowel matrix	92
4.22	Inspection of the content of the response matrices	96
4.3	Hierarchical clustering of vowel distances, grouped by language	97
4.31	Labeling clusters according to distinctive features	99
4.32	Relative length of branches and inter-language differences	106
4.33	Branch length as indication of tightness of clustering	107
4.34	Phonological properties	109
4.35	Limitations of hierarchical clustering analysis	112
4.4	Preliminary individual scaling to obtain additive constants	112
5.	The 6-dimension PARAFAC solution for all 35 listeners	120
5.1	Distinctive feature interpretation of the 6-dimensional solution	120
5.11	Backness	124
5.12	Height	124
5.13	Rounding	126
5.14	Overview of the space of 3 curved dimensions	126
5.2	Linear regression versus special strategies .	128
5.21	Backness	130
5.22	Height	137
5.23	Rounding	139
5.3	In defense of six dimensions	145
5.31	Formal evidence for six dimensions . .	145
5.32	The interpretation of alternate solutions	147
5.4	Personal variation within the 6-D PARAFAC solution	158

6.	PARAFAC and PARAFAC solutions for individual languages	164
6.1	German	165
6.2	Turkish	173
6.3	Thai	184
6.4	English	196
6.5	Swedish	
6.51	Problems in obtaining the Swedish solution	206
6.52	Acoustic sources for Swedish perceptual dimensions	214
6.6	Discussion	219
7.	Conclusion	226
7.1	Discreteness and feature interdependence . .	226
7.2	Binary vs. multivalued features	
7.3	Phonemic contrast and perceived distances . .	231
7.4	Summary	240
Appendix I.	Part A. Raw dissimilarity matrices for each listener	242
	Part B. Raw dissimilarity matrices summed for each language	248
Appendix II.	Co-ordinates for selected solutions . .	251
Appendix III.	A note on the I/O ratio as an index of solution stability	264
Bibliography		267

ACKNOWLEDGEMENTS

There probably exist people who do research -- perhaps even good research -- all by themselves. But this dissertation did not happen this way. Many there are who contributed their time, ideas, effort and support, all to my benefit.

Peter Ladefoged introduced me to the idea of phonetic research in general, to vowel perception in particular, and to the insidious delights of the laboratory computer. I feel very privileged to have had his guidance throughout the planning and execution of the experiment. I am especially grateful for his detailed, critical comments on the first draft.

Vicki Fromkin, whose frustrating office of thesis advisor is hereby brought to a merciful close, provided constant encouragement. Would that every graduate student had such a friend. I thank her also for making me keep sight of theoretical issues in phonology, both in the experiment itself and in the writing.

George Allen did his best to teach me the importance of careful, thorough experimental design and analysis. This work would have been a poorer one without the benefit of his viewpoint.

The phonetics lab was the scene of many spontaneous, stimulating discussions among students; I thank Mona Lindau, Steve Krashen, George Papcun, and Lloyd Rice for sharing their ideas and suggestions with me. Lloyd was, in addition, a most patient coach and troubleshooter who saw me through many a frustration in programming and operating the PDP-12.

Of all my student colleagues, Richard Harshman had by far the greatest impact on me, both as a friend, and as a scientist. The concept of this dissertation could not have existed without PARAFAC, and its execution was enhanced by many profitable hours with Rich, with my taking away insights into MD-scaling, factor analysis, and the creative use of mathematics and statistics in studying human behavior. I regard my all too brief collaboration with Rich as the most rewarding aspect of my stay at UCLA.

Lee Cooper was very generous in helping me use COSCAL effectively. Willie Martin and Larry Grant were always on hand whenever some lab equipment needed fixing or building. Edie Manahan and Jeff Strauss, respectively, typed and proof-read -- willingly, it seems, many tedious tables of numbers. I thank collectively my experimental subjects, who set new records for the endurance of boredom. Gerry Greene's superb evocation of the essence of laboratory research graces the cover.

I am grateful to the Woodrow Wilson Foundation, the UCLA campus computing network, and especially Marie Terbeck for indispensable financial support.

Transforming the manuscript, which ranged from nearly illegible to genuinely bewildering, into a presentable draft was the work of Kathy, who remained my wife. Typing the final version from the revised first draft, a good 50% longer and equally indecipherable, was once again done by Kathy, who is still my wife. There must be some way for me to show my gratitude for her faith in me and her constant support. I'll try to think of something.

Dale Terbeck
Chicago
July 1977

Abstract

The present work is a study of the perception of vowel quality which **focuses** on three main issues. (1) What are the perceptual attributes (dimensions, features) according to which listeners perceive vowels? (2) What in the acoustic stream can be said to underlie those attributes? (3) To what extent are the perceptual dimensions and their acoustic correlates language-specific or language-universal?

The data consist of triadic comparisons performed on 12 monophthongs in the context [b**ə**b_], by speakers (35 in all) of five different native languages: English, German, Thai, Turkish, and Swedish. The resulting 35 dissimilarity matrices were analyzed by PARAFAC, a specialized 3-mode factor analysis program. PARAFAC solutions include not only the dimensions of the vowel space, but coefficients representing the relative weight attached to those dimensions by each listener. PARAFAC solutions cannot be rotated without degrading the predictions of the listeners' data. PARAFAC dimensions must therefore be interpreted without rotation and may thus be regarded as psychologically real dimensions. This powerful analysis tool, therefore, is the source of claims made about "true" perceptual categories (question 1).

Question 2 is investigated by means of multiple regression analyses, treating each perceptual dimension as a dependent variable, and the acoustic measurements made on the stimuli as independent variables. These analyses are supplemented by discussion of possible perceptual strategies that account for the patterns on the perceptual dimensions in instances where the linear regression model gives unsatisfactory results.

Question 3 is treated by first analyzing all listeners taken together, then each language group separately. The patterns of personal weighting coefficients in the former analysis, and the differences in the dimensions themselves that arise with the latter condition, provide evidence relevant to language-universal vs. language-specific hypotheses.

The three main results are the following: (1) Perceptual dimensions correspond strikingly well with categories in use by linguists for the description of vowels, although the vowels often appear in ternary or continuous patterns on those dimensions. Binary oppositions appeared occasionally, but not always. (2) Dimensions which vary in detail from one language to another, but which would be described by the same general label, may be based on different acoustic cues. (3) The perceptual differences found across languages may depend upon abstract phonological properties, in addition to differences in phonetic inventory.

Chapter 1

Introduction

1.0 Two themes recur frequently in modern linguistic theory. One concerns the representation of language in the human mind, or the "psychological reality" of linguistic patterns. The second is the universal as opposed to language-specific aspects of these patterns. The present research is an experimental project dealing with both those themes as revealed in the mental representation of vowel quality.

1.1 The abstract nature of mental representation of speech sounds

The basic assumption is that mental representations for speech sounds must be abstractions related to both the articulatory activity that produces sounds and the sound waves that result. It is rejected at the outset that the mental representations might be equivalent to any specific stored motor commands or stored acoustic patterns.

Any claim that a certain pattern of motor gestures constitutes the mental representation of vowels can only be seen as a straw man: the same vowel in different consonant contexts will be approached by means of different gestures. Furthermore, it has been known for some time that alternate gestures can produce very similar vocal tract shapes. Ladefoged (1971, p. 70) credits Sweet (1890) with the observation that a given tongue height can be produced with an open jaw and longitudinally contracted tongue, or closed jaw and flattened tongue. Cine-x-rays of several speakers producing monosyllables (Ladefoged, et al, 1972) show that such variation is idiosyncratic. The essence of vowel quality cannot be captured by an enumeration of possible gestures for each vowel.

A more plausible claim is that vowels are represented mentally in terms of vocal tract shape -- a record of oral cavity cross-sectional area variation from lips to glottis. But this suggestion is seriously damaged by the fact that alternate shapes can give rise to extremely similar acoustic outputs. Some striking, albeit indirect evidence is found in Ladefoged (1967, Chapter 2). In that study, eighteen trained phoneticians were asked to mark cardinal vowel charts to correspond to a set of tape-recorded vowels. For seven vowels with "unmarked" rounding (i.e., front unrounded or back rounded), the worst case of disagreement as to placement on the chart involved less than 4% of the chart area. But

for three other vowels, there was considerable disagreement in the assessment of lip gestures and in the placement of the vowels on the chart as well. Half of the subjects thought that one vowel (labelled "H", transcribable perhaps as [ɔ̃], midback unrounded) had lip rounding; half thought it had spread or neutral lips (Figure 47, p. 135). In addition, the mistaken assessment by eleven of the subjects of vowel "J" as rounded was accompanied by a spread in judgments that covered 27% of the vowel chart area.

In theory, an experienced listener should be able to judge the lip activity associated with a vowel -- the third dimension, quantized in Ladefoged's experiment into four degrees of constriction -- and then mark the remaining plane according to the auditory quality due to the tongue position for the vowel. (Although actual tongue position is not linearly related to position on a vowel chart, it is fair to say that the two are largely isomorphic.) But expert listeners, as Ladefoged has shown, cannot necessarily separate the auditory effects of rounding and tongue position. Where one listener may hear a sound as rounded and relatively front, another will judge it to be unrounded but with a more retracted tongue position. This directly demonstrates that a given vowel may be perceived in different ways. The converse, that different articulations could give rise to very similar sounds, is a highly plausible inference.

Acoustic similarity in the face of articulatory differences has found its way into linguistic descriptions, notably in Jakobson's feature flat, which reflects lip rounding as well as constriction at the other end of the oral cavity (i.e., pharyngeal constriction) (Jakobson, Fant, Halle, 1952). The same is true of the feature grave, which is traditionally used to distinguish [t̪], with a high frequency noise burst, from [p] and [k], which have noise bursts of much lower frequency. More recently, Stevens (1971) has defended the view that phonemes are characterized by regions of acoustic stability which may be produced by articulations which vary over a large range. This viewpoint is compatible with the position adopted here: the rejection of articulation as the main source of the mental representations of sounds.

What is the result of considering various acoustic characterizations as potential psychological representations for vowels? The most naive possibility is that vowels are identified by referring to actual speech waves stored in the brain. But a speech wave includes information relevant to fundamental frequency, voice quality, rate of speaking, and speaker identification (c.f. Abercrombie 1967, Ladefoged 1967, Chapter 2). Vowel quality information must be extracted from a speech wave and separated from the other aspects just

mentioned. Besides, even if one speaker were to produce a given vowel repeatedly under the same conditions, the vowel tokens would differ in fine detail.

The traditional acoustic characterization of vowels is in terms of the first two or three formants: the two or three lowest vocal tract resonant frequencies associated with the production of each vowel. (See, for example: Fant, 1968, 1970; Flanagan, 1972; Ladefoged, 1967, 1972, 1962; Joos, 1948.) But absolute formant frequencies cannot serve as the mental representations of vowels either. Formant frequencies depend upon overall vocal tract length as well as the shape of the articulators. This was first discussed by Joos (1948), who considered plots of formants 1 and 2 for different speakers and found them different in actual frequency values, but similar in pattern. Peterson and Barney's (1952) measurements of formant frequencies of English vowels demonstrated a marked difference between male and female formants due to vocal tract length differences.

Ladefoged (1967) reports an experiment carried out earlier Ladefoged and Broadbent, (1957), in which the synthesized phrase "Please say what this word is" is followed by a synthesized CVC for listeners to identify. Changing the "personal quality" of the introductory sentence, by raising or lowering the absolute frequency of the formants contained in it, caused the identification of the stimulus word to change. For example, raising the first formant of the sentence caused [bɪt] responses where the original response to the same stimulus was [bɛt] (p. 110, Figure 44).

Clearly, the absolute value of formant frequencies is a fragile criterion to use in fixing vowel quality. Apparently, adjustment to a current speaker's vowel formants is rapid and automatic. The psychological representation of vowel sounds clearly must be more abstract than records of motor patterns, oral tract shapes, speech waves, or formant frequencies.

Within the framework of this study, the concept of the psychological representation of vowels will be limited to the area of perception and exclude any direct references to vowel production. But it is not unreasonable to consider the perceptual aspects of vowel quality as prerequisite to the understanding of production aspects. Ladefoged (1972) argues that vowel targets might be acoustic ones, rather than positional ones referring to the vocal tract. Various gestures might be invoked to approximate an acoustically based mental image of a vowel.

We are therefore interested in the representation of sounds which appears in Box C on Figure 1.1, which traces the relevant steps in the perceptual operation.

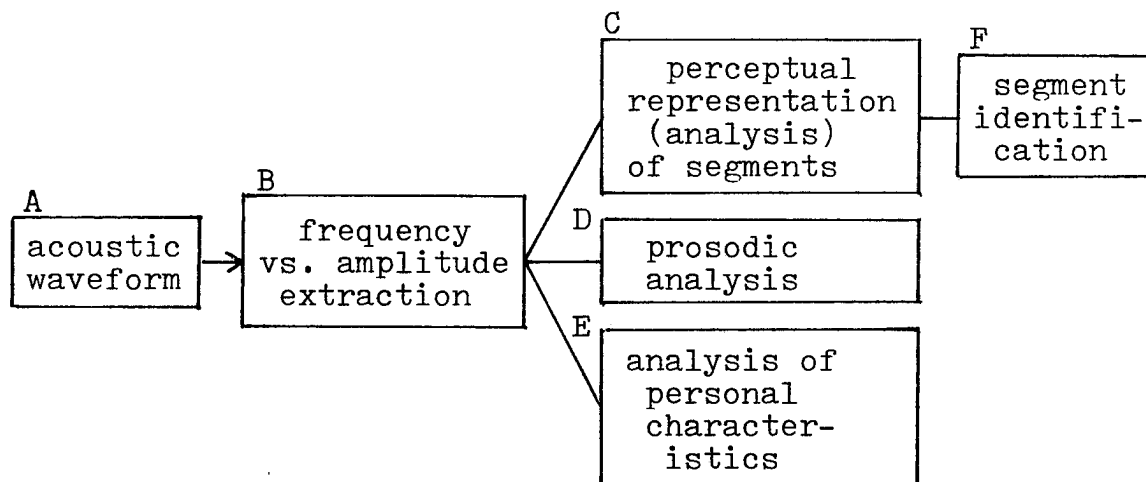


Figure 1.1: One component of a model for speech perception.

Sound waves (A) are transmitted by way of the eardrum and the bones of the inner ear through the oval window into the cochlear fluid. The function B, frequency vs. amplitude extraction, is performed at the interface of the mechanical elements in the cochlea (tectorial membrane and hair cells) and the neurons which join together to form the auditory nerve. (More details about the behavior of the ear in transmitting sound to the brain can be found, for example, in Ladefoged 1962, Gulick 1971, Kaplan 1972, Flanagan 1972.) E and D represent functions that parallel the function of extracting phonetic segments. Box C represents the perceptual analysis that eventually results in the decision made in Box F: the identification of the segment in question.

This sketch is obviously not a complete model of the perception of speech. It is only a portion of a model which must extend to morphological, syntactic and semantic data and also must include confirmation and rejection of predictions based on previously perceived material. Nevertheless, a model for perception must incorporate the chain of events described in Figure 1.1.

1.2 Some evidence for the reality of perceptual features

This research was based on the premise that vowel perception requires several channels, or criteria, or features, or properties. Each incoming vowel is assessed

according to these properties, and the resulting collection of values with respect to the properties identifies the vowel.

In linguistics, phonetic segments are grouped into classes by intersecting properties, known as distinctive features. One such set of features for vowel quality (Ladefoged, 1971) includes Height, Backness, Rounding and Tension. The vowel sound in the word "fee" would be classed as High, Front, Nonround, and Tense. Two other well-known sets of features are those of Chomsky and Halle (1968), and Jakobson, Fant and Halle (1952). (These two systems are conveniently referred to as CH and JFH.) Although there are differences between these two systems, they are effectively identical in describing vowels. High, Low, Back, Round and Tense in the CH system are parallel to Diffuse, Compact, Grave, Flat and Tense, respectively, for JFH. The feature Low or Compact is necessary to categorize three degrees of vowel height, since the features are taken to be binary. A high vowel such as [i] is [+High -Low], a mid vowel [e] is [-High -Low], and a low vowel [æ] is [-High +Low]. (Ladefoged advocates multivalued features, such that [i e æ] are regarded as 2, 1, 0 Height, respectively.) Other authors, such as Abercrombie (1967), Hockett (1962), Pike (1945), who do not explicitly work with distinctive features nevertheless use labels for vowel attributes that serve much the same purpose.

With the exception of the JFH features, the labels for vowel quality are based on articulatory parameters. Height and Backness were once thought to refer to the position of the highest point on the tongue; Round refers to the activity of the lips. Ladefoged (1967, p. 73) credits Essner (1947) with the first published account of the close relationship between articulatory labels and the formant structure of vowels. The JFH feature system, and Fant's own revised feature system (Fant, 1970) use labels that refer more directly to acoustic properties.

What unifies all the above schemes for describing vowels is the fact that they all rely on intersecting properties that group vowels into classes. These feature systems suggest various classifications that could operate in the perception of vowels. If "Height" and "Backness," for example, were assumed to be perceptual properties, the output of the cochlea would be subjected to two analyses, one which assigns values according to the feature called Height, and one according to the feature Backness. The vowels [i] and [u] would both be classified as High (or "+High" or "2High" or "+Diffuse" or "+Close," depending

on whose system is preferred) but assigned different values for Backness. [ɪ] is front (-Back, etc.) and [u] is back (+Back, etc.). These two classifications together with whatever other features are necessary would form the basis for the ultimate decision to regard the sound as [ɪ] or [u].

Prior to discussing the approach for discovering and identifying perceptual features which underlies this research, it may well be worth mentioning the possibility that vowels are not identified according to features at all, but rather as gestalts -- unanalyzed or unanalyzable wholes. Perhaps [ɪ] is recognized as [ɪ] because it is simply qualitatively different from other vowels; perhaps it shares no perceptual characteristics with any other vowels. In this view, there is no concept of degree of similarity between vowels. Vowels are either the same or different.

This approach is completely untenable. The work reviewed in Chapter 2 is offered as evidence that perception of vowels depends upon categories such as those proposed by linguistic feature systems. But other considerations lead to this conclusion as well.

For example, if listeners are asked to identify vowels out of context, the errors made are not random. Peterson and Barney (1952) had 70 listeners identify ten tape-recorded American vowels, 152 tokens each (two tokens from each of 76 speakers). If no two vowels had anything in common perceptually -- if all were equally "different" from one another -- then the errors would be randomly distributed over all the incorrect responses. There would be no basis on which to predict any particular errors. As might be expected, a random distribution of errors was definitely not the outcome. Vowels that were adjacent to the intended vowel, in terms of both articulation and acoustics were by far the more likely errors (op. cit., Table I, p. 182; in Lehiste 1967, p. 125). For example, the vowel [ɪ] was heard by listeners as [ɛ] about 7% of the time. No other single error accounted for even .5% of the responses to [ɪ]. When [ɛ] was presented, [ɪ] received 3% of the responses, and [æ] received 9%. Again, no other response achieved even .5% of the total responses to [ɛ]. Back vowels showed the same sort of asymmetry in their responses.

These results demonstrate, if it need be demonstrated, that some vowels are more confusable, that is, more similar to one another, than are other vowels. And this is tantamount to saying that there are auditory qualities, or properties, which vowels share to a greater or lesser degree, i.e., features. Errors in backness were considerably fewer than height errors involving vowels identical in backness.

Assigning the values front and back is less subject to error than assigning values for height. Also, errors of one degree of height (or first formant similarity) were generally much greater than errors of two degrees of height. This study not only supports the idea of distinctive features in perception, but supports the plausibility of the well-known features called Height and Backness as well.

In a study by Wickelgren (1965), subjects were asked to listen to six CVC items differing only in vowel quality, copy them down, cover their answers, then write them down again in the correct order. Confusions were tallied in a vowel-by-vowel matrix, but only those stimuli were counted that were copied correctly the first time. It is not perceptual confusions that are being studied, as in the Peterson and Barney experiment, but confusion in short term memory. If the stored information were in the form of whole phonemes, then the recall errors ought to have been randomly distributed over the vowels in the set. But the results were strikingly systematic, as can be seen from studying Figure 1.2 (a graphic rendering of Wickelgren's Table VI, p. 568).

The figure lists the six stimulus vowels [i ε æ υ ʌ α] horizontally and shows the conditional probability of each error vertically. Notice that each topmost vowel (the most likely error for each vowel) differs from the intended sound along the feature of Height, and, with the exception of the [æ] response to [i], differs in only one degree of Height. ([i υ] = high, [ε ʌ] = mid, [æ α] = low.) The next most probable errors, in general, involve either a different error of Height, or an error in Backness with Height correct. An example of the first error type is the [i] response to [æ], and an example of the second is the [ε] response to [ʌ]. Finally, the two least likely errors for each vowel differ from the target in two features. In the [ε] column, for example, the two least probable errors [α] and [υ] share values on neither Height nor Backness.

This evidence strongly suggests that the short-term storage of vowel sounds is accomplished in terms of distinctive features. Forgetting, when it occurs, affects feature values, not entire segments. This result does not directly imply that perception is accomplished by means of feature decisions. But if segments were perceived as wholes, it is unlikely that those segments, once identified, would then be broken down into features. The work of the speech perception mechanism, after all, is to map incoming sounds (through various intermediate stages) onto meanings. If there is a level of representation of whole phonemes (or whole phonetic segments), it would seem highly inefficient not to chunk them directly into representations of whole words or morphemes.

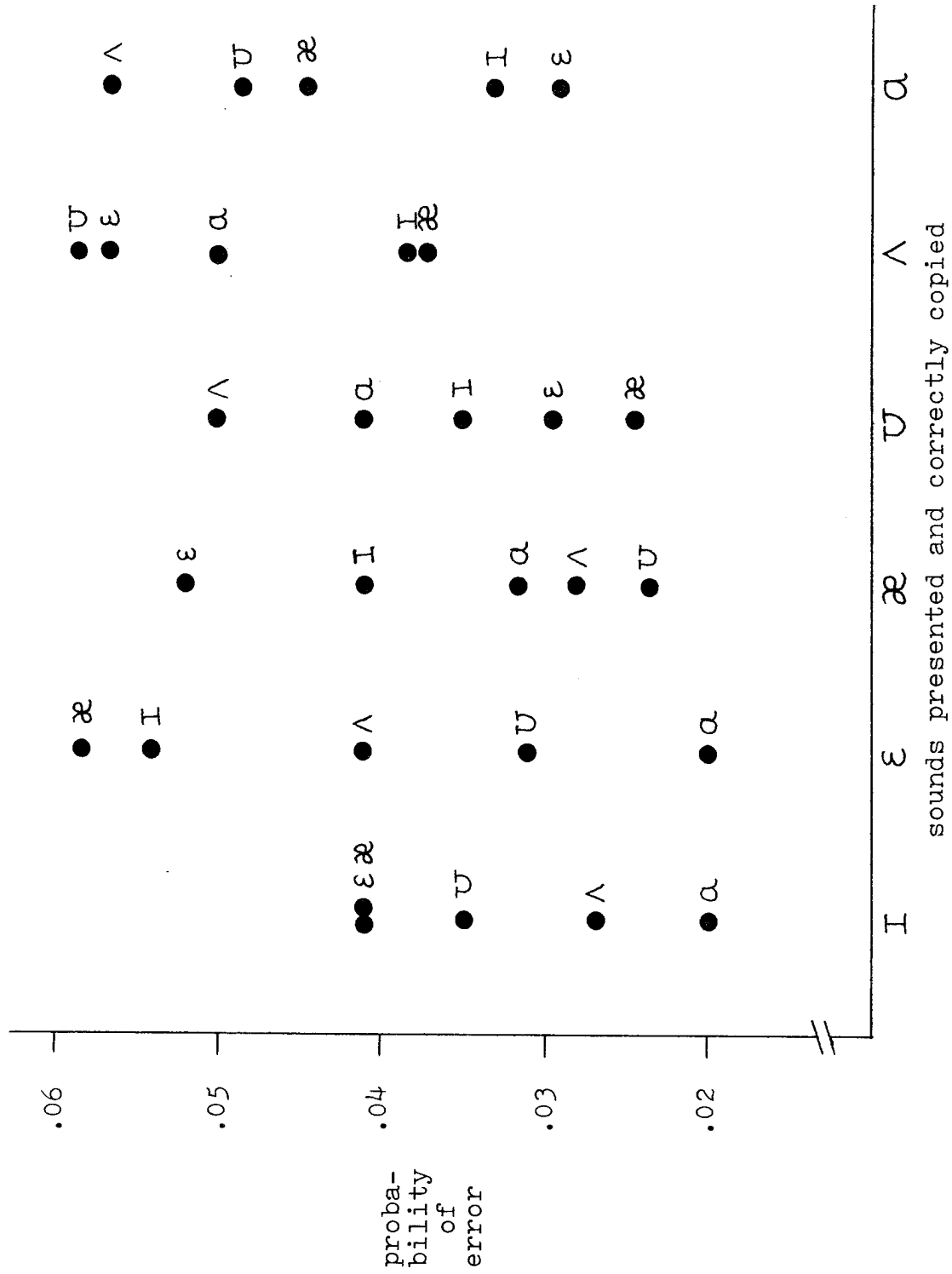


Figure 1.2: Probability of error in recall of vowels. From Wickelgren(1965), Table VI, p. 586.

1.3 The notion of perceptual distance

Postulating a feature analysis as the basis for perception of vowels has an important implication. Namely, it suggests the idea of defining "distance" between sounds in the perceptual domain. Sounds that have all but one feature in common can be thought of as "closer to" each other than sounds which share only one feature. This can be given a spatial interpretation, as shown in Figure 1.3.

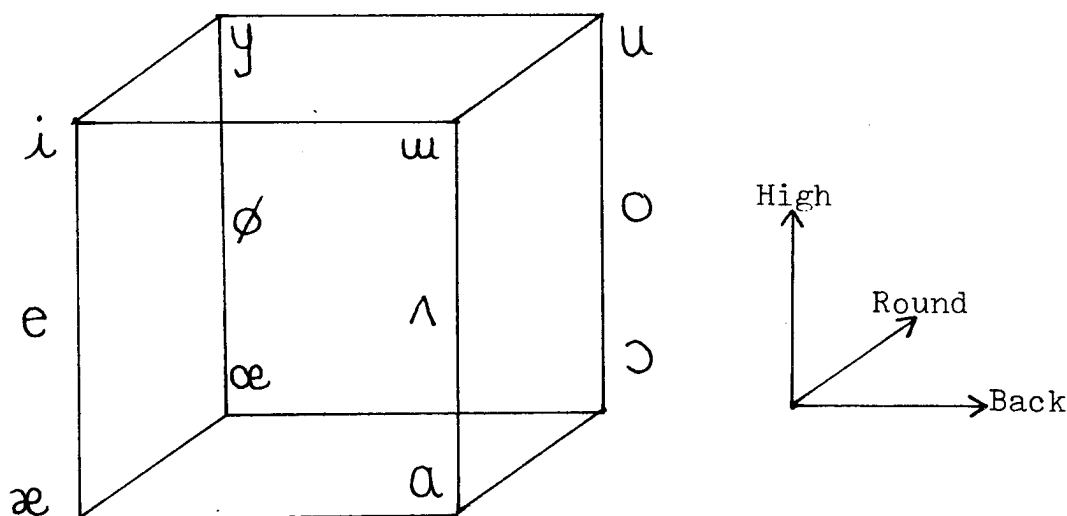


Figure 1.3: Spatial display of 12 vowels arranged according to 3 linguistic features.

In this figure, the traditional phonetic features Round, Height and Backness are drawn as three-dimensional co-ordinates, and twelve vowels have been arranged on a cube according to their feature values. Each side has a length of $\underline{1}$. The Height feature, the only one which is not binary, is arbitrarily regarded as equally "long" as the other features, in spite of the existence of a mid value of Height.

Since the difference between $[\dot{\text{i}}]$ and $[\text{y}]$ is only one feature value, or the length of one side, the distance between the two is $\underline{1}$. $[\text{æ}]$ and $[\text{a}]$ are similarly distinguished by one feature, and their distance is also $\underline{1}$. The distances $[\dot{\text{i}}] - [\text{e}]$, $[\text{ø}] - [\text{œ}]$, and others involving vowels that differ by only one degree of height would be equal to $\frac{1}{2}$. What of the distance from $[\dot{\text{i}}]$ to $[\text{u}]$? Two features are involved, and the distance must be greater than 1. However

that distance is computed, it is the same as the distance from [a] to [u], or [y] to [ʊ]. Finally, some vowels are three features apart: [i ɔ], [æ u], [y a], and [œ ʊ]. The difference of three features implies an even longer distance.

The use of three dimensions in this illustration is in no way intended as a limitation on the number of theoretically possible features. Features of Tenseness, Duration, Retroflexion, or Nasalization, to name other possibilities, could easily be added. However difficult it is to visualize a space of more than three dimensions, any distance measure one adopts can be extended to any dimensionality.

There are two intuitively reasonable ways to compute the distances involved. One is by counting features, that is, taking a "city block" approach (c.f. Attneave, 1950). The distance from [e] to [a] would be the distance from [e] to [æ] to [a] (or, equivalently, [e] to [ʌ] to [a]). The second is to compute straight-line Euclidean distances from one point to another. In this case, the [e a] distance would be $(\frac{1}{2}^2 + 1^2)^{1/2} = 1.12$. [ɪ] to [a] and other two-feature distances would be $2^{1/2} = 1.4$. The exposition and analysis methods used operate in terms of a Euclidean model. Some practical and theoretical consequences of that approach will be mentioned as such remarks become appropriate.

Distances between phonemes can be defined in terms of acoustic and articulatory properties as well. The first and second formants of a set of vowels could be plotted against one another, allowing an acoustic definition of vowel distance: the (relative) length of a line connecting two vowels on the F1-F2 plane. Similarly, by adopting suitable conventions for measuring tongue positions or tongue shapes associated with the articulation of sounds, articulatory distances could be given. But in the case of articulation and acoustics, there seems to be little reason to do so.¹ In the perceptual domain, however, the concept of distances between sounds is crucial in investigating features without a priori assumptions about what those features are. A basically unprejudicial way of studying perceptual features is to experimentally derive a set of distances between vowels, and subsequently try to map them into a spatial display. This latter process, by now a well-established psychometric technique, goes under the general name of multidimensional scaling (MD-scaling, for convenience). The derived space must then be examined for axes which are interpretable in terms of known properties of

1. Perhaps a notion of "articulatory distance" would help explicate the description of phenomena such as vowel harmony.

Chapter 2

Methods and Results: A Review

2.0 By way of introducing the present approach to the study of perceptual space, this chapter is organized in the following way. First, a description of several possible data-collection methods is given (2.1), leading to the decision to use triadic comparison. Second, a brief description of MD-scaling methods is provided (2.2), followed by a review of four studies of vowel quality (2.3). A more powerful scaling method ("three-mode" scaling, as developed by Carroll and Chang, 1970, and Harshman, 1970) is then introduced in Section 2.4 as a way to investigate the previously unresolvable question of "psychologically real" perceptual dimensions and to consider previously unasked questions concerning universal vs. language-specific perception. The chapter concludes (Section 2.5) by presenting the results of some tests of the method using both synthetic data and real data collected by other investigators.

2.1 The collection of perceptual data

There are a number of ways to derive a set of perceptual distances for speech sounds, which are reviewed by Pols et al (1969). They mention identification (= confusion) tasks, semantic differential,¹ distance estimation, and triadic comparison tasks as the available methods. Since that time, Singh (1974) has published an analysis of fricative consonant perception based on reaction time, which counts as a fifth method.

Three very practical criteria directed the choice of experimental method used in this study. First, the method used should involve as little a priori input on the part of the experimenter as possible as to just what the perceptual dimensions might be. Second, it should engender as little idiosyncratic behavior from subjects as possible. Finally, the method ought to be one that has been used somewhat successfully by others. Problems of testing and evaluating a new method could only distract attention from the perceptual questions under investigation. The use of three-mode MD-scaling was a sufficient methodological challenge in itself. This third criterion eliminates choice reaction time (where distance is inversely related to the amount of time it takes to decide whether or not two stimuli are the

1. Semantic differential data appear as matrices of stimulus-by-semantic scale, not as distances between vowels. Factor analyzing such a matrix is equivalent to applying MD-scaling to a distance matrix.

same or different). This technique was insufficiently known to the author at the time the research was planned and executed, and no precedent in the literature was found.

All three criteria led to the rejection of the semantic differential technique (Osgood et al, 1957). In this technique, listeners are asked to rate stimuli according to a number of scales with semantic labels such as "light-dark," "sharp-dull," etc. If there are sufficiently many scales, and if they are sufficiently varied in content, it should be possible to tap most if not all of the ways that listeners relate sounds to semantic labels. Finally, since no vowel scaling experiments in the literature were based on the semantic differential, it was deemed too costly to explore, relative to the main goals of the research.

It seems only fair to mention in this context Eli Fischer-Jørgenson's provocative study (1967) in which listeners were asked to associate vowels with scales "bright-dark," "full-not full," "saturated-not saturated," "light-heavy," "thin-thick," and a few others. Other listeners matched vowels with colors. The purpose was to test Jakobson's hypothesis (1941) that speech sound perception and color perception were closely related. She found that "the vowels [i e ε y u ø æ o ɔ] showed a gradual change from the color with the highest specific brightness, yellow, ([i e ε y]) over red ([ø æ o]) to blue ([ɔ p ɔ]), which is a specifically dark color (p. 669)." Brightness of colors, then, is associated mainly with the first formant of vowels, with some influence of the second. This study is fascinating in its own right, but it does not offer evidence that any particular ordering of vowels represents a perceptual dimension.

Of the three remaining methods, triadic comparison and distance estimation require that subjects overtly assess the similarity between sounds. This is a somewhat artificial task, compared with the identification type experiment. In an identification task, subjects simply listen to stimuli and identify them. Responses are entered in the appropriate cells in a stimulus-by-response matrix (= confusion matrix). Small off-diagonal entries represent relatively large distances, and large entries represent small distances. The identification task is certainly more closely related to what one does in the ordinary process of hearing and understanding speech.

But there are two difficulties in converting confusion matrices into vowel distances. The first is that confusions are often not symmetric. The number of times stimulus a evokes a response b may be very different from the number

of times stimulus b evokes response a. This is easiest to demonstrate if the confusions are induced by low pass filtering, say, at 1000Hz. Under this condition, [i] will sound very much like [u]; the reverse, however, is not true at all. If the experimental measure of similarity is changed by considering the vowels in a different "direction," then a spatial model would not seem to be applicable.

The usual solution (Shepard, 1973) is to average corresponding values across the main diagonal, though this hardly overcomes the objection. Van der Kamp and Pols (1971) compare several methods of analyzing an asymmetric matrix. Two of them involve obtaining different spatial configurations for responses and stimuli. It is not at all clear how such solutions ought to be interpreted. A third method is to make the matrix symmetric by treating the asymmetry as response bias (Wagenaar, 1968). They found that analyzing such a symmetric matrix by Kruskal's (1964) nonmetric scaling procedure gave essentially the same result as when triadic comparisons on the same vowels (Pols, et al, 1969, described later in this chapter) were analyzed by Kruskal's method. In spite of the similarity of results of the two methods, analyzing away all asymmetry as response bias requires some explanation.

The second disadvantage to the use of confusion matrices as a source of intervowel distance is that it is possible to derive matrices which have a large number of zeros, that is, with relatively few confusions. These zeros will correspond to very large, equal distances in the derived distance matrix, the closer one gets to the limiting case of n-1 dimensions for n vowels.

One way to handle this problem is to devise an analysis procedure that is not disturbed by excessive zeros. Shepard (1973) obtains good results by assuming that the relationship between distances in the space and values in the confusion matrix are not linear, but exponential. Applying this method to Miller and Nicely's (1955) data on consonant confusions (16 consonants, all unfiltered noisy conditions summed into one matrix), a two-dimensional space was obtained in which sounds are clearly grouped according to nasality and voicing.

Shepard also analyzes Peterson and Barney's (1952) vowel confusion data (ten vowels, 152 presentations each, responses summed over 70 listeners). In this case, a two-dimensional space, which reflected basically height and backness, accounted for 97% of the variance in the data. However, the large number of zero or near-zero values in the confusion matrix allows the program considerable leeway in assigning co-ordinates. Fifty-four cells of the 100 have values of 3 or less, where each row sums to over 10,000

responses. While one cannot help but be impressed that such interpretable results emerge from Shepard's reanalyses, it is nevertheless true that the data he treated did not contain at all as much information about perceived differences between segments as one could hope for.

A second way of treating the problem of too many "holes" in a confusion matrix is to create a means of filling them in. Van der Kamp and Pols do so by using a confusion condition that promotes a large amount of errors: their subjects had to identify vowels which lasted only 8 milliseconds, or one pitch period. This resulted in few zero cells, but the asymmetry problem remains. [e], for example, was heard as [æ] 22 times, but [æ] was heard as [e] only 8 times.

Many of Miller and Nicely's extreme confusion conditions, such as the -18dB signal to noise ratio, resulted in the filling in of many cells. But in this particular condition, many off-diagonal values were larger than the diagonal ones. [ʃ] was correctly perceived as [ʃ] 7 times under this condition, whereas it was perceived as [s] 14 times and [b] 20 times. [g] was perceived as [g] 14 times, as [b] 26 times, and as [k] 29 times. What sort of distance model will account for that? Clearly, there may be difficulties in adding too much confusability.

In spite of the naturalness of the identification task, an overt judgment method was used. Listeners are presented with more than one sound and make a response based on their perceived similarity or difference. Two difficulties are thereby avoided completely: (1) The asymmetry problem effectively disappears. (2) There are no confusion conditions. The stimuli are presented under good conditions of frequency response and noise level intended to promote correct identification. This insures that the dimensions which ultimately emerge from the analysis are not artificially generated by any confusion-producing scheme.

There are three overt methods to choose from. The first is free magnitude estimation, in which the listener is presented with a pair of sounds and requested to give a number reflecting the perceived distance between the sounds. Often there is an anchor value given. (If [ε æ] is a distance of 3, what is the distance for the pair [o α]?) This is the method used in Hanson's (1967) main experiments. A more convenient method is the rating scale. (If 1 means "very similar" and 9 means "very different," what number from 1 to 9 gives the relationship between [u o]?) Mohr and Wang (1968) used rating scales in their experiments.

The last method, and the one adopted here, is triadic comparison (c.f. Torgerson, 1957, for a thorough discussion). This is the simplest overt task a listener can have: three sounds are presented, and the subject judges which two are most similar, and which two are most different. No response as to degree of similarity is required. Pols et al (1967) used triadic comparisons; the design of the current work draws heavily on theirs.

The conversion of triadic judgments to dissimilarities is described in detail in Chapter 3. The consequence of directly interpreting those dissimilarities as Euclidean distances is illustrated in the re-analysis of Pols et al's data presented later in this chapter.

2.2 The use of 2-mode MD-scaling

Once the triadic comparisons have been made, and the dissimilarities between vowels have been obtained, the data are ready for submission to an MD-scaling program. This section will describe two scaling models,² the assumptions that each makes of the data, and the strength of claims that can be made in the interpretation of the results. In both cases, data submitted to the program is in the form of a 2-dimensional array of vowel-by-vowel information, hence the term "2-mode" scaling. A recent, thorough review of scaling methods is Shepard et al (1972), Multidimensional Scaling, Volumes I (theory) and II (practice).

Traditional MD-scaling is generally traced to Torgerson (1957). This method requires that the dissimilarities matrix be transformed into a best guess matrix of absolute Euclidean distances. If the data are dissimilarities based on triadic comparisons, an additive constant must be estimated and added to each cell in the matrix (c.f. Section 4.4 for justification). A further transformation suggested by Torgerson is a stretching of large distance values, which corrects for three effects: (1) a ceiling effect for large distances due to the mechanics of deriving the dissimilarities from the raw responses; (2) a regression towards the mean for all distance values as the data are summed over all the listeners; and (3) underestimation of large distances in the listeners' judgments.

The new absolute distances are transformed into scalar products and then submitted to the program. The output

2. Shepard's exponential fitting method, which was devised for the skewed distribution of values toward zero, found in many confusion matrices, would be a third.

consists of a set of co-ordinates in an n-dimensional space. n is specified by the experimenter for each analysis. At this point the skill of the investigator comes into play, as he selects one of the n dimension spaces and attempts to interpret the results.

The ultimate decision as to which solution is correct is usually based on interpretability of the space. One rotates the space to find axes which correspond with known properties of the stimuli. If a 2-dimensional result is interpretable, and a 3-dimensional space for the same data is not, then the 2-dimensional space is likely to be regarded as "correct."

A supplementary criterion is the number of dimensions vs. a measure of goodness of fit of predicted distances to experimental distances. A standard measure is

$$1 - \frac{\sum (x' - x)^2}{\sum (x - \bar{x})^2}$$

where x' values are predicted values based on output co-ordinates, and x values are data values. (Torgerson's b* method involves double centering the response matrix, reducing the denominator to $\sum x^2$. This measure will be called "fit," after Cooper, 1971.) It is expected that fit will improve with number of dimensions, but there will be a point of diminishing returns, seen visually on a fit vs. dimensionality plot as elbows in the curve.

The second scaling method involves considerably fewer assumptions about the properties of the data matrix. We refer to non-metric scaling as introduced by Kruskal (1964), implemented by Kruskal and by Young and Torgerson (1967). This very powerful method derives a set of co-ordinates based on a set of input distances, but finds the best-fitting monotonic transformation of the data as well. Computation of the additive constant and the non-linear stretching of distances is no longer necessary, since the program is free to stretch or compress any distance values, large or small. The only requirement is that rank order of data values be preserved.

The evaluation of a non-metric scaling solution is the same as for a non-metric one: rotate the space in hopes of finding interpretable axes, and consult the fit curve for a likely limit to the number of dimensions involved. Kruskal (1964) computes a measure called "Stress" which decreases with goodness of fit of the space. Stress is equal to

$$\frac{\sum (x' - x)^2}{\sum (x - \bar{x})^2},$$

where x values are the optimally monotonic transformations of the original data. Stress is used in exactly the same way as "fit," except that the sought-for "elbow" in a stress curve is concave upward, stress decreasing as more dimensions are extracted.

The disadvantage of both these methods is that while the configurations are stable, the co-ordinate systems are not. By way of illustration, imagine that a number of people estimate the distances between ten American cities, after which their data are pooled. We would expect either of the above scaling approaches, metric or non-metric, to derive a space that bears some resemblance to the way those cities appear on a map of the United States. But what determines the orientation of the space? It is completely arbitrary. And the fact that there are standard rotation algorithms -- such as rotation to simple structure, where the co-ordinate axes are rotated to maximize the number of near-zero loadings -- makes it no less arbitrary. There is no information in a pooled data matrix that could bear on any choice of rotation. Only the constellation of points has psychological reality. Any axis in the space that correlates with properties of the stimuli can only be proposed as a possible candidate for a psychologically real dimension. Any claim that a particular dimension represents a psychologically real property, according to which the experimental judgments are made, is simply not justified.

2.3 Four 2-mode studies of vowel perception in light of the problem of rotational indeterminacy

The following four studies, all using overt similarity or difference judgments, are reviewed with this particular point as the main focus.

The most extensive study of vowel quality using MD-scaling was done by Hanson (1967). He reports on 74 "experiments," that is, 74 variants or replications of the basic technique: obtaining listeners' judgments of similarity between Swedish vowel sounds. The experimental variables included number of stimulus vowels, synthetic vs. natural speech, experimental tasks, subjects' hearing ability, and filtering conditions for the stimuli. Our discussion will be based on experiments 22-25, which Hanson regarded as the main experiments, and experiments 54 and 55.

Experiments 23-25 consist of replications by the same subjects, on successive days, of experiment 22, which is summarized as follows:

1. The stimulus set consisted of the nine Swedish long vowels [i e ε a o u y ʌ ø], in Swedish orthography "i e ä a a o y u ö."
2. The stimuli were synthetic vowels of approximately .5 sec. duration and constant F0, produced by an OVE II synthesizer.
3. The task was paired comparison, in which listeners gave numerical estimates of the perceived difference between each pair of vowels.
4. Vowels were presented by tape recorder in a randomized series.
5. Each pair of vowels was judged twice, once in the order ab and once in order ba. Every listener therefore heard $n(n-1)=72$ presentations.
6. The anchor stimulus pair, [o ɔ], was assigned a value of 5.
7. All listeners were native speakers of Swedish.
8. Data for all listeners were summed, producing one vowel-by-vowel dissimilarity matrix.

Two metric analyses were performed: Thurstone's centroid method (Thurstone, 1947) and principal components analysis (Harman, 1960). We presume that the two sets of co-ordinates given for each experiment, labelled F and F', correspond to the centroid and principal components analyses, respectively. Since the F and F' solutions differ only slightly, we have arbitrarily chosen to discuss the solutions labelled F.

Hanson obtained solutions of 1-4 dimensions. He presents the 3-dimensional solutions, mainly on grounds of interpretability. The 3-dimensional results for all four replications are plotted in Figures 2.1 and 2.2. Points labelled A, B, C, D represent experiments 22, 23, 24, 25, respectively. Even without attempting to rotate the four replications to optimum congruence, it is clear that they all reflect the same configuration of points. (The results discussed below are essentially similar to those for experiments 54 and 55, which represent triadic and paired comparisons, respectively, on a set of vowels pronounced by Hanson.)

The first observation is that the plane of Figure 2.1 is a deformation of the F1-F2 plane (Figure 2.3). Perceptually, the distances between successive high vowels [i y ʌ u] appear more equally spaced than the F1-F2 plane would

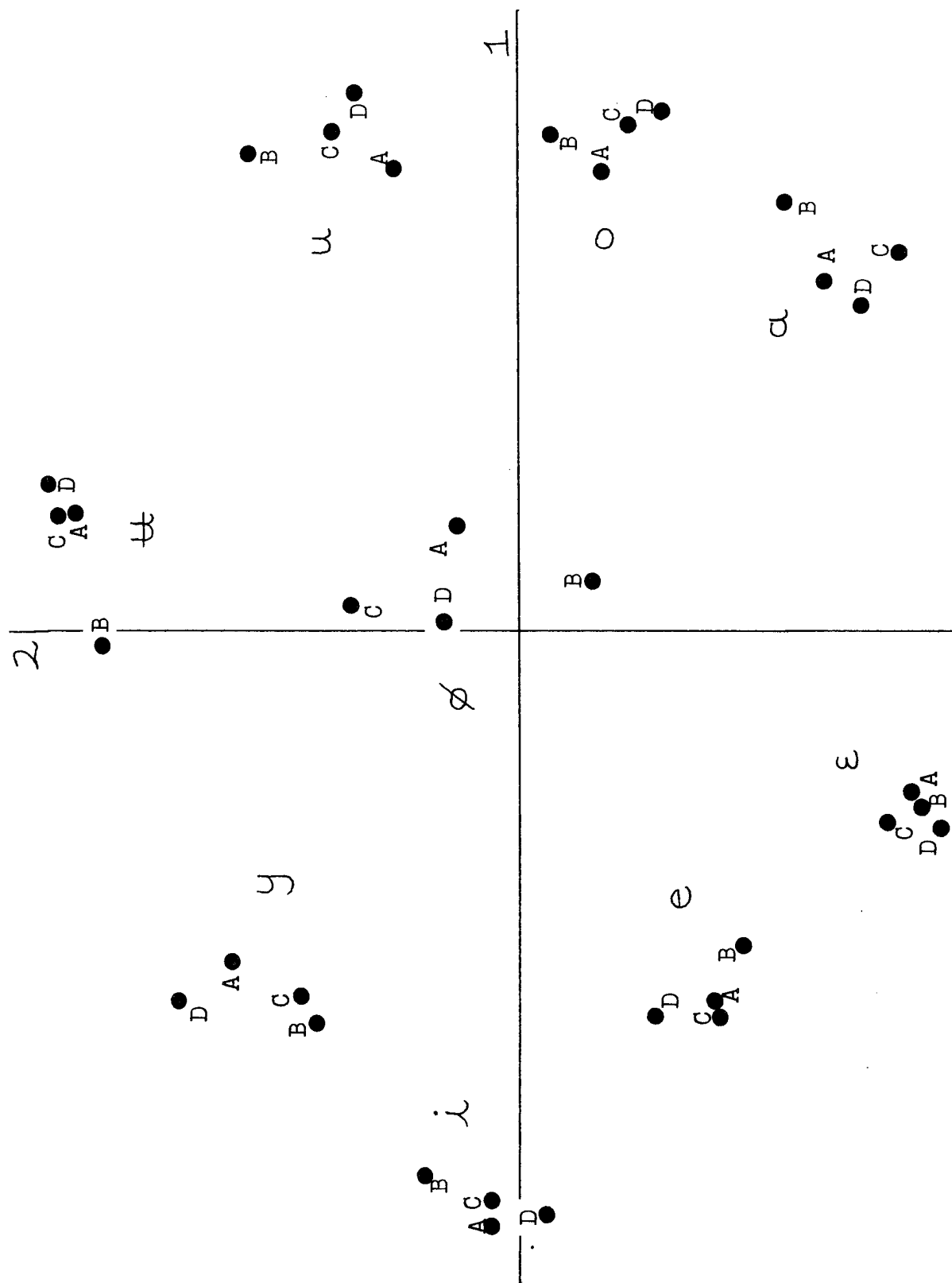


Figure 2.1: Dimensions 1 and 2 of Hanson's (1967) MD-scaling study of 9 Swedish tense vowels. A, B, C, D represent experiments #22, 23, 24, 25 respectively.

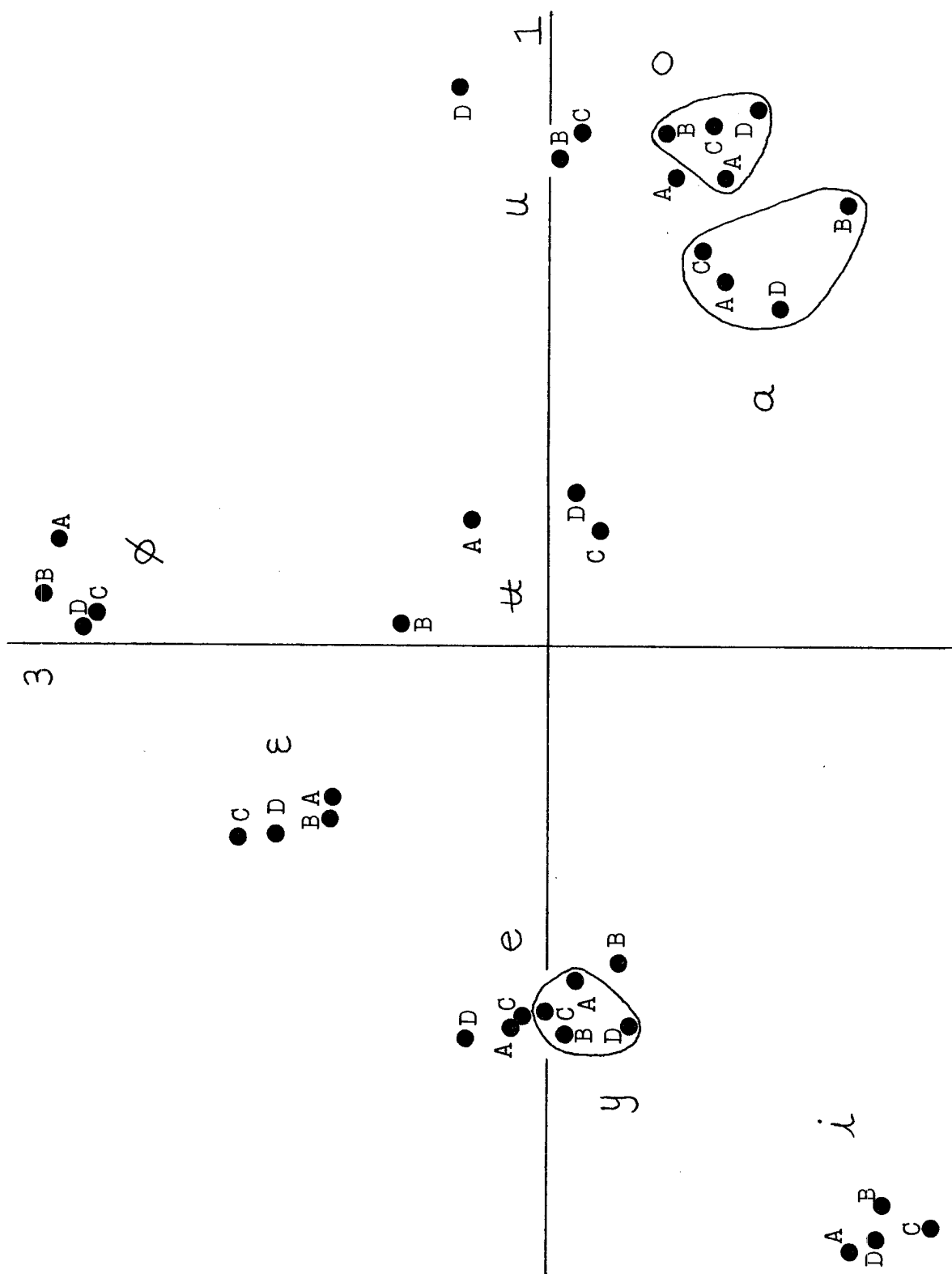


Figure 2.2: Dimensions 1 and 3 of Hanson's (1967) MD-scaling study of 9 Swedish tense vowels. Labelling as in Fig. 2.1.

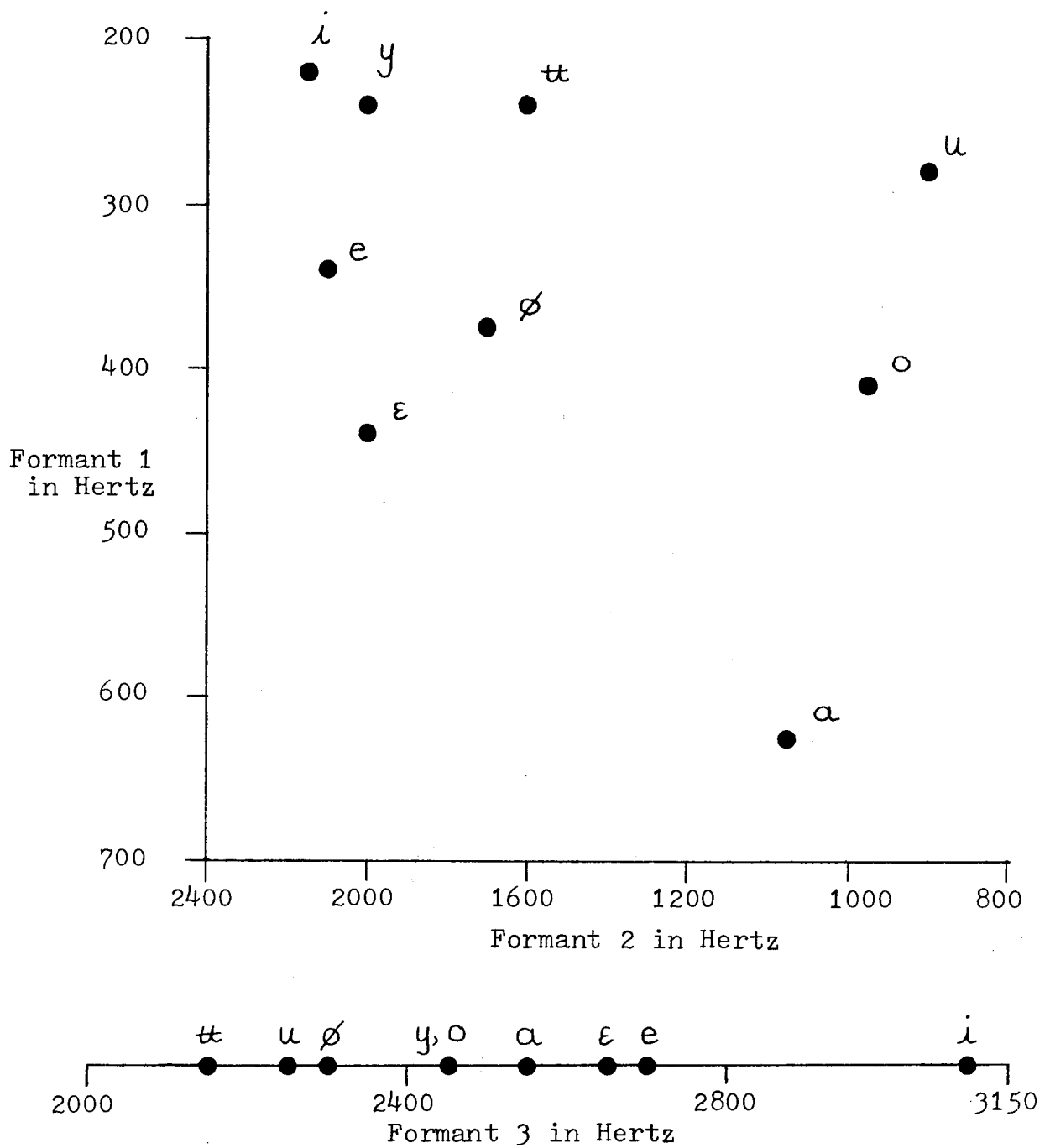


Figure 2.3: Formants 1,2, and 3 of vowel stimuli for Hanson's main experiments.

suggest. Also, the distance between extreme high vowels [i u] is underestimated, causing a curve in the pattern of high vowels. The same comments hold for mid vowels, which are more evenly spaced from one another on the perceptual plane of Figure 2.1 than on the formant plane. Also, the distance from [e] to [o] is less than the distance [e] to [ø] to [o], continuing the curvature.

Given this discussion, consider cutting the plane according to distinctive features. Figure 2.4 shows some possibilities for doing so. The points representing the vowels are approximate mean co-ordinates for the plane of Figure 2.1. The most obvious cut is between back [u o ɔ] and nonback. Round/Nonround is also straightforward. The boundary has been drawn so that [ɔ] is labelled Round; a more accurate transcription of the Swedish low back vowel, and the stimulus as well, is [ɒ], that is, [ɔ] with some rounding. If one argues that rounding means "closely rounded" such that [ɒ] is excluded, that is possible as well. The cut can be just as easily drawn on the other side of [ɔ].

The partitioning of the space into High, Mid and Low is just as simple, if straight lines are not demanded. The high vowels [i y ɤ u] obviously appear near the top of the figure and the low vowels toward the bottom. Labelling [ɛ] as low reflects the fact that [i e ɛ] appear in a sequence which parallels the sequence [u o ɔ]. Nothing is gained in the context of this discussion by insisting that [e] and [ɛ] are both mid.

These are by no means the only possible ways to divide the space according to linguistic descriptions. On the plane of dimensions 1 and 2, still, a counterclockwise rotation of the co-ordinate system establishes a projection with [i y] at one end and [o ɔ] at the other, approximating Lindblom and Sundberg's (1971) palatal-pharyngeal direction. And on the plane of the first and third dimensions (Figure 2.2), a 45° rotation orders the vowels according to three clusters: [i y e ɛ ø] front, [ɤ] central, and [u o ɔ] back. Doubtless the 2-3 plane, to say nothing of other rotations not easily observed in Figures 2.1 and 2.2, would yield other plausible interpretations. It hardly seems necessary for Hanson to postulate a "perceptual contrast" factor, dimension 3, which has no relationship to any known linguistic descriptions.

When Hanson reanalyzed the results of experiments 22-25 using Kruskal's nometric method (1964), he was able to interpret the dimensions of the 3-dimensional space as rounding, height, and backness, in spite of a curious unwillingness

to rotate the solution (p. 142-144). The "perceptual contrast" factor had disappeared.

It is worth commenting on Hanson's choice of the three-dimensional nonmetric solution over the two-dimensional one: "...the third dimension corresponds to F3 and a two-dimensional specification is therefore not an acceptable solution (p. 45)." But the two-dimensional solutions each have one dimension that corresponds at least as well to F3 as do the touted third dimensions of the 3-D spaces. More puzzling yet is his insistence that F3 is important. It is true that D3 (and D2 from the two-dimensional solutions) are interpretable as good rounding dimensions. Hanson, however, says, "It [dimension 3] corresponds perfectly to a p/f feature [plain/flat, that is, rounding] and F3 (p. 145)." While the pattern of vowels on F3 (Figure 2.3) does find unrounded [i e] at one end, the lack of a convincing gap on F3 between round and non-round vowels vitiates any idea that F3 corresponds "perfectly" to rounding. And F3 corresponds only imperfectly to the perceptual dimension as well.

Given the embarrassing richness of alternate descriptions for Hanson's results, the one most important question to be asked is this: What indeed are the properties of the sounds that might be called perceptual dimensions? It is clear, at least in the case of dimensions 1 and 2, that the F1-F2 plane is in some sense the source of the listener's responses. If it were not, there would be good reason to conclude that they were not responding to vowel quality at all. But what are the perceptual equivalents of distinctive features? On what basis does the listener classify and compare vowels? No answer is possible. There is simply no intrinsic orientation of the space. Rotating the space to a pattern that corresponds to a known distinctive feature (or formant, perhaps) is no demonstration that such a pattern is perceptually real. An argument of the form "x, therefore x" has little to recommend it.

The next major MD-scaling study of vowels to appear in print was that of Pols et al (1969). They had 15 native Dutch subjects (and three non-native) make triadic comparison judgments on 11 Dutch vowel sounds [i e ε a a ɔ o u y ø œ]. The vowels were spoken by a native Dutch speaker, but equalized for duration, subjective loudness level, and pitch artificially. The entire experiment was run using an elegant computerized system whereby listeners pressed buttons to play the stimuli within each triad in any order they wished, and to signal their judgments, which were recorded by the computer. Judgments were summed across listeners.

Kruskal's non-metric scaling technique was used on the summed vowel-by-vowel matrix (Kruskal, 1964 a, b), and on the basis of interpretability and the elbow in the fit-dimensionality curve, the 3-D solution was presented as the most reasonable one. Figure 2.5 shows one plane of the solution in three dimensions. It is possible to interpret the plane using the labels Round/Nonround, High/Mid/Low, and Back/Nonback, although the configuration itself is not as satisfying as the D1-D2 plane of Hanson's solution. [y] is inexplicably close to [u], and [e] and [ɛ] are much closer to [ø] and [œ] than would be predicted from the F1-F2 plane of the stimuli (Figure 2.6). In order to gain a better intuitive grasp of the space, a wire model of the space was constructed and examined. But no rotation resulted in a plane more interpretable than Figure 2.5.

The position of the mid vowels is accounted for by the acoustic properties of the sounds. A physical space was derived by factor-analyzing an 11x18 matrix of 11 vowels by 18 amplitude measurements, each for successive 1/3 octave bands in the vowel formant range. When the 3-dimensional space resulting from this analysis was compared with the 3-dimensional perceptual space (Cliff, 1966), the correspondence for all vowels was good, except for [y]. The authors suggested two conclusions: (1) the F1-F2-F3 space is not the best 3-D physical space for the description of vowels; (2) [y] requires a fourth acoustic dimension to account for its position in the perceptual space. Unfortunately, the stimulus token is not to be faulted; all or most of the 15 listeners (the text is unclear) are said to have identified the [y] token as a Dutch [y], likewise [u] and [i].

Here, as with Hanson's results, it is clear that the acoustic properties of the sounds underlie the listeners' judgments. But there is no evidence that any rotation of the space to conform to linguistic descriptions represents dimensions or criteria that listeners use.

An investigation of the perception of American vowels was made by Singh and Woods (1971). In that study, 18 American listeners performed paired comparisons using a 7-point rating scale, rather than the free magnitude estimation used by Hanson in his main experiments (22-25). The stimuli were the 12 American vowels [ɪ ɪ e ɛ æ ʌ ɔ ɒ ʊ u ɜ] spoken in isolation. All stimulus pairs were recorded in a randomized order by a female native speaker. Although it is reported that the microphone was six inches from the speaker's mouth, nothing is said about the rate of presentation, the duration of the stimulus tokens, the relative amplitude of the stimulus tokens, their fundamental frequency,

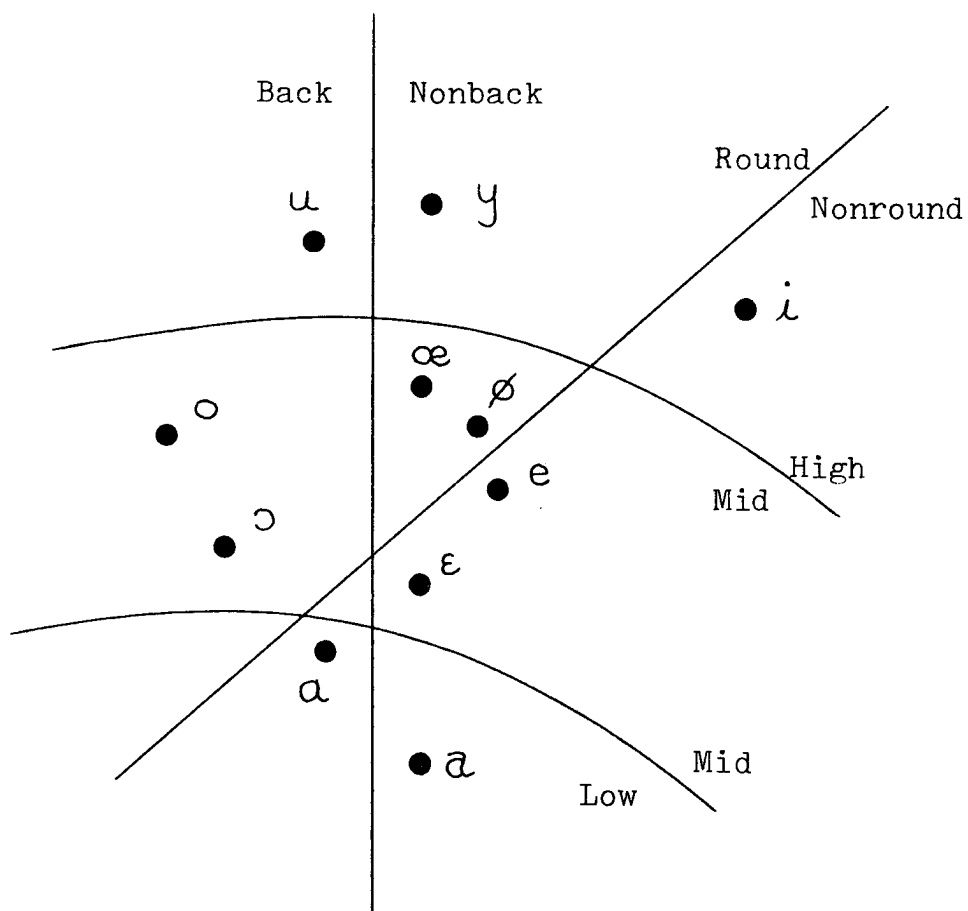


Figure 2.5: Two dimensions from the 3-dimension solution for 11 Dutch vowels given by Pols et al (1969). A possible distinctive feature interpretation is superimposed.

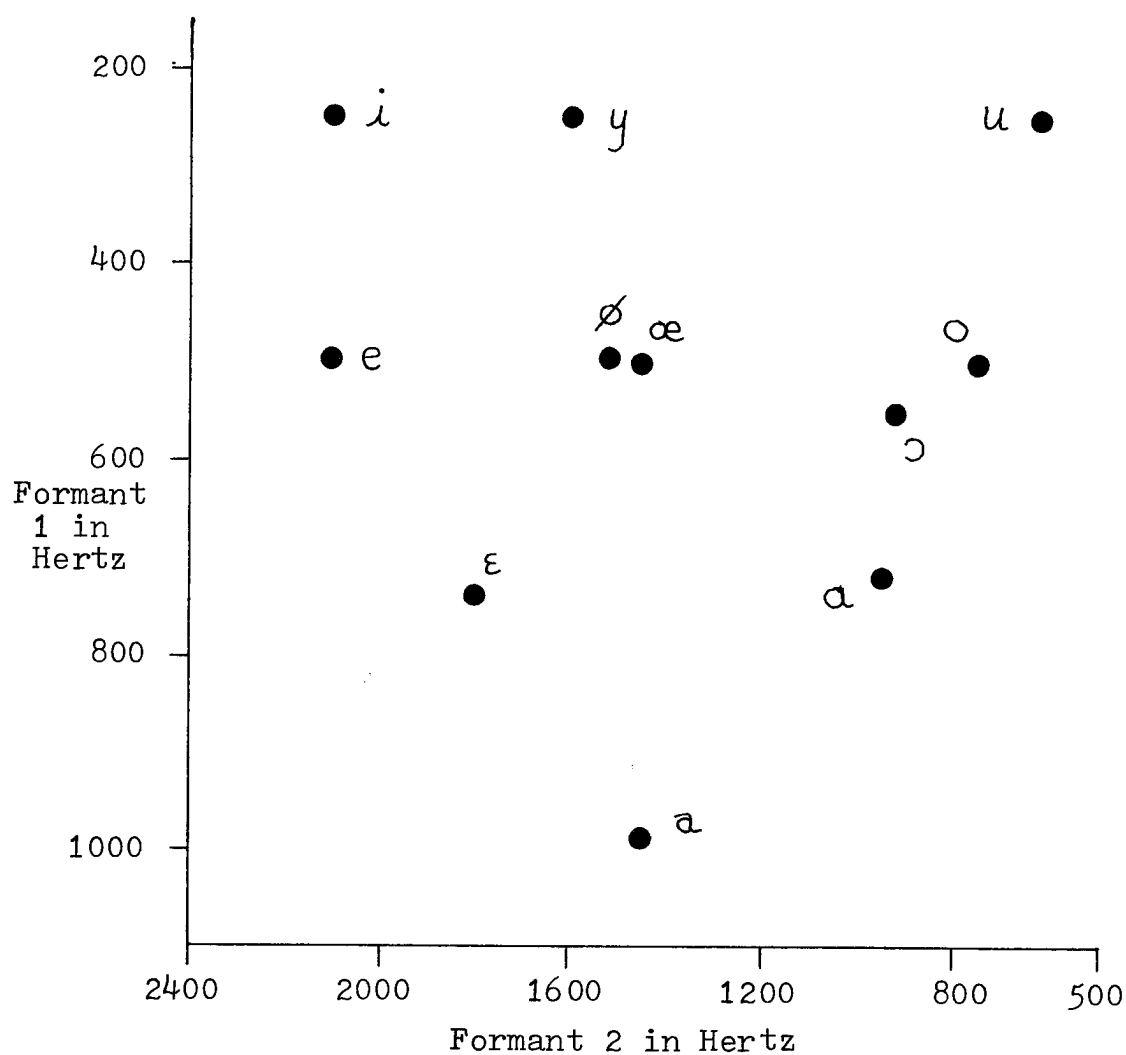


Figure 2.6: Formants 1 and 2 for 11 Dutch stimulus vowels (Pols et al 1969).

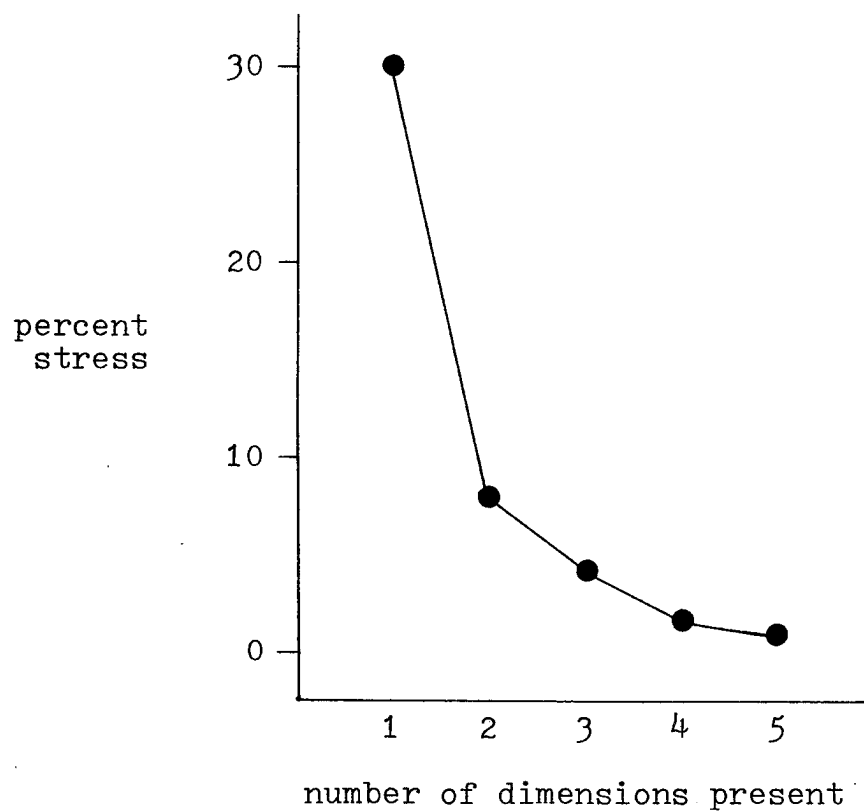


Figure 2.7: Fit against dimensions extracted for Singh and Woods'(1971) scaling of 12 American vowels.

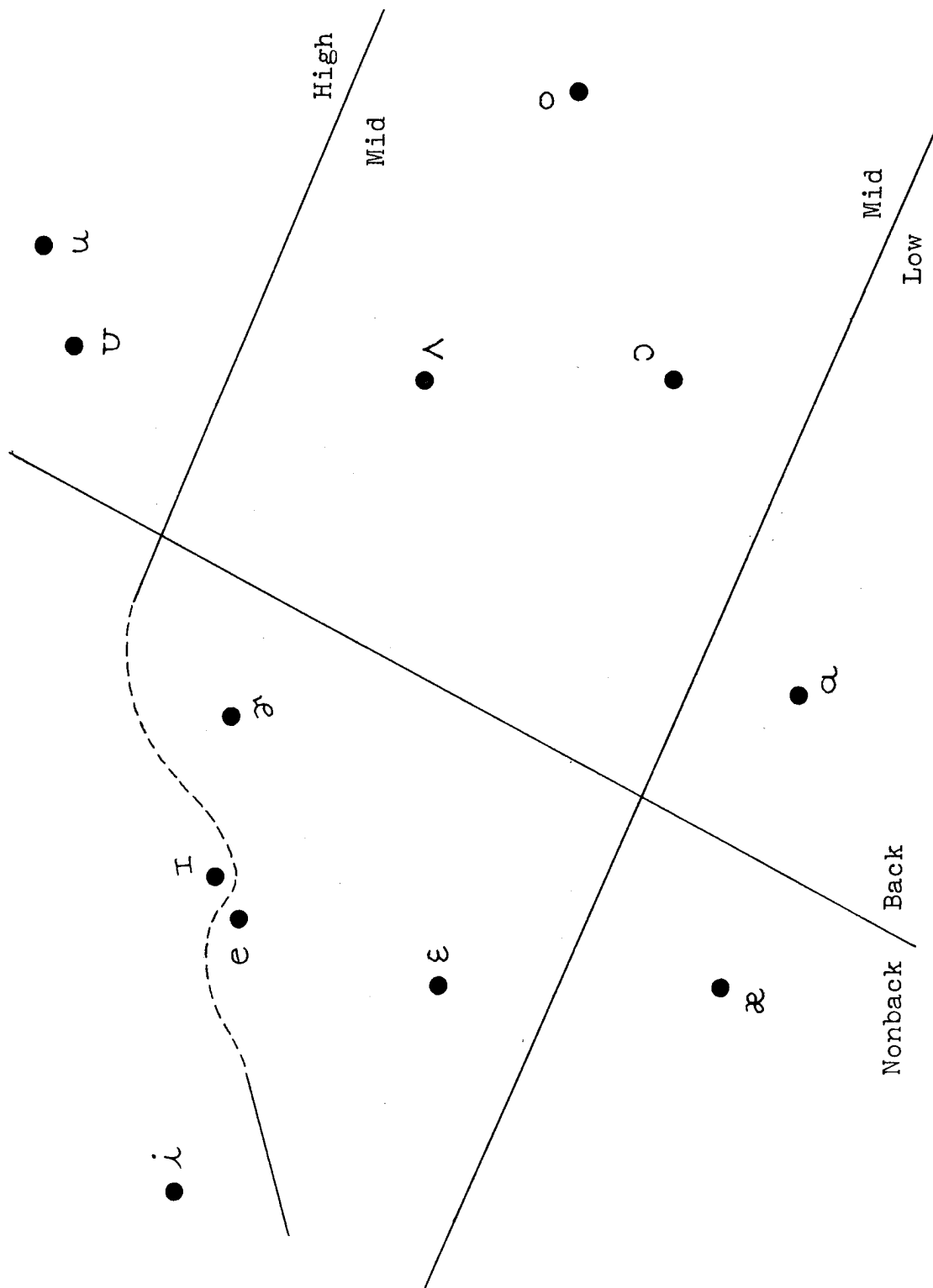


Figure 2.8: Best orthogonal rotation for Height and Backness in Singh and Woods' 3-dimension solution for 12 American vowels, with superimposed feature interpretation.

to what extent the vowels [u o e i] were diphthongized, or the amount of variation in any of the above parameters during the recording of the entire stimulus tape.

The summed judgments were submitted to Kruskal's non-metric analysis, and results were obtained in 1-5 dimensions. Only their analysis of all 12 vowels will be treated in any detail. In choosing the "correct" solution, the authors find that "the stress functions [Kruskal's goodness-of-fit measure - DT] ... did not suggest conclusively which configuration was most likely to yield the best interpretation (p. 1863)." That is, there was no single prominent elbow in the plot of stress vs. dimensions extracted. Apparently this circumstance justified abandoning the use of stress information entirely: the solution discussed, the one in three dimensions, lies on a relatively straight portion of the stress curve, between elbows at 2 and 4 dimensions (c.f. Figure 2.7). In light of this, it is difficult to understand the import of their conclusion that the stress figures "provided excellent criteria by which to judge the final configurations (p. 1866)." The judgments as to which solution to discuss was made entirely on the basis of interpretation. At least, the 3-dimensional solution was chosen over the 2-dimensional solution on that basis. No mention was made of the 4-dimensional solution at all. In all fairness, it must be pointed out that a set of analyses was performed on 11 vowels, with [æ] deleted. For these reduced data, a plausible elbow in the stress curve at two dimensions did support the authors' decision to interpret the 2-D solution.

Interpretation was guided by an explicit attempt to relate directions in the perceptual space to the linguistic features Tenseness, Backness (called Tongue Advancement by Singh and Woods), Height, and Retroflexion. Schoneman and Carroll's method (1970) was used to rotate, translate and expand or contract the 3-D perceptual space to make it as similar as possible in orientation (and, irrelevantly, in absolute values of co-ordinates) to 3-D spaces defined by two a priori distinctive features sets: (1) Height-Backness-Retroflexion, and (2) Height-Backness-Tenseness. Rounding was never considered. Basically, they found that Height, Backness, and Retroflexion were the features that fit the data best. No rotation of the space resulted in a credible Tenseness dimension.

Figure 2.8 is a plot based on the best orthogonal rotation for Height and Backness in the 3-D space. A binary Back/Nonback cut is easy to make, as drawn on the figure, although an equally plausible line could be drawn with [ɑ] classed as nonback. Information about the exact phonetic quality of the [ɑ] stimuli would have been useful, specifi-

cally, whether it was pronounced more like [a] than [ɑ]. This is hinted at rather covertly by the Backness ("Advancement") values used in the target space of distinctive features: [ʌ ɔ ɒ ʊ u] were specified as 0.0, [ɤ ɑ] as .5, and the front vowels [i ɪ e ɛ æ] as 1.0. The authors note that these values are based on "traditional phonetic features (p. 1862)." The postulation of a low central vowel phoneme apparently derives from the tradition exemplified in Trager and Smith (1957), which postulates [æ ɑ ɒ] as three low vowel phonemes. But this view is incompatible with the claim that "the complete set of American English vowels (p. 1861)," excluding diphthongs, was used as stimuli. The set is simply not complete without [ɑ], and it is reasonable that their vowel labelled "/a/" (p. 1862, Table I) be interpreted as back [ɑ] whether the backness feature allows two values or three.

Height is straightforward, for back vowels, at least; [ʊ u] are high, [ɔ ɒ] are mid, and [ɑ] is low. Only a priori beliefs can determine whether [ɛ] is mid or low. The desire to class [ɪ] and [i] as high causes some difficulty. The line running between [ɪ] and [e] is obviously not based on the natural spacing of vowels on the plane.

Dimension 3, shown as the vertical direction of Figure 2.9, nicely separates [ɤ] from the rest of the vowels and is therefore an unambiguous retroflex (= extreme third formant lowering) dimension. However, by rotating the co-ordinate axes about 45°, the space can be partitioned in the way shown on the figure.

One of the new directions classes [ɤ] apart from [i ɛ ɪ æ e ɑ ɔ ʊ ʌ], and those apart from [u ɒ]. The vague term "high frequency prominence" (HFP) can be introduced as a possible acoustic characterization. [u] and [ɒ] generally show heavy attenuation of the third and higher formants. The third formant of [ɤ], in contrast, is not only of high amplitude but intrudes unusually far into the second formant range. This particular direction in the space has been pointed out not to show how farfetched the possibilities of rotation can be, but because the results for the English listeners in Chapter 6 replicate it and support it as being perceptually real. For this reason, the HFP direction is drawn so as to set off [ɤ], not [ɤ] and [i]. Grouping [ɤ] and [i] would be just as defensible, if HFP included the phenomenon of F2 appearing very close to F3.

The second rotated dimension of Figure 2.9, separating the vowels into the two groups [i ɪ ɛ æ e ɑ] and [ɤ ɔ ʊ ʌ u ɒ], is difficult to name using an articu-

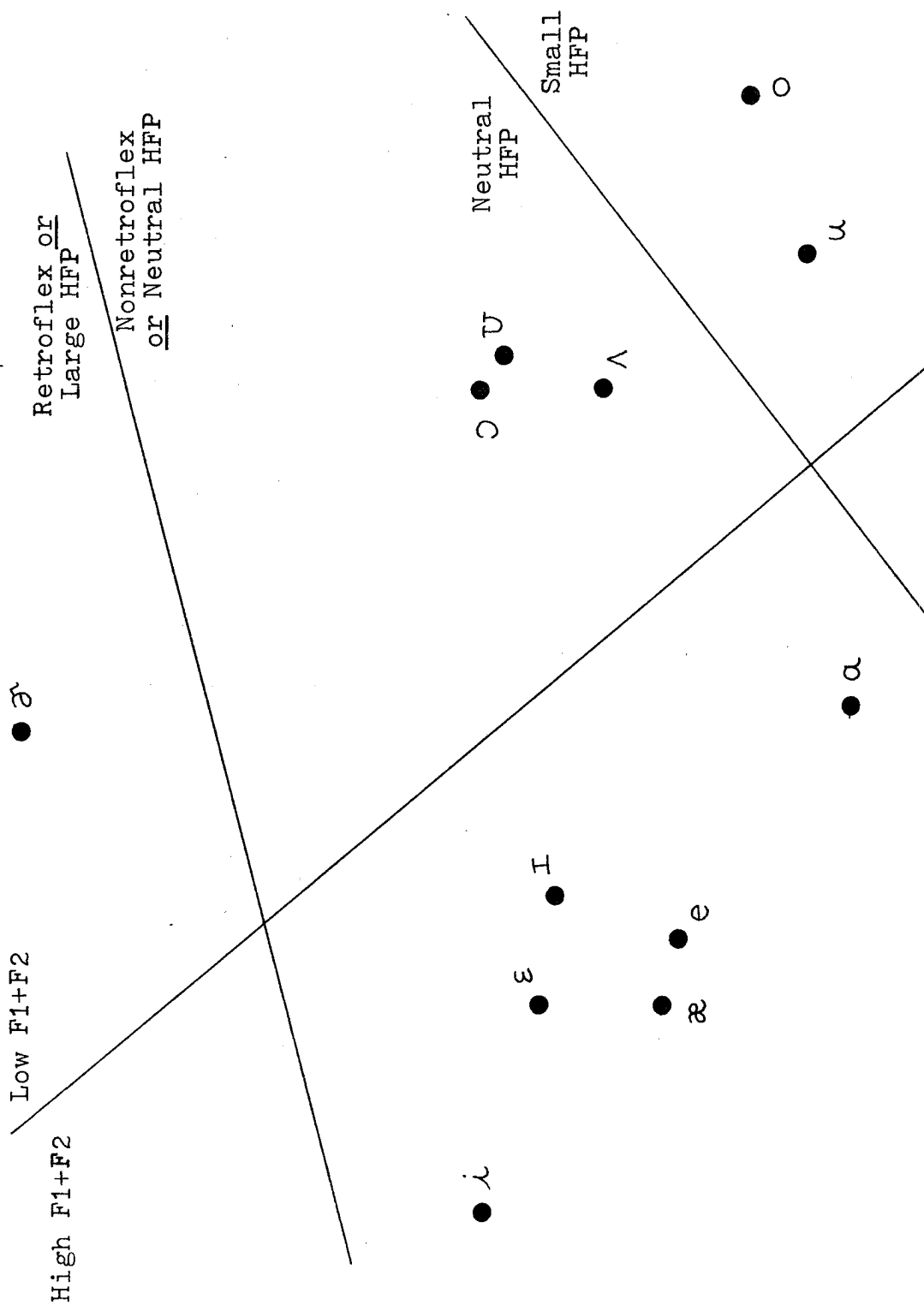


Figure 2.9: Dimensions 2 and 3 of Singh and Woods' solution for 12 American vowels, with superimposed feature interpretation. HFP means high frequency prominence.

latory label. But a first guess at an acoustic measure that corresponds to this feature is a (weighted) sum of the first and second formants, which is the way the dimension is labelled on the figure. Without knowing the actual formant frequencies of the stimuli, there is no point in speculating as to what the weights might be.

But the wealth of multiple interpretations of the 3-D space for Singh and Woods' vowel data, to say nothing of what might be found in a 4-D analysis, shows this conclusion to be premature if not false. Singh and Woods conclude that "the dimensions of the perceptual space of the vowel system seem to be essentially the same [as those found by Pols *et al*, Hanson, and Shepard - DT] -- tongue advancement and tongue height (p. 1866)." Naturally, it is of some interest to see that the configuration of vowels is similar in gross outline to the patterns found by those authors. But these experiments do not and cannot reveal what the dimensions are if "dimension" means one of the specific perceptual qualities according to which listeners classify and make discriminations between sounds.

Singh and Woods also leave many questions unanswered about the perception of American vowels. Perhaps the properties Tension, Duration, Rounding, and Diphthongization would find some reflection in the perceptual space, if a complete set of vowels including the diphthongs [ai au ɔɪ] were to be presented in a natural phonetic context.

The remaining study based on listeners' judgments of "distance" between vowel sounds is that by Mohr and Wang (1968). They asked listeners to perform paired comparisons of vowels using a rating scale. They found that degree of phonetic training (none, one introductory linguistics course, or a recently completed phonetics course) and size of rating scale (10 point or 6 point) had no demonstrable effect on their results, so the data from those conditions were pooled.

Their experiments differed in an important respect from the previous three: the stimuli were chosen as representative of certain phonetic distinctions that were of interest to the authors and were not sets of vowels native to the subjects. The two vowel studies had as stimuli [i a u ĩ ã ũ] and [i e æ y ø œ u o ɔ]. The first experiment investigated the attribute nasality as a potential perceptual quality, and the second used front rounded, front nonrounded, and back rounded vowels "to extricate the oppositions rounded-unrounded and front-back which are tied up in English (p. 32)."

No scaling analyses were performed on the data. Rather, five binary features were taken as a priori categories: high,

mid, labial, palatal, nasal. Similarity judgments were then averaged over sets of vowels having particular features in common. This procedure can be traced to Greenberg and Jenkins' (1952) modest but pioneering study of stop consonants. For example, the mean distance for all vowel pairs of which both members were [+nasal] was 3.43, while the mean distance considering only [-nasal] vowels was 4.61. This particular result was taken as support for the perceptual reality of the asymmetry of the feature nasality: +nasal involves the addition of a phonetic property, -nasal its absence; the extra quality nasality shrunk the perceived distances between nasal vowels, whereas the lack of nasality resulted in larger distances. The features High, Mid and Palatal (this feature presumably distinguishes [i y] from the rest of the set), on the other hand, show no such differences in mean distance ([+high], 3.42; [-high], 3.50; [+mid], 3.76; [-mid], 3.67; [+palatal], 3.07; [-palatal], 3.01). These figures support the status of tongue position features as equipollent rather than privitive (c.f. Trubetzkoy, 1939). Neither value of the features can be said to represent an added gesture. Mean distances with respect to the feature [±labial] show some slight drawing together of [-labial] vowels ([+labial], 3.10; [-labial], 2.78). This is the reverse of what was expected, since [+labial] involves the extra gesture, and the effect is explained away by considering the interaction of the features labial and palatal.

The immediately preceding discussion shows some interesting points which can be raised, if the description of the space is agreed on in advance. But comparing averaged distances over pairs with certain feature values provides no evidence that the features represent perceptual criteria, plausible as they sound. Actually, there is a risk -- in theory, at least -- of using a priori features only to discover that there is no spatial configuration which can be rotated to conform to the features at all.

This can be demonstrated by considering the 6-vowel set of nasal and non-nasal vowels. Based on Mohr and Wang's pooled data matrix, Table VI, two conclusions can be drawn. (1) [i u ɑ] are more distant from each other than are [ɪ ʊ ɔ̃]. Mean non-nasal vowel distance is 4.62; mean nasal distance is 3.42. (2) Mean distances for vowel pairs that differ only in nasality ([i ɪ] [u ʊ] [ɑ ɔ̃]) is 2.12. Attending only to these conclusions and the data that lead to them, one could conclude that there was a nasality dimension present. Figure 2.10 is a sketch of a possible space for which those two conclusions hold true, but for which no rotation is possible that separates nasal from non-nasal vowels. Mohr and Wang would doubtless assert, correctly, that this is not the correct perceptual space.

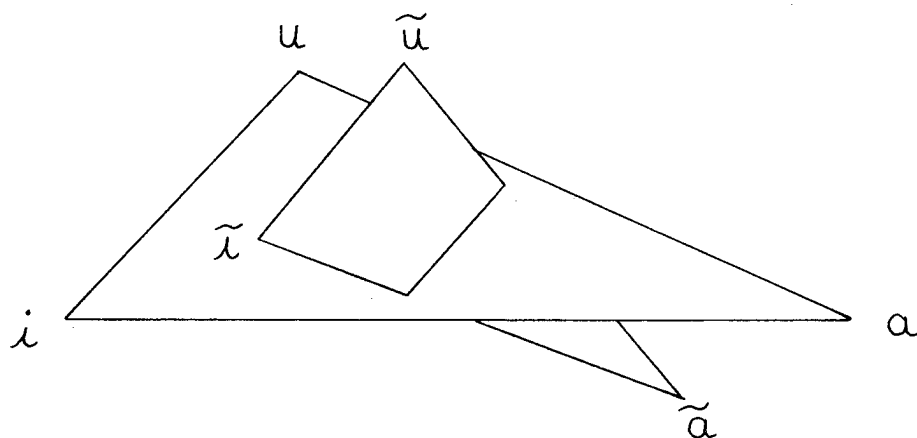


Figure 2.10: Sketch of a possible space in which the above two conclusions about Mohr and Wang's 6-vowel data hold true, but in which there is no nasality dimension.

But simply looking at various mean distance values cannot show them to be right. A more careful look at all inter-vowel distances is necessary to determine whether or not something like Figure 2.10 is a correct representation. MD-scaling affords that careful look.

At the time this chapter's research was done, non-metric analysis programs such as MD-SCAL (Kruskal, 1964) or TORSICA (Young and Torgerson, 1967) were not available to the author. Instead, Mohr and Wang's data were analyzed with COSCAL (Cooper, 1970), which assumes that the values in the response matrix represent metric, additive-scale distances.

The results in two dimensions for the 6-vowel experiment are plotted in Figure 2.11. The configuration is hardly unexpected. The rotation is arbitrary, and any scheme for classifying one vowel position as against the other two is no worse than any other. With these few vowels, it is just as easy to defend the position that there are no "dimensions" at all: $[i]$, $[u]$, $[a]$, may simply be regarded as qualitatively different from each other.

In three dimensions, the same plane recurs, but, in addition, the nasal and non-nasal vowels are separated, suggesting that nasality is a viable candidate for a perceptual dimension. A problem arises, however, in scaling a

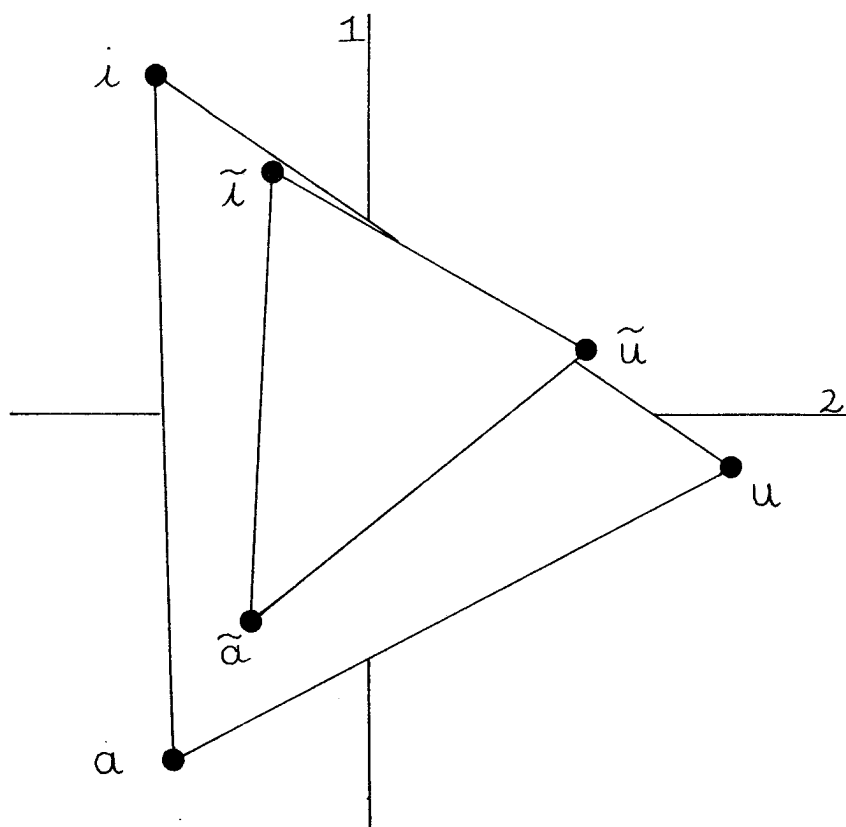


Figure 2.11: COSCAL analysis in 2 dimensions of Mohr and Wang's 6-vowel data.

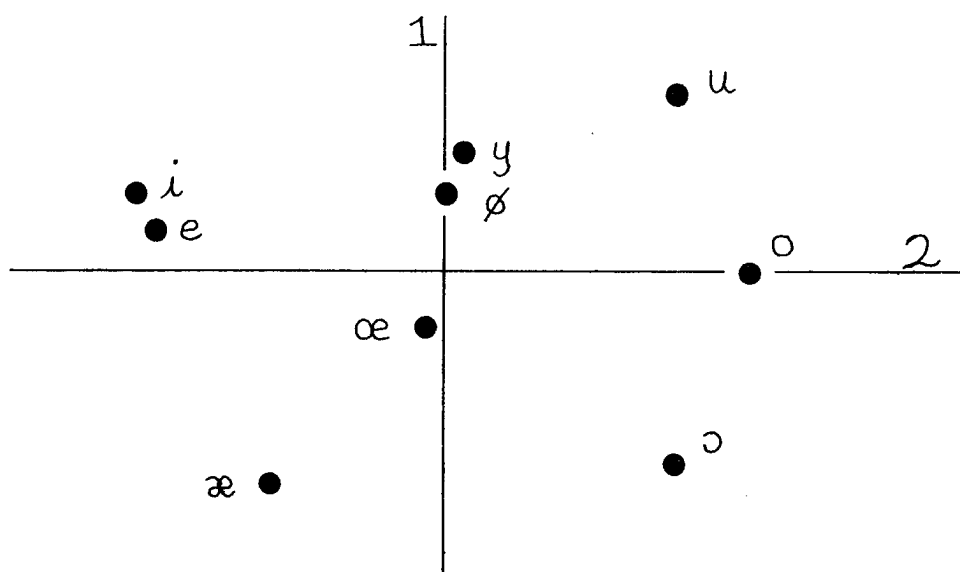


Figure 2.12: COSCAL analysis in 2 dimensions of Mohr and Wang's 9-vowel data.

6-point space in three dimensions. The input to the program consists of 15 independent cells in the distance matrix ($=6 \times 5/2$); the output consists of 18 co-ordinate values (6 vowels x 3 dimensions). There is no distillation of information, no reduction. One cannot in good conscience regard the 3-D space as an adequate solution. Besides, the increase in fit from 1 to 2 to 3 dimensions (.567, .973, .999 for the three spaces, respectively) is insignificant after two dimensions. But if the 3-D structure is examined, there is a rotation that corresponds to nasality.

Mohr and Wang's 9-vowel experiment was analyzed in the same way. The improvement in fit from 1 to 4 dimensions (.76, .93, .97, .98) suggests little improvement beyond three dimensions, and perhaps even three is too many. As it happens, the 2-D solution is almost exactly equivalent to one plane in the 3-D space, and the remaining dimension makes a binary distinction between [y ø œ] and the other six vowels.

The 2-D space appears in Figure 2.12. The vertical dimension forms a satisfactory scale of height, except for the fact that [i e] and [y ø] are too close to justify a high/mid feature cut. [œ] can be regarded as a lower (or "lax") mid vowel, in which case it is situated reasonably enough on the height scale. Alternately, [œ] is "low," and appears close to the other two front rounded vowels due to a tendency, perhaps, of the listeners to hear them as extremely similar. All front rounded vowels are non-native sounds for the subjects.

Horizontally, the vowels can be classed either according to front/back [i e æ y ø œ] / [o u ɔ], or according to round/nonround [o u ɔ y ø œ] / [i e æ]. A ternary acoustic feature of F1+F2 is also a possible hypothesis, although it cannot be tested directly, because Mohr and Wang did not publish any acoustic data for their stimuli. In the results for German (Section 6.1) it is found that a weighted sum of F1 and F2 correlates well with a perceptual dimension that distinguishes [u o] from [y ø œ] from front unrounded vowels.

Dimension 2 of Figure 2.12, then, is supported by the results of a more powerful technique, namely PARAFAC. But insight into the perceptual process is lost if it is automatically assumed that some known feature (such as rounding or backness) -- and a binary one, at that -- affords the best description of the dimension. From this point of view, the attempt to "extricate the oppositions rounded-unrounded and front-back" was not only unsuccessful, since either feature is applicable, but misleading in its very dependence on given features.

2.4 Three-mode MD-scaling

A profound advance in the theory and implementation of MD-scaling has come about within the last few years, with the independent yet simultaneous development of INDSCAL at Bell Laboratories (Carroll and Chang, 1970) and PARAFAC at UCLA (Harshman, 1970). Both programs are equivalent, insofar as their application to the present research is concerned. In using the name "PARAFAC" exclusively throughout the research, we are expressing nothing more or less than familiarity with the "local" product, and certainly intend no disrespect to the authors of INDSCAL. PARAFAC-INDSCAL is of crucial interest as it provides a solution to the problem of rotation of axes in traditional or non-metric scaling.

The input to PARAFAC³ consists not of one stimulus-by-stimulus matrix, but a number of them. In the research at hand, the third mode was persons; there was no summing across subjects to obtain one aggregate matrix. The output from this program includes a set of co-ordinates which assign the stimuli to points in a space. But it also includes a set of weights, or loadings, one for every dimension extracted, for every person. These weights express the relative salience of the dimensions for each listener. If person A has a weight of .6 on dimension 1 and .3 on dimension 2, then A's space is produced by multiplying all the co-ordinates of dimension 1 by .6, likewise the co-ordinates of dimension 2 by .3.

The crux of the PARAFAC model is that the axes of a solution are not rotatable; if the space is rotated, the fit of the model to the data becomes worse. Maintaining the separate identity of all data in the third mode (i.e., the subjects) locks the program into the best fitting orientation of the space, in addition to finding the general spatial representation of the stimuli. It is not correct to refer to "the" configuration of points here, because the program is actually deriving as many configurations of stimuli as there are listeners whose data was analyzed. PARAFAC assumes that each listener uses the same set of perceptual qualities, or dimensions, but that individuals may differ in the relative importance they attach to each dimension.

3. The application described here is a special case of the general properties of both PARAFAC and INDSCAL. Both programs perform factor analyses involving cases by measures by occasion (i.e., person). In MD-scaling, the cases-by-measures data are replaced by stimulus-by-stimulus matrices. INDSCAL is further generalizable to more than three modes.

A further advantage to PARAFAC is that it provides a new supplementary criterion for determining the true number of dimensions used by the listener. PARAFAC, like all scaling programs, is an iterative procedure, in which an initial set of co-ordinates is successively improved until the fit is maximum. PARAFAC is typically run using random numbers for starting co-ordinates. It has been proved (Jennrich, 1972, Harshman, 1972a) that for noise-free data, all random starting configurations in n dimensions will converge to the same solution as long as n does not exceed the number of dimensions actually used by the listeners. If listeners are using three dimensions, then all 1-D solutions will be identical, all 2-D solutions will be identical to each other, and likewise for 3-D spaces. But in four dimensions, uniqueness of solution will break down, and different sets of starting values will lead to different dimensions. As we shall see, this criterion is not infallible in connection with real, i.e. noisy data.

The PARAFAC approach expands considerably the range of possible hypotheses which can be investigated. It is possible to claim that only certain axes in the perceptual space represent perceptually real dimensions. It is therefore possible to investigate the acoustic correlates of the dimensions with more assurance than is possible with 2-mode (vowel-by-vowel only) scaling. Finally, individual subjects' reliance on the various dimensions can be compared. This last fact pertains directly to the possibility of studying language-specific vs. language-universal aspects of vowel perception.

The most extreme universalist hypothesis (H1) is that all languages, therefore all listeners, perceive vowels in terms of the same dimensions, or features, and that any variation from person to person is idiosyncratic. This is not a straw man hypothesis. The fact that different languages may use radically different sets of vowels does not falsify the universal feature hypothesis. It is certainly possible for speakers of a language with three vowel heights and speakers of a language with four vowel heights to use the dimension of height as a perceptual cue in identifying given vowels, even though the placement of vowels on that dimension will obviously differ. Continuing this example, H1 predicts that if a speaker of either language compares a pair of vowels from either language, one component of the similarity judgment will be Height; furthermore, speakers of both languages will attend to the same acoustic information with reference to Height.

A less extreme stand regarding universal features (H2) is that the features of the vowel space are universal, but that speakers of one language systematically attach more weight to some features than do speakers of another language. Rounding, for example, might be proposed as a feature more salient to German speakers than English ones, since German has rounding contrasts independent of backness and English does not.

A completely language-specific hypothesis (H3) is that speakers of different languages might have different ways of assessing the similarity of sounds, that is, language-specific features. If language A were shown to use height and backness as two main vowel dimensions, that result in no way implies other languages must therefore operate in a similar way. Perhaps speakers of language B assess vowels in terms of a palatal/pharyngeal dimension and a velar/non-velar dimension, similar to those dimensions proposed by Lindblom and Sundberg (1971). Other, more subtle differences could exist, as well. Language C might assess backness by reference to F2 alone; language D, however, may attend to F2-F1 for backness estimates.

PARAFAC appears to be an ideal tool for comparing the perceptual space for different languages. If speakers of different languages made similarity judgments for a common set of vowel stimuli, the entire matrix of vowel-by-vowel-by person data could be analyzed to choose between hypothesis H1 and H2. If there were consistent differences in the personal weighting factors across languages, then H2 is to be preferred over H1. If personal variation is not associated with language, but is idiosyncratic, then the absolute universal hypothesis (H1) is preferred. A blend of these two results is also possible. Idiosyncratic weightings may appear for some dimensions, and systematic differences by language may appear for others. In any case, of course, the program is constrained to find one set of dimensions all listeners share.

If some languages (or persons) have underlying perceptual spaces that are qualitatively different from others, the joint PARAFAC analysis will be a distortion of the truth. H3, the language-specific hypothesis, must be tested by analyzing language groups separately. If languages A and B used different sets of dimensions, then the separate PARAFAC analyses for the two languages will be different. An extreme result supporting H3 would be the case in which two analyses had no perceptual scales in common at all. Or, there may be a common core of dimensions, used by all speakers, supplemented by dimensions unique to different languages.

In addition to the universal- vs. language-specific nature of the results, it should be possible to focus on the dimensions themselves. With PARAFAC, the investigator is challenged to interpret the axes in the space directly, with no recourse to rotation.

One consequence of the fixed nature of the PARAFAC axes is that the relationship between perceptual dimensions and acoustic structure is investigated differently than with two-mode analyses. The appropriate question with two-mode analysis (e.g., TORSCA, etc.) is the following: Is there a direction in the perceptual space which corresponds to acoustic measure X? Answering such a question involves performing a multiple linear regression, with the acoustic data as the dependent variable and the co-ordinates for the vowels in the space as the independent variables. This is equivalent to rotating the space to discover the axis most similar to the acoustic variable.

But the appropriate question with PARAFAC is reversed: Is there some combination of acoustic variables which is capable of predicting the values on each PARAFAC dimension X? To answer this question, the acoustic variables must be treated as independent, and the PARAFAC dimension as dependent. To the extent that a particular simple acoustic measure, say the frequency of formant 1, is itself a perceptual dimension, the two approaches will give the same result. But with PARAFAC, to propose an acoustic parameter as a perceptual one as well is not a convenient, a priori decision, but a testable hypothesis.

The present research is organized exactly along the lines outlined above. Speakers of different native languages performed triadic comparisons on a set of vowel sounds. Details of the experimental method are in Chapter 3; an examination of the quality of the collected data and a preliminary overview of the results by means of hierarchical clustering (Johnson, 1967) make up Chapter 4. Chapter 5 presents the results of the PARAFAC analysis of all subjects, testing hypotheses H1 and H2. Analyses of each language separately appear in Chapter 6; these analyses contribute to the exploration of the language-specific hypothesis H3. In both Chapters 5 and 6, an attempt is made to relate the derived dimensions to acoustic data in addition to linguistic labels. Chapter 7 is a brief summary and concluding discussion.

2.5 Testing PARAFAC on real and synthetic data

Now that the broad outlines of the research have been sketched, it is necessary to conclude this chapter with three important methodological questions relative to the application

of PARAFAC to our data. These three questions can be discussed best in the context of four preliminary analyses, two on synthetic data, and two on data from other investigators.

1. What effect does noise in the data have on the interpretation of PARAFAC solutions? A set of data were constructed that had no spatial structure at all, in order to test PARAFAC with pure noise. The "data" consisted of a computer-generated simulation of 12 listeners who performed an entire triadic comparison experiment on 12 "vowels" randomly. Additive constants were fitted to each matrix as described in Section 4.4. After converting the new "absolute" distance matrices to scalar products, PARAFAC solutions in 1, 2, 3, 4 and 5 dimensions were obtained. Table 2.1 summarizes the results.

Table 2.1: Summary of a test of PARAFAC using random data. Each row represents a different solution.

<u>Number of dimensions</u>	<u>Fit</u>	<u>Number of negative person weights</u>	<u>Number of times that solution appeared</u>
1	.06	6	2
2	.11	11	1
2	.13	5	1
2	.14	9	1
3	.21	3	1
3	.21	3	1
4	.30	2	2
4	.30	3	1
5	.42	3	3

The first expectation was that Fit values would be low, as indeed they are. It was also expected that negative person weightings would appear. They do, but the negative person values become fewer for the 3, 4 and 5-11D spaces than for the 1-D and 2-D solutions. Finally, it was expected that each solution would be different. But the more dimensions extracted, the more likely the chance of a unique solution. Extrapolated to the case of real data, these last two results sound a potential warning: excessive noise may cause a solution of too many dimensions to appear formally acceptable -- more acceptable, perhaps, than a solution of fewer dimensions.

2. How important is it to apply Torgerson's systematic stretching of larger distances? Put another way, will the analysis be weakened if this step is not carried out? It was thought worthwhile to observe the effect of omitting that step with synthetic data. A two-dimensional space of 12 points was selected for this test. That space was a source for 12

derived spaces, each of them different from the original in relative stretching or compression of their co-ordinate axes. Some spaces were therefore thicker and shorter than the original, some taller and thinner; but the relative placement of points on each dimension was never changed.

A PDP-12 computer was used to simulate a complete set of triadic comparison judgments for each of the 12 distortions of the space. Dissimilarities matrices were computed for each set of judgments. These matrices represent error-free data from 12 hypothetical listeners who share perceptual dimensions but differ in the relative salience they attach to them.

Additive constants were fitted to each matrix, and scalar products were computed, as in the noise analysis above, before submitting the data to a 2-dimensional PARAFAC analysis. The positions in the resulting space (filled dots) are compared with the original points (open circles) in Figure 2.13. The spaces are obviously very similar; the axes have been superimposed directly without rotation. Most important in terms of the question at hand is that the three points towards the center of the plane are more widely spread out than their targets. This is equivalent to having distances between peripheral points contracted somewhat. Notice, though, that the interpretation of the results (whatever it is) hardly differs from the interpretation of the original space. If the effect were much larger, an extra dimension might appear, warping the 2-D space into a bowl-like surface. This could be counteracted by non-linearly stretching the dissimilarities matrices before analysis. But there is no way to determine the amount of such stretching, except by analyzing the unstretched dissimilarities data and measuring the amount of distortion imposed on the solution.

In this case, the 2-dimensional PARAFAC result is unquestionably the right one, since the space was known in advance. Theoretically, PARAFAC should demonstrate the correctness of the 2-D solution by deriving unique solutions in one and two dimensions only. This expectation is borne out with error-free synthetic data. But while the triadic judgments were error-free, the very quantization of responses into "more similar" and "more different" adds error to the distance matrix. Figure 2.14 demonstrates this problem. In judging which pair from the triad ABC is closer together, the answer is BC. But in judging triple ADE, the pair DE has the smallest distance. In computing the final difference matrix (c.f. Section 3.6), the "most similar"

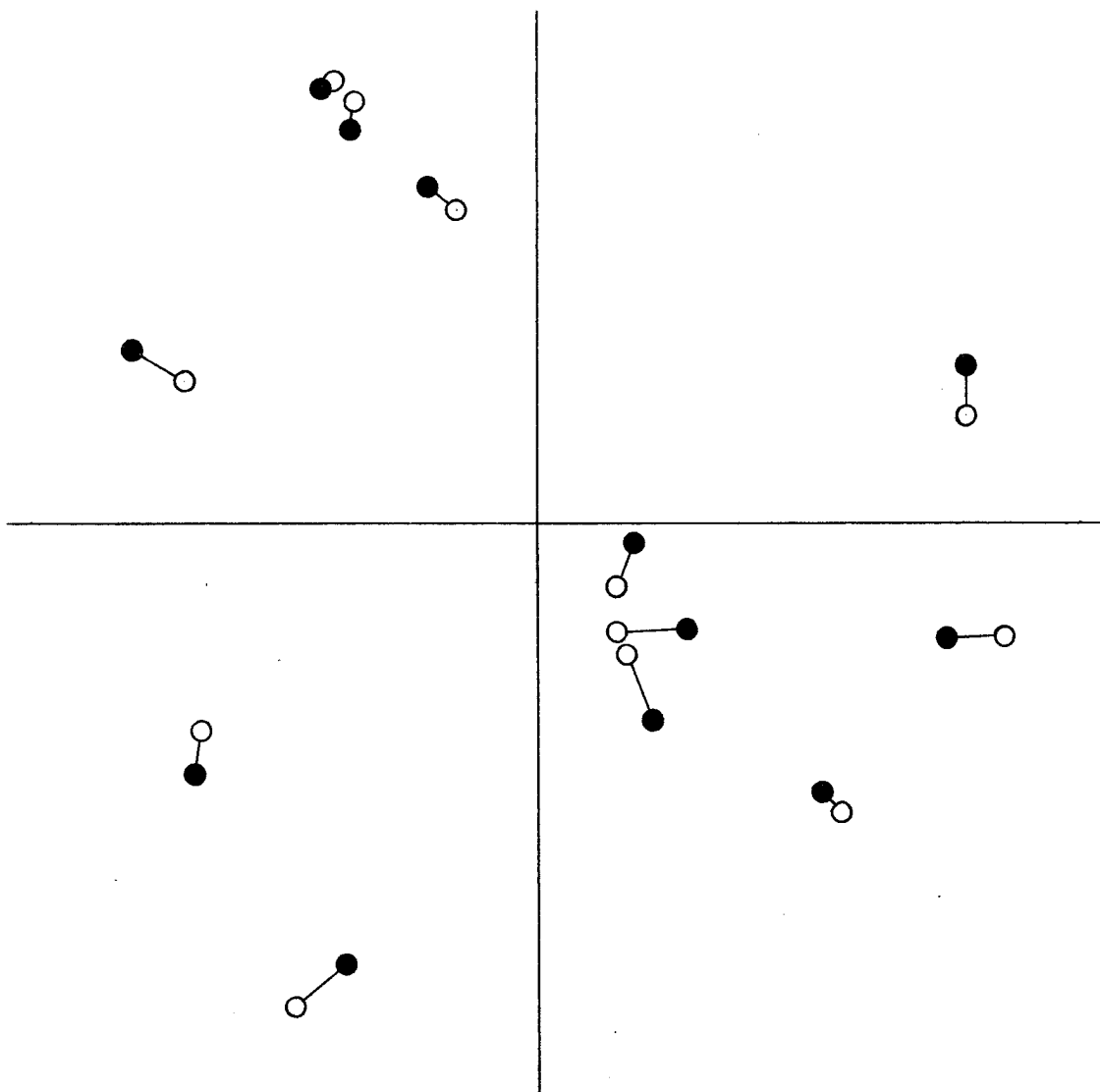


Figure 2.13: Comparison of target space (O) with PARAFAC result based on synthetic, noise-free triadic comparisons (●).

● A

● C

● B

● D

● E

Figure 2.14: Hypothetical space to illustrate that error-free triadic judgments do not result in error-free distance matrices.

judgments-BC from triad ABC and DE from triad BDE-will have the same weight. Furthermore, AE is correctly judged "most different" for triple ADE, although distance AD is very similar to distance AE. Yet the judgments for pairs AD and AE in the triad ADE have the same weight in computing the response matrix as do distances AB and AC, which are clearly more distinct from each other.

In order to see what effect the use of undistorted triadic judgments and their inherent noise has on PARAFAC's uniqueness-of-solution property, analyses were run in three and eventually four dimensions. The one and two-dimensional solutions were expectedly unique, but so were the three- and four-dimensional solutions. The third dimension of the 3-D space was uninterpretable, although the person weights on that dimension were roughly related to the size of the additive constants used for each "listener." Apparently the program was trying to account for some slight difference in the overall size of the subjects' matrices caused by adding different additive constants to each one. In this way, the dimension can be thought of as "real," although it is not accounting for any dimension in the target space.

The fourth dimension cannot be interpreted in any way. It simply appears to be true that there can be best-fitting solutions for the error present, especially, perhaps, when almost all the variance in the data is accounted for. (Fit in 1-4 dimensions is .64, .96, .97, .98.)

There is one further aspect of the 4-D solution, which requires its rejection: Four of the personal weighting factors for that solution, as output by PARAFAC, are negative. This represents a breakdown of the ability of the program to analyze the input as MD-scaling data. When the program is used to analyze factor analytic data, that is, objects by measures by persons, the personal weights are meaningful whether positive or negative. But in the MD-scaling application, the personal weighting factors, as they have been described earlier, are actually the square roots of the factors printed by PARAFAC. The algorithm is the same in either case, and there is no procedural constraint against negative values. So if negative values appear in the analysis of MD-scaling data, the solution is formally defective. Where very small negative weights may be charitably regarded as zero and therefore acceptable, large or numerous negative values reduce the credibility of the solution.

Given four different criteria, then, for determining the number of dimensions in the solution, there are four different answers. The fit curve clearly supports the two-dimension solution. The criterion of interpretability supports either the 2-D space or the 3-D space, though the latter possibility is marginal. The requirement that personal weighting factors be positive sets an upper bound of three dimensions. Finally, the uniqueness criterion is of no use, supporting as it does at least four dimensions.

Based on this study of synthetic data, it must be concluded that the use of undistorted triadic comparisons results in an interpretable solution, if one can pick the right solution to interpret. Choosing the solution may be hindered by conflicting guidelines.

All these considerations lead to the third question, i.e., (3) How difficult is it to interpret the results of scaling real data, with unknown noise, an unknown number of dimensions, and an unknown target space?

I am indebted to Dr. Louis C. W. Pols for his kindness in making available the individual listeners' dissimilarity matrices for the triadic comparison experiment reported in Pols et al (1969). The interpretability of this set of data ought to provide a reasonable set of expectations for the interpretation of our own data. The dissimilarities matrices from the 15 native Dutch listeners were converted to Euclidean distance matrices and analyzed by PARAFAC in 1, 2, 3, 4 and 5 dimensions.

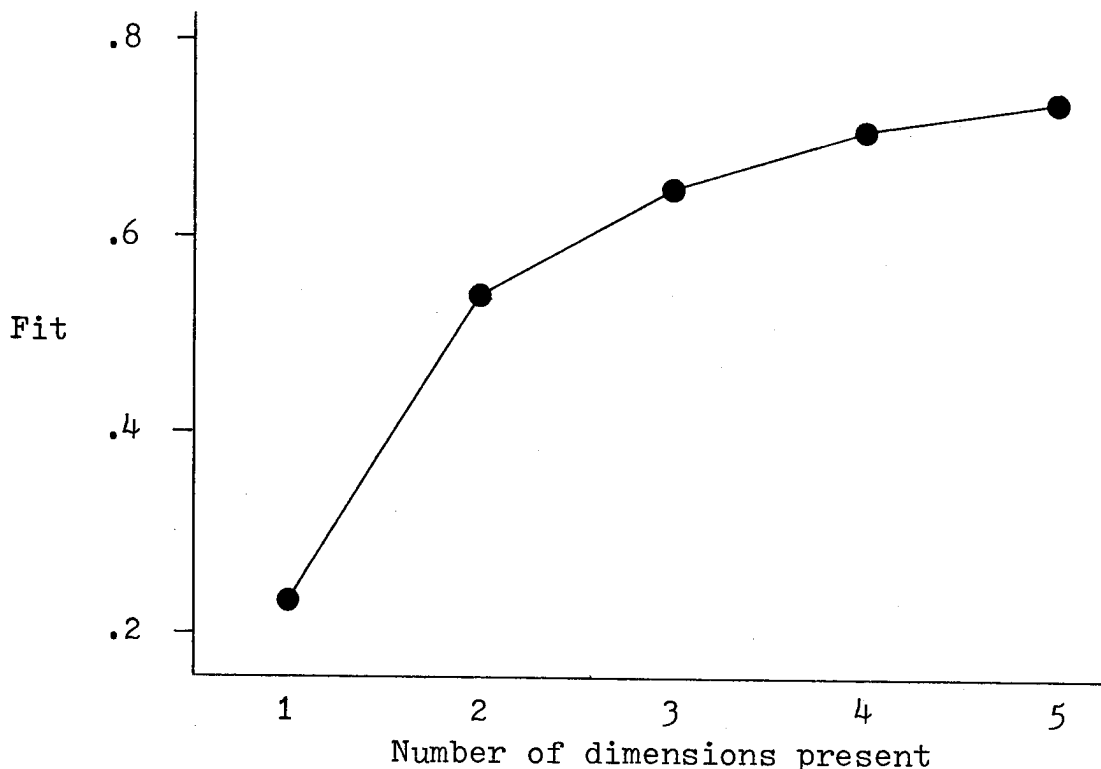


Figure 2.15: Fit vs. dimensions extracted for PARAFAC analyses of data from Pols et al on triadic comparisons of 11 Dutch vowels by native listeners.

The fit vs. dimensionality curve (Figure 2.15) reveals little. The not very distinct elbow at two dimensions suggests only weakly that the 2-D space is the proper one to consider. Another uninformative criterion is the existence of negative person weights. As it happens, there are one or two negative person values in the 3, 4, and 5 dimension solutions. It seems unreasonable to discredit these solutions merely because one dimension of each of them is not applicable to one or two out of fifteen listeners. Furthermore, the unacceptable values are small, relative to the other values, c.f. the 4-D solution which appears in Appendix II. The person weights in the appendix are normalized square roots of the PARAFAC values; the unacceptable weights are represented as coefficients of i , the square root of -1 . The weight of $.086i$ for person #7 is the smallest of all the person weights. The $.268i$ for listener #9 is more serious, but those two values are still by far the smallest in absolute magnitude for their particular dimension.

The remaining "objective" criterion, that of uniqueness of solution, suggests 4 as the correct number of dimensions. Only in five dimensions do there appear two different solutions when different random starting co-ordinates are used. As will be seen, there is a reasonable way to interpret the entire four-dimension space. The two 5-D solutions each consist of that 4-dimension subspace plus one uninterpretable

dimension. It appears that the uniqueness criterion is a good guide in this case, in spite of potential problems caused by local optima and spurious uniqueness due to noise.

Figure 2.16 represents one plane from the 4-D solution. Only one of the axes -- High/Low -- is easily interpretable according to traditional features. It is not a binary scale, but the high vowels [i u y] do lie at one end, with low [a ʌ] at the opposite end, leaving mid [ɛ ɔ o e ø œ] spread out between those extremes. The vertical dimension on that figure, Highlow/Mid, groups [ɔ o e ø œ] in a tight cluster, separate from [a ʌ ɛ u i y]. There is no apparent explanation for the grouping of [ɛ] with low vowels. To interpret this dimension directly is to claim that mid vs. extreme vowel height is a real perceptual feature, distinct from the feature high vs. low.

An alternate interpretation comes from considering the plane as a whole. The most striking feature about the plane is that it is not uniformly filled. Rather, the vowels lie in such a way that a smooth, curved line can be drawn that passes close to each of them. The curve drawn on Figure 2.16 was fitted by sight to illustrate this point. Instead of interpreting each dimension, it is proposed that the perpendicular projections of each vowel onto the curved line be regarded as a single perceptual scale of Height. Under this interpretation, the Height scale is ternary, separating high [u i y] from mid [ɔ o e ø œ] from low [a ʌ]. ([ɛ] remains an anomaly.)

Figure 2.17 is the more difficult of the two planes to interpret. The horizontal direction would be a reasonable backness scale, but for the placement of [a] and [ʌ] too far to the right and [y] too far to the left. It appears that the unrounded nature of [a] and [ʌ] and the rounding of [y] account for this shift, leaving back rounded [u ɔ ɔ] at the left of the scale, with front or unrounded [y ɛ ʌ œ a ø e i] spread out towards the right. The vertical dimension is a mystery. It truly looks like a random ordering of vowels. The label "e/ʌ" on the figure does no more than name the extreme ends of the scale.

As with Figure 2.16, the vowels tend to lie on a curved line (again, drawn by sight), and the interpretation is simplified by considering the projections of the vowels onto the curve as a single warped dimension. The immediate advantage of this approach is that the vertical e/ʌ dimension is effectively eliminated, since the order of vowels along the curve is very similar to the order along the horizontal scale. Only [a] and [i] are re-ordered. The curved scale, of course, still does not reflect any simple linguistic feature.

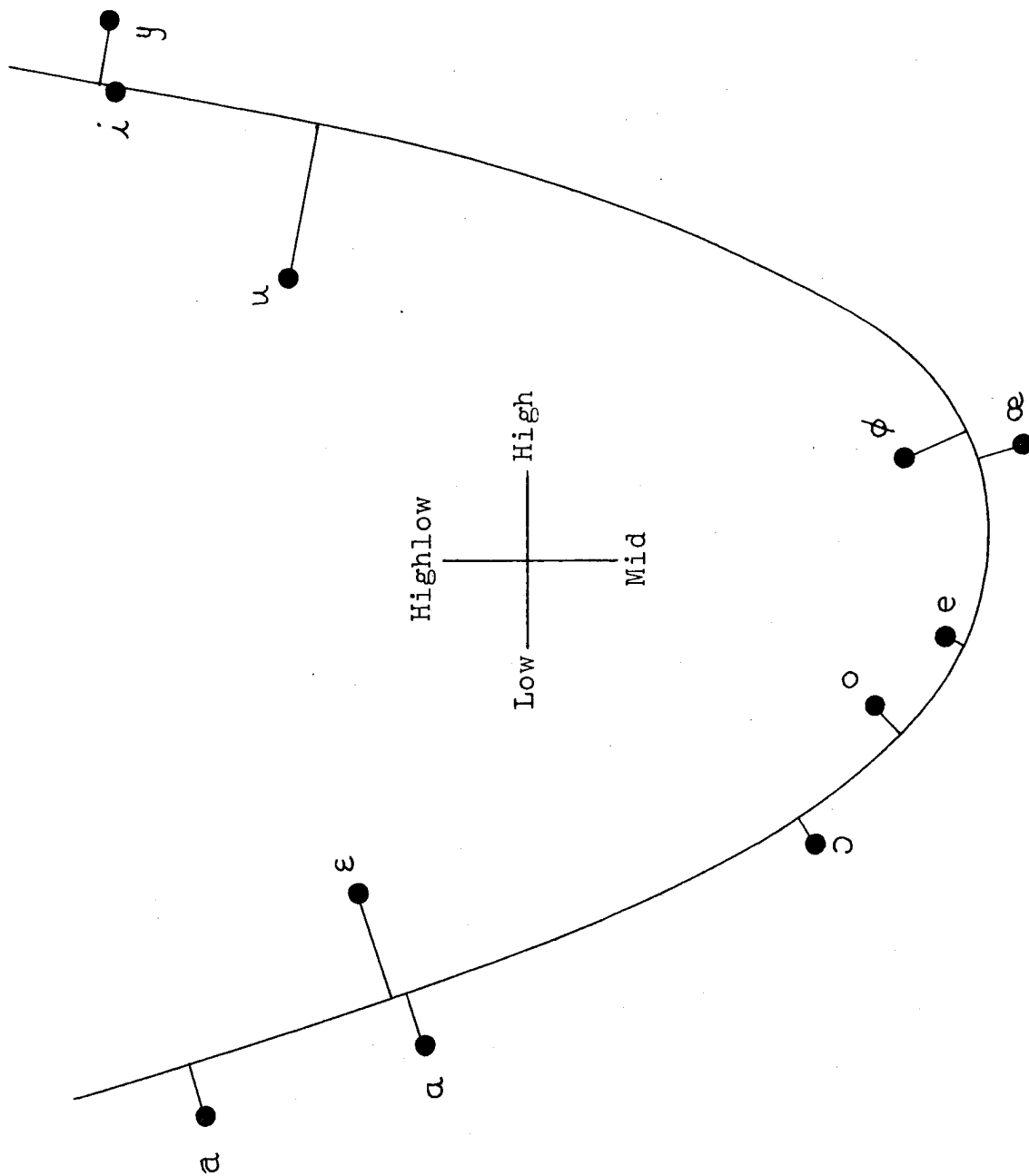


Figure 2.16: Two dimensions, High/Low and Highlow/Mid, of the 4-dimensional PARAFAC solution for Pols' et al data on Dutch vowels.

The interpretation of the two planes as two curved dimensions is strengthened by considering whether there exist vowels that would lie well away from the curves. What vowel, for example, could be predicted to lie half way between [a] and [i] on Figure 2.16? Those two sounds are described as low back and high front, respectively. A mid central vowel would be the prediction, but the [œ], which is acoustically very similar to a mid central vowel, is at the bend in the curve. Similarly, a vowel "between" [u] and [i] would certainly be [y], but on Figure 2.17 [y] is not off the curve -- in the center of the space -- but rather on the curve between [u] and [i]. (Another possible prediction is [ɤ] or [ʊ], that is, nonfront and unrounded rather than front rounded. It would be admittedly difficult to predict the location of those vowels.)

Another kind of support for the curved scales as perceptual dimension comes from stepwise multiple regression analyses, in which proposed dimensions are regarded as dependent variables, with acoustic measures as independent variables. The appropriate test is to compare how well the acoustic properties of the data predict the vowel positions along the curved scales, as opposed to how well those properties predict the original PARAFAC dimensions. If it could be shown that all four dimensions of the PARAFAC solution were weighted combinations of acoustic measures, one could conclude that there do not happen to be any traditional linguistic features that reflect those dimensions. The curved scales would be weakened in credibility. But if the regression analyses showed the curved scales to be better predicted by the acoustic data, the curves would be strengthened as possible "real" perceptual dimensions.

The acoustic variables used in the regression analyses (published in Pols *et al.*, 1969) were the frequencies of the first three formants, supplemented by the first three formant amplitudes if the frequency information was insufficient to provide a good prediction. Any linear combination of these variables which one might propose as the acoustic correlate of a particular dimension -- such as F2-F1 as a correlate of Backness -- is automatically included. Frequency and amplitude of the fourth formant, not regarded as potentially useful, were excluded.

The results are summarized in Figure 2.18. The vertical scale represents the value of the multiple correlation coefficient (= multiple r), and each vertical sequence of dots represents a series of regressions done for a different perceptual scale. The lowest point in each column gives the correlation value for the single acoustic measure that

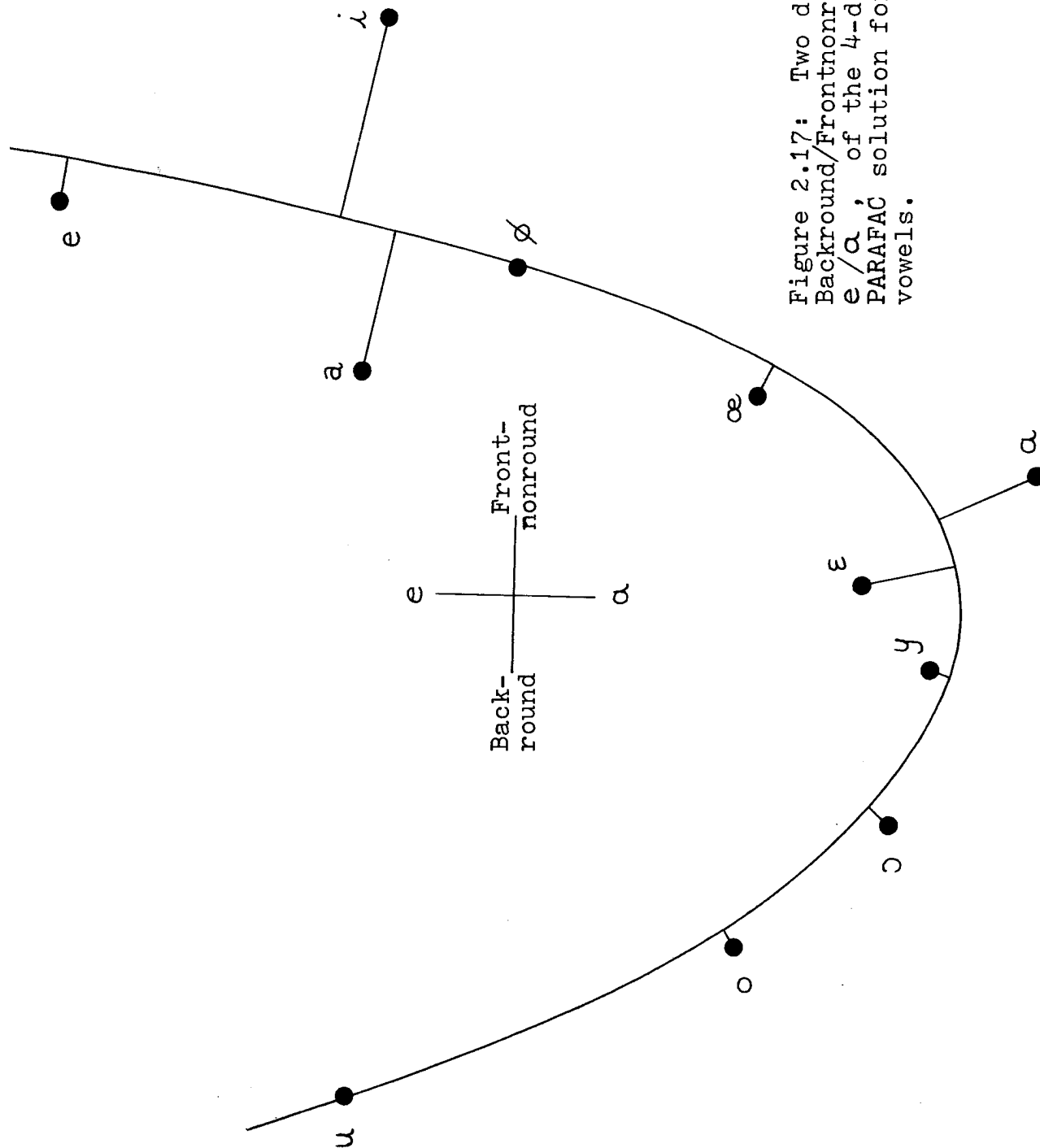


Figure 2.17: Two dimensions, Backround/Frontnonround and e/α, of the 4-dimension PARAFAC solution for Dutch vowels.

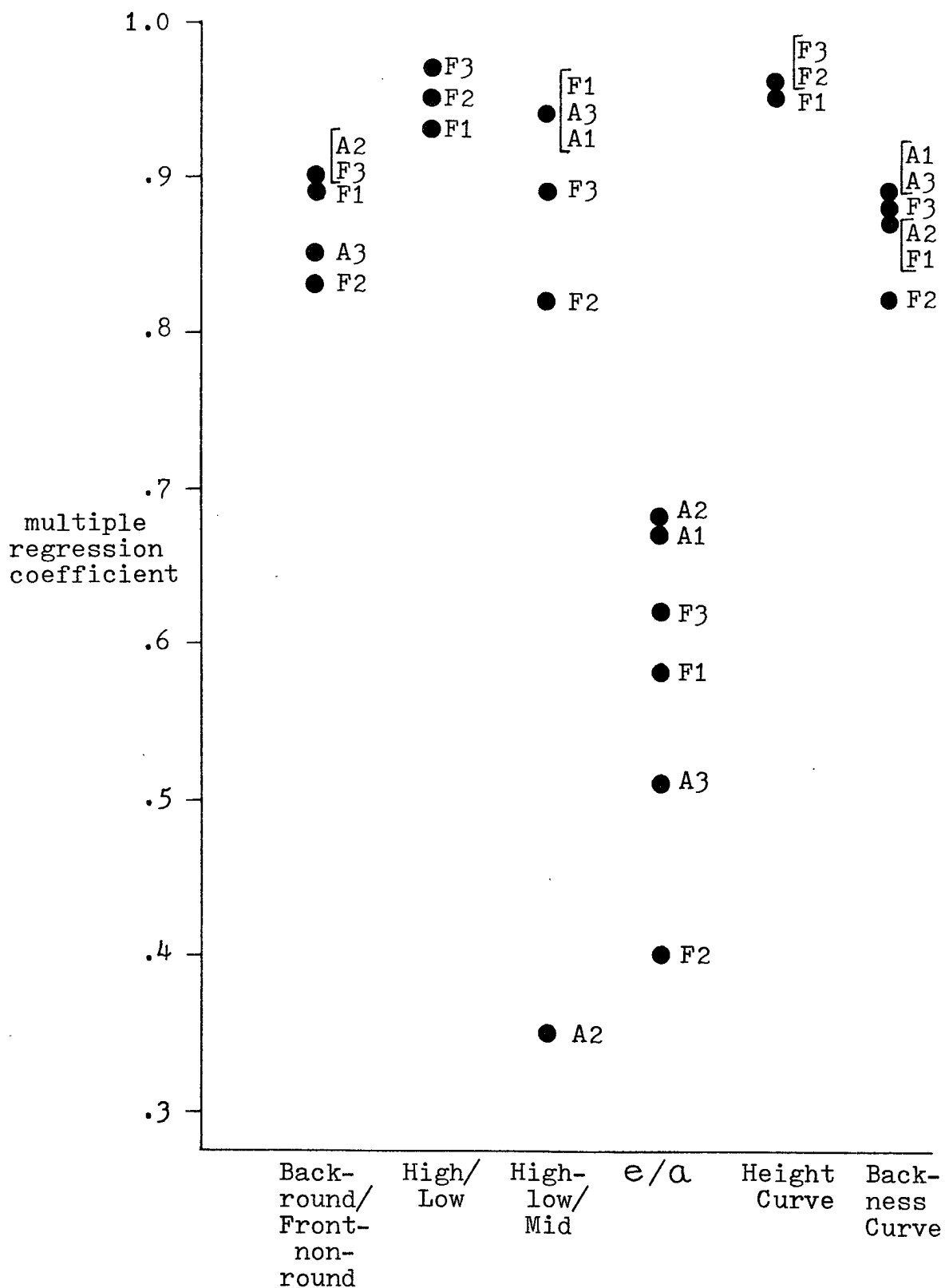


Figure 2.18: Stepwise linear regression analysis comparing acoustic measures to dimensions in the 4-dimension PARAFAC solution for Dutch.

best matches the perceptual scale. Each successive dot from the bottom upwards marks the multiple $\bar{r}$ which obtains when the named variable is added to the list of independent variables.

High/Low is the PARAFAC dimension that is best predicted by the acoustic variables; $\bar{r}=.93$ for High/Low and F1 alone. No useful improvement is made by adding other variables. Highlow/Mid is not predicted well by any single variable; the best is A2($\bar{r}=.35$). When F2 is included, $\bar{r}=.82$; with F3 added to those two variables, $\bar{r}=.89$. This appears to be a respectable correlation, but the dependence on A2 is not to be trusted. In the 2-variable regression, setting the regression coefficient of F2 arbitrarily at 1 for convenience, the coefficient for A2 is 127. The A2 values ranged from 20 dB to 33 dB, the values for F2 from 620 to 2100 Hz. To make the best prediction of the Highlow/Mid dimension, the amplitude values are inflated by a factor of 127 to a range of 2540 to 4190

before being combined with F2 values. A similar inflation occurs with the 3-variable prediction (for which the relevant expression is $F2 + 112A2 - .3F3$). The fact that A2 is weighted so heavily, together with the fact that neither A2 nor F2 (nor F3) are known to be related directly to height casts doubt on Highlow/Mid as a separate perceptual variable.

The alternative is to consider the Height curve as a single perceptual scale. In fact, the height curve is slightly better related to F1 ($\bar{r}=.95$) than is the PARAFAC height dimension.

As for the plane of Figure 2.17, Background/Frontnonround is based mainly on F2 ($\bar{r}=.83$); with all variables included, $\bar{r}$ reaches only .90. But e/α is not well predicted at all, even by all six variables at once; $\bar{r}$ in that case is only .68. If the e/α dimension is regarded as a perceptual scale, then there is no apparent clue in the acoustic space that shows what listeners attend to in placing vowels on the scale. Interpreting the Background/Frontnonround curve is again a plausible alternative. That curve shows a correlation with F2 ($\bar{r}=.82$) that is insignificantly smaller than the correlation of F2 with the PARAFAC Background dimension.

On Figure 2.19, the two curves are straightened and plotted against one another. This figure demonstrates a further nonlinearity in the solution. The curved dotted lines represent a superimposed grid of formant frequencies, drawn by sight (c.f. Figure 2.6, a linear F1-F2 plot). Formants 1 and 2 are clearly the source of the listeners' judgments, but the relationship between acoustic and perceptual

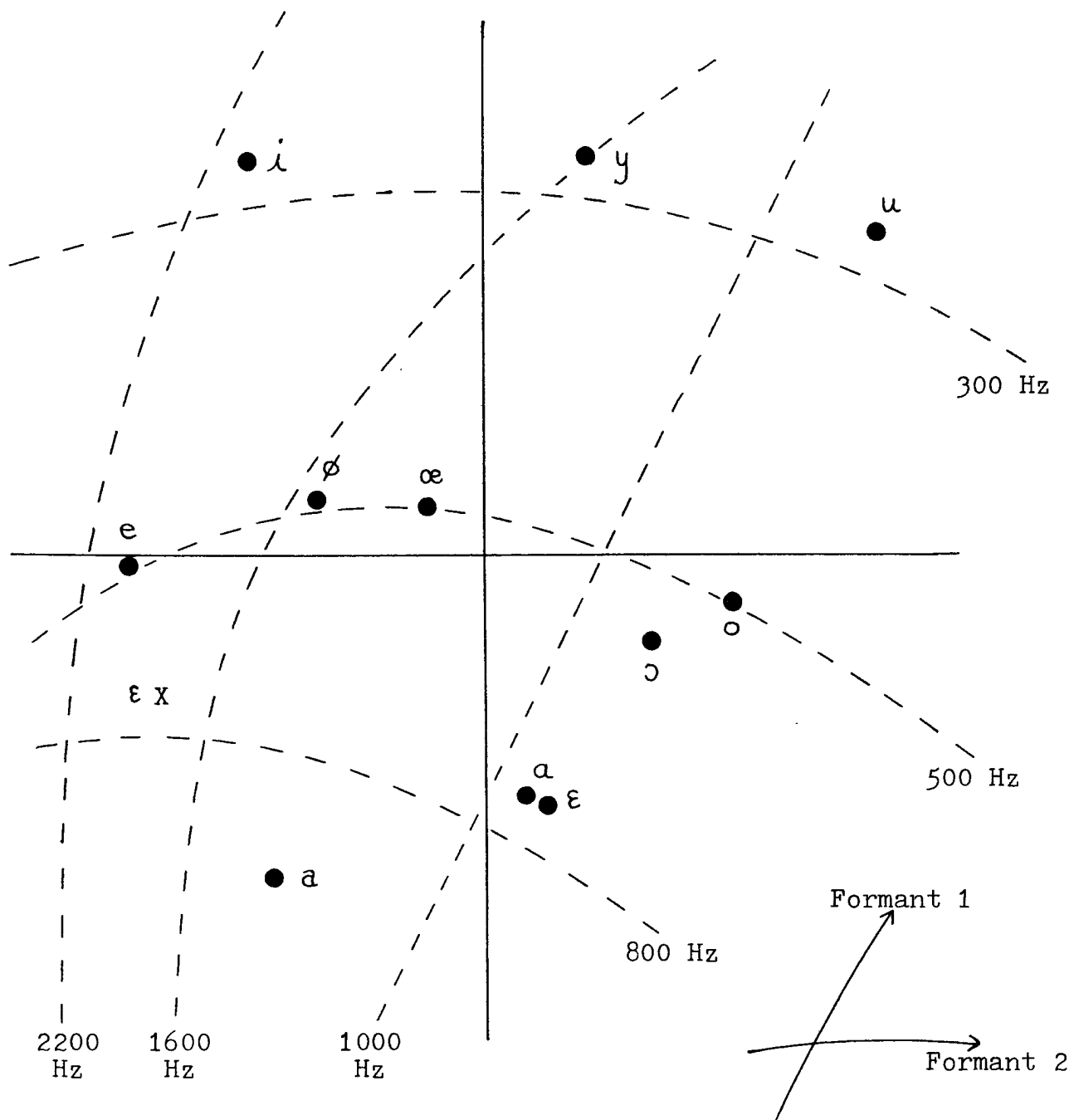


Figure 2.19: Straightened curves of Figures 2.16 and 2.17 plotted against one another. Superimposed is an appropriately deformed plot of formants 1 and 2.

distance is not constant over the vowel space. With F1 there is a gradual compression of perceptual distance as formant frequency increases. Perceptually, the difference between 300 and 500 Hz. is about the same as that between 500 and 800 Hz. As for the second formant, perceptual distance depends considerably upon the current value of F1. For a low F1, high F2 values give rise to relatively large perceptual distances. With a mid F1 (about 500 Hz.), perceived distance between vowels with high F2 are somewhat smaller. Little can be said about what happens at high F1 values, since the crucial vowel [ɛ] is perceptually way out of place (X marks its true acoustic position). No explanation for this anomaly is presented.

The 4-D solution has been described in some detail because that solution is "correct," according to the uniqueness-of-solution criterion. A final consideration is the possibility that the 1, 2, 3 or 5-dimension spaces might be equally easy to interpret, or perhaps easier. The existence of two different solutions in 5 dimensions argues against either as being correct, in addition to the fact that each solution has one uninterpretable dimension. (But see Section 6.3 for a problematic case of non-uniqueness in too few dimensions.) It remains to compare the 4-D solution with the three smaller ones.

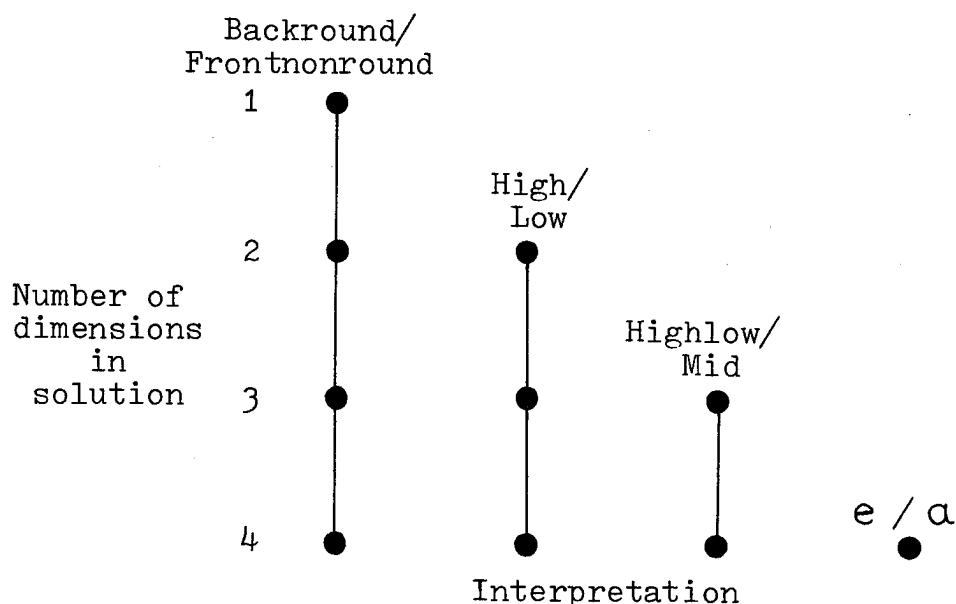


Figure 2.20: Evolution of the 4-dimension PARAFAC solution for Dutch.

Figure 2.20 summarizes the first four solutions; each solution is a subset of the next largest one. When only one dimension is requested, the result is the Backround/Frontnon-round scale, which appears in every larger solution as well. In a 2-D space, the second dimension is High/Low. It is not until the 3- and 4-dimension solutions that the Highlow/Mid and e/ α dimensions emerge, respectively. The 3-D space is therefore somewhat easier to interpret, since it does not contain the troublesome e/ α scale. The 2-D space, by the same token, excludes the Highlow/Mid scale, leaving only the two dimensions that are interpreted satisfyingly in the 4-D space.

In a sense, it might be argued that the curve of fit-by-dimensionality should have been accepted as a good guide in the first place, since the elbow in the fit curve was at two dimensions. But the elbow in the curve was not that sharp, and more importantly, this re-analysis was performed as a test of the uniqueness property. There is considerable satisfaction in being able to account, in some way, for all the unique solutions.

Another set of data, which will not be examined in detail, is found in Hanson (1967): the complete record of responses of 20 listeners to paired presentations of nine Swedish vowels (his experiment 22). Listeners heard each pair twice. In our reanalysis, the means for repeated pairs were computed; then each listener's matrix was fitted with an additive constant.

It was dismaying to find that the solutions above three dimensions were largely uninterpretable, even though unique solutions were obtained through five dimensions. Noise in the data could conceivably account for that, and it was in fact found that Hanson's individual data were exceptionally noisy. Since each listener heard each pair twice, a correlation coefficient (Pearson's r) based on the two lists of responses would show how reliably each listener was able to repeat his judgments. For eight listeners, r was .32 or less; which by a matched T test gives $P > .05$. The remaining twelve listeners had r 's of .43 to .80; variance accounted for ($=r^2$) therefore ranged from poor (.18) to fair (.64).

PARAFAC analysis was repeated with the eight worst subjects deleted. The data appeared to be tamed thereby to some extent; there were two different solutions at four dimensions, suggesting that the space is properly a 3-dimensional one. Unfortunately, only one of those three scales is easy to interpret -- a clear Backness dimension. The other two dimensions, when plotted against each other, form a plane which contains height and rounding information

but the two dimensions themselves are not interpretable. No attempt at finding curved dimensions was successful. Since the data from even the best listeners were rather inconsistent, it did not seem worthwhile to delete more subjects in hopes of improving the results. Co-ordinates for the 3-D space appear in Appendix II.

Two further attempts at handling these data were unsuccessful. One was to obtain a solution in five dimension, on the chance that the 4-D solutions were equally bad projections of a larger space. Four different sets of starting co-ordinates resulted in the same solution, which consisted of the 3-D space plus two other uninterpretable dimensions. If some larger solution might be the truly appropriate one (5 dimensions? 6?), the 3-D solution will probably re-occur as a subspace and cannot be explained away.

The other approach was to analyze the data by PARAFAC2 (Harshman, 1972b) an extension of the PARAFAC model which allows the co-ordinate axes to become oblique if the fit is thereby improved. (PARAFAC2 is discussed somewhat more thoroughly in Section 6.5.) The analyses in two and three dimensions (more would have been prohibitively expensive) offered no improvement over the PARAFAC solutions.

Bricker (1972), reanalyzing Hanson's data in an early demonstration of INDSCAL, refers to Hanson's data as "limited," a very conservative way indeed to describe data so laden with noise. Bricker's study did not present any detailed justification for choosing the 3-D solution, though his results were similar to those given above: a backness dimension, and two others not so obvious.

The reason Bricker's analysis is not identical to the present one is that the additive constants were estimated differently; the data submitted to INDSCAL and PARAFAC were therefore not identical.

The two PARAFAC analyses, based on data from Pols et al and Hanson, differ considerably in their ease of interpretation. The Dutch data were much more easily dealt with, even though it was necessary to propose that curved lines drawn through a space be seen as perceptual dimensions. Hanson's Swedish data, on the other hand, would not yield to this or other attempts at interpretation. The behavior of these two data sets is taken to support two decisions made earlier in this chapter: (1) to use triadic comparisons (as did Pols et al) rather than paired comparisons (as did Hanson), and (2) to perform no new non-linear stretching of large distances in triadic-based dissimilarities matrices. Any underestimation of large distances ought to be captured by a curved dimension based on the original PARAFAC dimensions.

Chapter 3

The Experimental Design

3.0 This chapter follows the design of the experiment from the selection of languages, stimuli, and listeners, through the experimental design and the subjects' completion of the task.

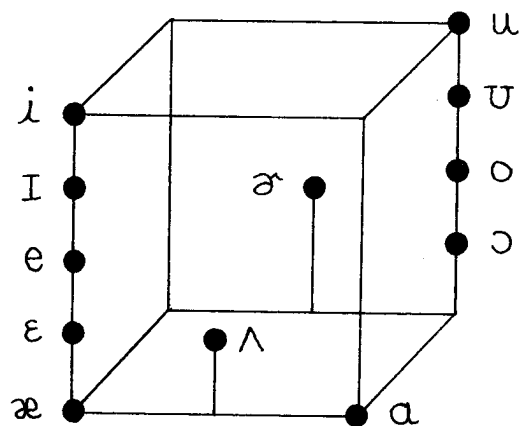
3.1 Languages used in the experiment

The selection of languages and of vowel stimuli were understandably dependent on one another. Languages and stimuli were selected so that each listener heard some sounds that were more or less familiar to him, and some sounds that were not. The languages chosen for the experiment were English, German, Thai, Turkish, and Swedish. Speakers of these languages were readily available, and, more importantly, the languages represent a wide range of variation in vowel sounds and vowel patterns within the framework discussed below. A brief sketch of the vowel systems of the test languages is presented here.

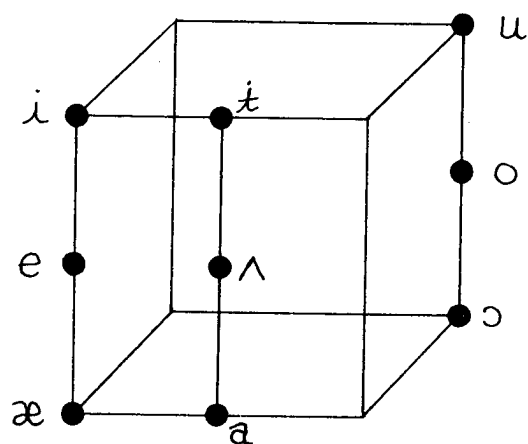
A simple way to describe the vowel patterns is by means of phonetic symbols arranged spatially, according to the traditional parameters of height, backness, and rounding, as in Figure 3.1 a-e. These symbols are intended to represent surface phonemes, fairly monophthongal, and are not intended to be phonetically precise. English (Figure 3.1-a) can be thought of as having twelve such vowel phonemes, although [e o u] generally terminate with high glides. This number reduces to eleven for speakers who make no distinction between [ɑ] and [ɔ]. [ɛ] and [æ], although often listed as English phonemes, following the tradition of Trager and Smith (1957), are more convincingly seen as allophones of [ɪ] or [ʊ] and [æ] or [ɑ], respectively. But these details do not affect the main fact that English requires only three linguistic dimensions to describe its vowel system: height and backness, primarily, and retroflexion to distinguish [ɻ] from all other vowels. Tenseness and rounding are predictable, given the height and backness specifications. If there is objection to distinguishing [ɪ] and [ɪ] or [u] and [ʊ] or even [e] and [ɛ] by means of the category of height, then tenseness can be used to make these distinctions.

Thai (Figure 3.1-b) uses a 3-by-3 symmetrical vowel system, with duration as an independent feature applicable to all vowels. If backness is thought of as a three-way feature, then rounding is redundant. Alternatively, if rounding is regarded as distinctive, then backness can be

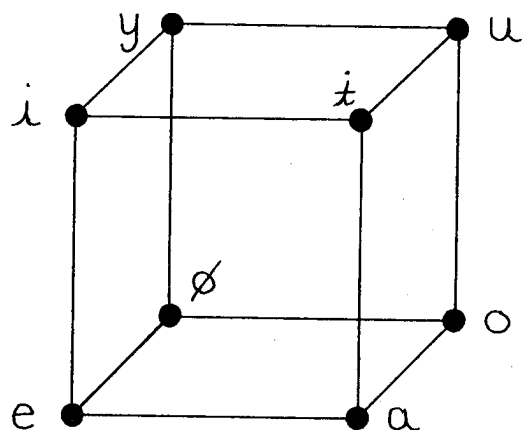
a. English



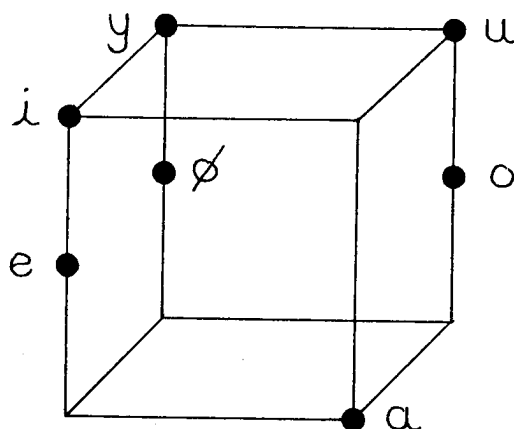
b. Thai



c. Turkish



d. German



e. Swedish

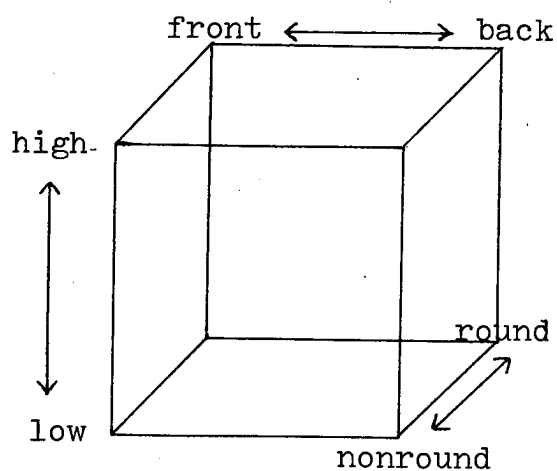
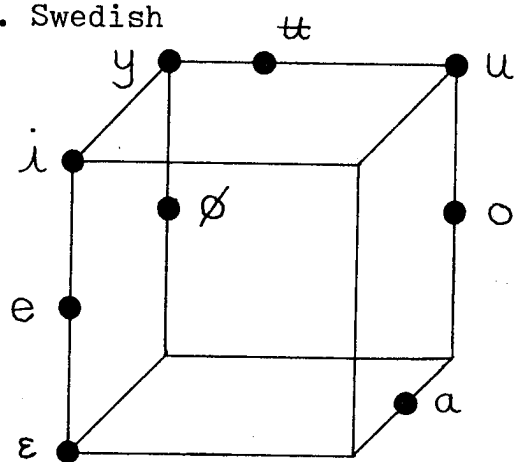


Figure 3.1: The vowels of the five test languages arranged according to the parameters of height, backness, and rounding. 59

represented at some level as binary, the distinction between back and central being redundant. The vowel system of Turkish likewise has a symmetrical system (Figure 3.1-c), but here all three features -- height, backness, and rounding -- are distinctive. The same three features are distinctive in German (Figure 3.1-d), although the system is not as neat as the Turkish one. German vowels may further be [$\pm$ tense], which affects duration and height; but tenseness is predictable from the syllable structure.

Finally, Swedish has a system of nine vowels: three non-rounded, three front rounded, and three back rounded (Figure 3.1-e). Again, vowels may be (predictably) [$\pm$ tense]. What makes Swedish somewhat unusual is the presence of the [ɥ], a second high nonback rounded vowel in addition to [y].

The simplified descriptions of the five test languages are not really adequate for comparisons between languages. To provide an overview of how the vowel symbols are being used in each language relative to the others, it is useful to consider plots of Formant 1 against Formant 2 for the vowels as they are produced by speakers of the languages (Figure 3.2 a-e). It is of course the patterns formed on the F1-F2 plane, rather than the absolute formant frequencies themselves that are important. Needless to say, there are other parameters along which vowels may differ, but these two have much to do with the linguistic scales of height and backness and are therefore extremely convenient measures. The vertical axes for Figure 3.2 a-e represent the frequencies of the first formant, increasing downwards; the horizontal axes represent the second formant, increasing towards the left. This orientation of axes places high vowels at the top of the plane and front vowels to the left.

The English data (Figure 3.2-a) are means for 76 naive male speakers as reported by Peterson and Barney (1952). The German data (Figure 3.2-d) are taken from Jorgensen (1969); the points plotted are means for 8 speakers, some linguistically naive, some sophisticated. The basis for Figure 3.2-e was a scatter diagram of F1 against F2 for 14 Swedish speakers, found in Fant (1959, reprinted in Hanson, 1967, p. 28). The points on the figure are approximate means for Fant's speakers, estimated by sight.

No comparable set of average formant values could be found for Thai vowels. Abramson's monograph (1962) discusses in detail the formant structure for only two speakers. Mean frequencies for the long vowels, spoken in isolation, for these two listeners are given in Figure 3.2-b.

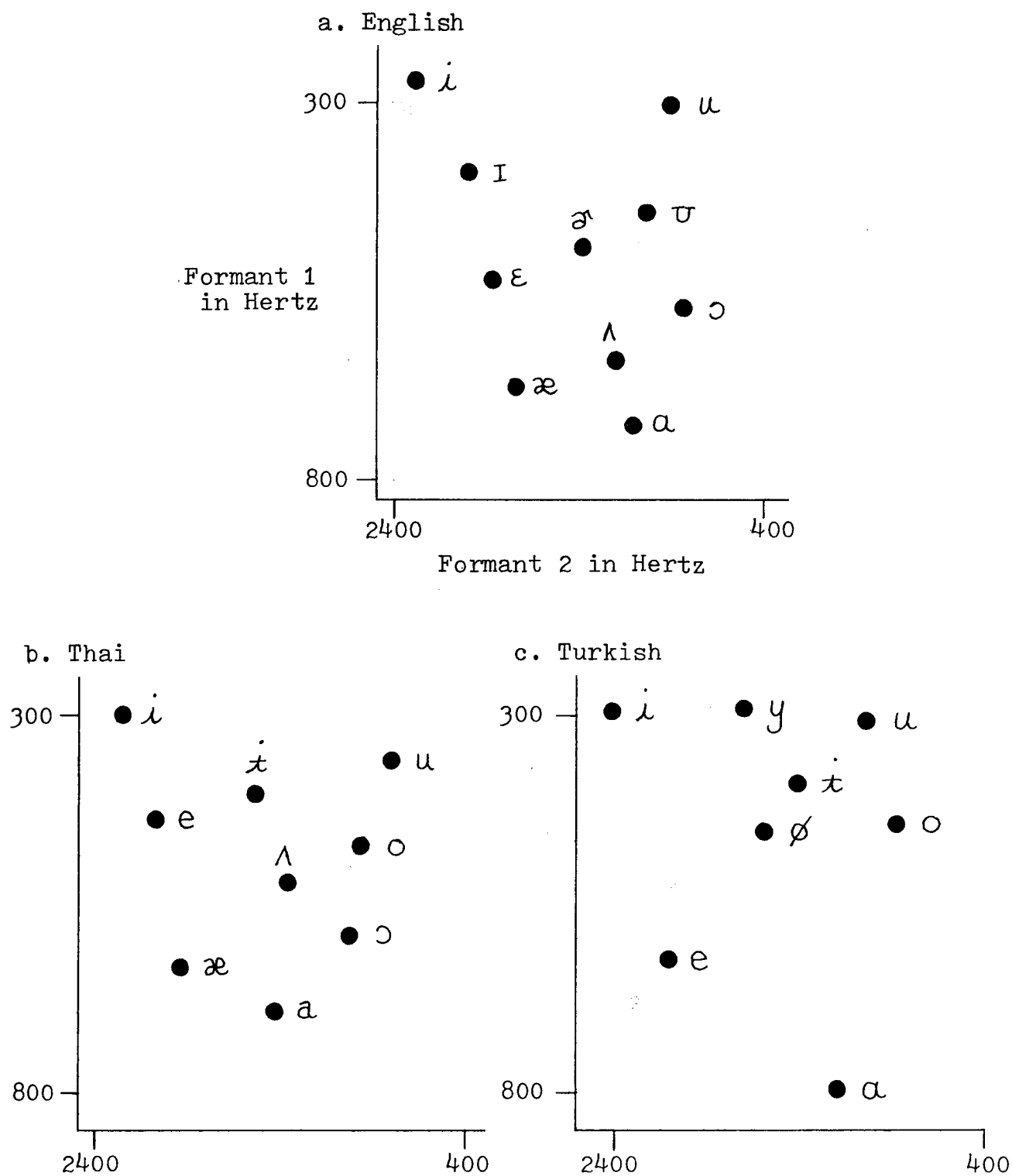


Figure 3.2: Vowels from the 5 test languages plotted according to formants 1 and 2.

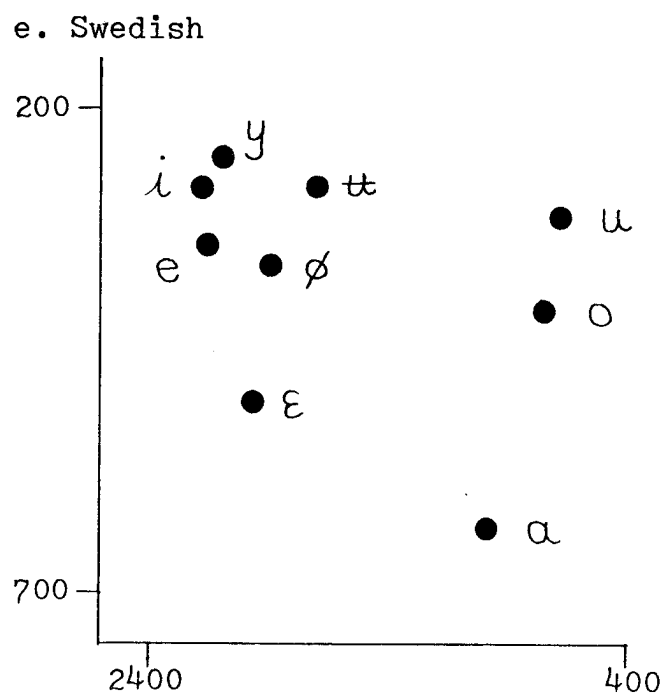
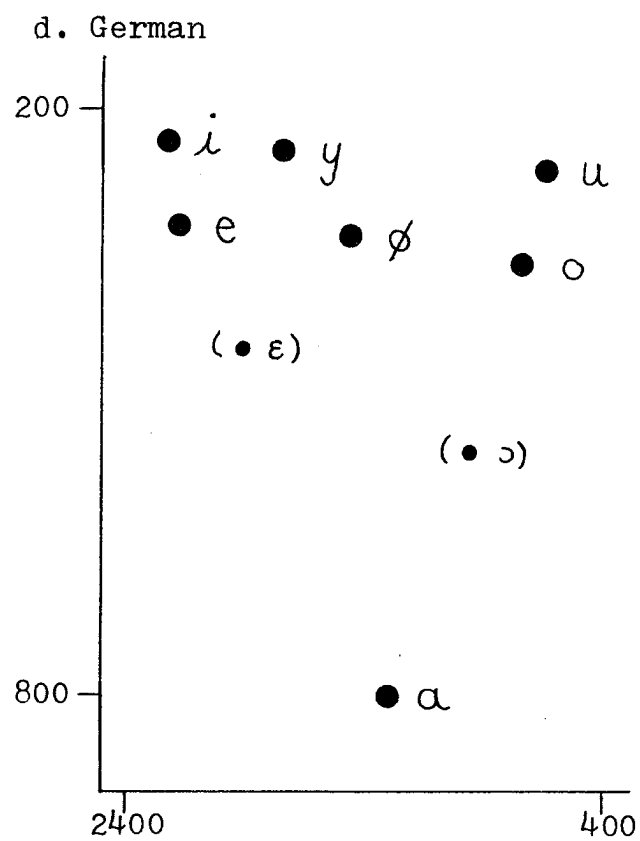


Figure 3.2 continued.

For Turkish, no published vowel formants could be found at all. Spectrograms were made of two native Turkish speakers pronouncing the eight vowel letters of the Turkish alphabet, which are homophonic with the eight Turkish vowel phonemes. (Each speaker pronounced four tokens of the vowels.) Figure 3.2-c plots the means for these two speakers.

In comparing Figure 3.1 with Figure 3.2, that is, the linguistic descriptions with the acoustic descriptions, there are two points to be made. First, the meaning of any one vowel symbol is not constant from one language to the next. By far the more interesting observation is that some vowels within one language appear extremely close to one another on a 2-formant diagram. A very clear example is the Swedish pair [i y] ; another is the Turkish pair [ɰ ø]. (This pair is much closer for one Turkish speaker than for the other.) The closeness suggests the possibility that there may be other parameters that are important for some distinctions, besides the first and second formant. This, of course, is precisely the sort of problem MD-scaling is intended to help answer.

3.2 Selection of the stimulus vowels

The difficulty in selecting the stimuli was one of finding a small set of vowels that could be handled within the scope of the experiment, and that still represent a wide range of vowel qualities. Too many vowels would fatigue the listeners, making their responses potentially less reliable; too few vowels would weaken any resulting claims about linguistic universals.

The number of stimuli was set at 12, requiring each listener to respond to 220 triads. Since two judgments are made per triad -- one "most similar" and one "most different" response -- 12 stimuli seemed a reasonable upper limit on the listeners' patience.

Given the five languages chosen, it was necessary to select the stimulus vowels so as to take advantage of the differences between them. This meant, first of all, ignoring parameters which do not serve to distinguish vowels from one another in the chosen languages, namely, Glottal source spectrum (creaky-to-breathy voice quality) and Nasality. Other features are excluded because they deserve investigation in their own right and would blur the focus of the experiment at hand. Diphthongization, Tone features, and perhaps Duration belong in this group. Tenseness was not used as a factor in selecting the stimuli because previous investigations (Pols et al, 1969, Singh and Woods, 1970) failed to find a perceptual dimension of Tenseness.

Considering the universal features proposed in Ladefoged (1971), four ways of classifying vowels remain. All of them -- height, backness, rounding, and retroflexion -- reflect mainly formant frequency variation. The vowels were chosen so as to vary widely over these four features. Retroflexion, first of all, was included since it is a distinctive feature for the vowel [ɘ] in American English, at least at the surface phonemic level, and it was thought to be of interest to see how speakers of other languages treated this vowel. To make identification of a retroflex dimension more positive, if one should emerge, another retroflex vowel was included, a low back vowel symbolized as [ɑ̣]. As for the other three features, if rounding and backing are regarded as binary, and height is regarded as ternary, there are twelve possible vowels to choose from:

	unrounded		rounded	
	front	back	front	back
high	i	ɨ	y	u
mid	e	ʌ	ø	o
low	æ	ɑ	œ	ɔ

To make room for [ɘ] and [ɑ̣], [ɔ] and [œ] were eliminated. Without them, there is still an adequate assortment of front rounded and back unrounded vowels, and the category of height is well-represented. Although this set covers a broad range of qualities, it leaves out the interesting [ɥ] vowel, which causes difficulty in the phonetic description of Swedish. After some preliminary analyses of the English, Thai and German data were carried out, it was found that for non-English speakers, the [ɑ̣] vowel was effectively identical to [ɑ], perceptually. The [ʌ], too, was very similar to [ɑ]. These two vowels were therefore deleted from the stimulus set for the Swedish listeners, and [ɥ] and [œ], both native to Swedish, were put in their places.

A plausible classification of the stimulus vowels is given in Table 3.1. These feature values cannot give precise phonetic descriptions of the sounds, but are generally adequate to indicate the relationships that exist between sounds. [ʌ] is a mid vowel only relative to [ɑ]; [œ] is a low vowel only relative to [ø]. The [ɨ] vowel used in this experiment is not a back vowel, but it is certainly non-front; calling it [+back] in this framework is the only way to contrast it with the other high unrounded vowel [i]. Notice also that [y] and [ɥ] have the same feature specification. An additional feature or the use of three values along the backing dimension would be necessary to distinguish these two vowels.

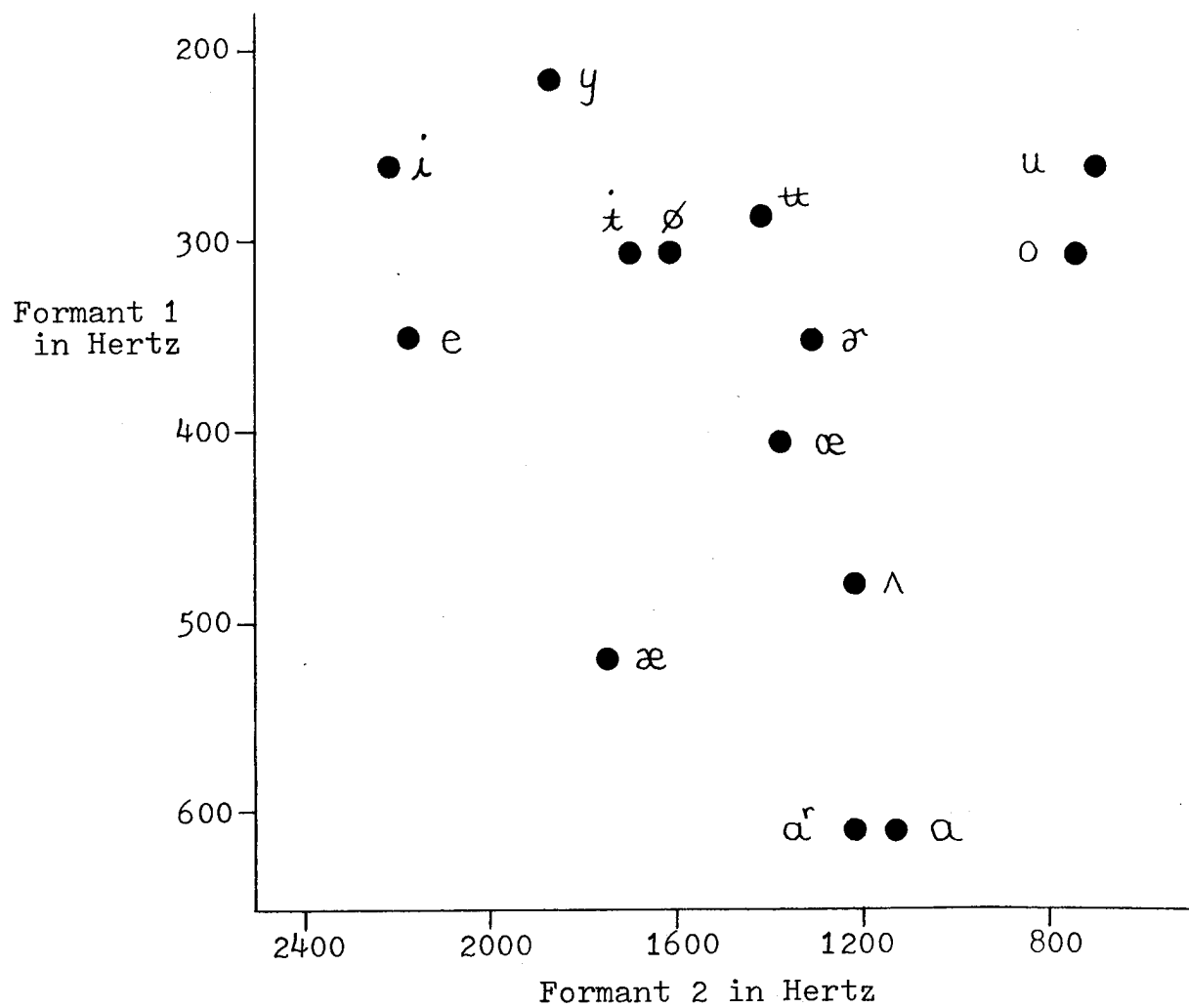


Figure 3.3: Stimulus vowels plotted according to their first and second formants.

Table 3.1: Distinctive feature chart for the stimulus vowels.

	Vowels used with all listeners										Swedish only	Others only		
	i	e	æ	ɑ	o	u	ɤ	y	ø	ɨ	ʉ	œ	ɑʳ	ʌ
High	1	2	3	3	2	1	2	1	2	1	1	3	3	2
Back	-	-	-	+	+	+	-	-	-	+	-	-	+	+
Round	-	-	-	-	+	+	-	+	+	-	+	+	-	-
Retroflex							+						+	

It is instructive to examine the plot of formants 1 and 2 for the stimulus vowels (Figure 3.3) in order to compare them more readily with the vowels in the listeners' native languages. (All vowels were spoken by the author; c.f. Section 3.3.) As was mentioned when the F1-F2 plots for the test languages were presented (Figure 3.2), only patterns can be compared, not absolute frequencies. The vowels labelled [i e æ ɑ ɑʳ u] form a rough triangle, as expected, and need no further comment. Although [ɤ] appears higher than the plot of Peterson and Barney's means would suggest, the [ɤ] stimulus was in fact the author's normal American vowel. [o] has an unusually low first formant, placing it very close to [u]. The vowel [ʌ] is more like an English [ʌ] phoneme than the Thai [ʌ] phoneme. [ø] and [œ] are higher than the vowels with those labels in the Swedish, German and Turkish graphs.

The placement of [y] and [ɛ] deserves perhaps more comment. The [y] stimulus lies about 1/4 the distance from [i] to [u], roughly comparable to the pattern for German and Turkish. But in Swedish, the vowel [ɛ] is in this area and forces [y] to move closer to [i]. To maintain the proper relationship between the stimuli [y] and [ɛ], [ɛ] was given a lower second formant than is usual for Swedish.

The final vowel that needs mention is [ɨ]. This stimulus appears much farther "front" than is typical for either of the two languages that use a [ɨ], Turkish or Thai. The extent to which Turkish and Thai listeners actually heard the [ɨ] stimulus as a native vowel sound is brought out later, in Chapters 4 and 6.

It must be kept in mind that a plot of formant 1 against formant 2 is not necessarily representative of perceptual distances, and the patterns of vowels on such graphs do not

indicate with any surety whether or not a vowel will be identified in a certain way. Yet the first two formants do provide a quick way to refer to vowels and give us some context in which to interpret Table 3.2, a list of the approximate native vs. non-native status of each of the stimulus vowels relative to the test vowels. Entries for [ɤ] appear in parenthesis because [ɤ] may not have been close enough to native standards, as mentioned directly above. [æ] entries appear for German, Turkish and Swedish, because there are allophones in those languages that are very similar to our [æ] stimulus, even though they have no separate [æ] phoneme.

Table 3.2: Approximate native/non-native status of the stimulus vowels relative to the test languages

	Vowels used with all listeners										Swedish only	Others only
	ɤ	y	ø	ɘ	æ	u	o	a	e	i	œ ɘ	ɑ ʌ
English					x	x	x	x	x	x		x
German		x	x	(x)	x	x	x	x	x	x		
Thai	(x)				x	x	x	x	x	x		x
Turkish	(x)	x	x	(x)	x	x	x	x	x	x		
Swedish		x	x	(x)	x	x	x	x	x	x	x	

3.3 Design of the stimulus tokens

Two main questions arose in preparing the stimulus tokens. One question is whether the stimuli should be natural or synthetic speech. Synthetic stimuli have the advantage of being very precisely controllable, but the potential disadvantage of sounding unhuman. Naturally produced stimuli, on the other hand, have a better chance of being human-sounding, but may vary in ways that were intended to be held constant.

Actually, there is little evidence that either kind of stimulus is superior to the other in this kind of experiment. Singh and Woods (1970) and Mohr and Wang (1968) used natural stimuli. Pols et al used effectively synthetic stimuli -- a single period taken from each of 11 natural vowel tokens was repeated to produce a sound of approximately 400 msec. Hanson (1967) used both synthetic (generated by OVE II) and natural speech. All these analyses gave reasonable results;

certainly there is no difference in any of the results that can be attributed to the synthetic or natural nature of the stimuli.

The most direct comparison of synthetic vs. natural speech in a perceptual distance experiment is provided by Hanson. His experiments #22 and #55 were performed using the nine Swedish vowels [i e ε a o u y ø ʉ] with listeners making numerical estimates of distance in a paired comparison task. The difference was that the stimuli in experiment #22 were synthetic, while those in experiment #55 were natural. Since the synthetic stimuli were apparently prepared first, there was no attempt to make the formant frequencies of one set of vowels coincide with those of the other set. Consequently, some differences in formant frequencies exist.

The simplest way to compare the two experiments is to compare their response matrices. Pearson's correlation coefficient r was computed for the average judged intervowel distances for experiment #22, compared with the average distances for vowels in experiment #55. In spite of the use of synthetic speech in one case and natural speech in the other, and furthermore some differences in the formant frequencies, the judgments agree very well across experiments ($r=.86$, $t=9.96$, $p<.001$).

This evidence allows the question of synthetic vs. natural speech to be treated as relatively unimportant. The second is a more substantive one: Should the stimuli be isolated, steady state vowels, or should they be embedded in a context? Here there is no evidence, since the relevant vowel studies up to this point have used more or less steady state vowels. Hanson and Pols et al used truly steady state vowels, as described above. The natural stimuli used by Singh and Woods and by Mohr and Wang were not identical from one token to the next, but they were nevertheless presented without any context and without any intentional variation of intonation or intensity.

To see why the presence or absence of phonetic context may be important, one must keep in mind the goal of discovering cross-language differences in perception, if any exist. It is not reasonable to expect cross-language differences unless the listeners are performing the task in terms of a "language set" or "speech mode." If the listeners are performing a purely psychophysical task, their responses should be based solely on the acoustic properties of the sounds, which are the same for all listeners. To what extent the listeners in the four studies mentioned above were engaged in a speech mode is not entirely clear. The natural quality of the stimuli

in Singh and Woods and in Mohr and Wang may well have been sufficient to give the listeners a speech set. There can be no doubt that the listeners in these experiments knew that the sounds they were judging were sounds of human language. As for Hanson, his subjects were asked "to notify the experimenter immediately if a signal had not been recognized as a known Swedish speech sound (p. 53)." No mention was made of any instance in which a listener could not recognize the stimulus as a Swedish vowel.

In Pols et al there are conflicting statements. At one point, the authors claim that "10 of the 11 signals were practically unanimously denominated by fifteen subjects [out of a total of 22 subjects - D.T.] as the vowels which were originally pronounced (p. 464)." This obviously implies that the stimuli were recognized as speech sounds. But on p. 466 this statement appears:

Since most of the subjects did not even realize that the stimuli were taken from speech sounds, we may assume that they did not use linguistic information in their judgments. In their opinion, they were presented with complex synthetic signals, and they based their decisions on physical cues present in the signals.

It is difficult to reconcile these two quotations, and Pols et al do not define "linguistic information" as used here.

It was thought advantageous to allow listeners in the present experiment to use whatever native phonetic or phonological prejudices they could bring to bear on the task. To make it as easy as possible to tap whatever cross-language differences may exist, natural vowel sounds were presented in a phonetic context. The constant frame [b a b _] was selected; each stimulus word was pronounced with stress on the test vowel (the second syllable) and with a falling intonation.

3.4 Preparation of the stimuli

Preparing the stimuli was a fairly tedious process. The recording was done from a sound-treated room onto a remote Ampex AG-350 tape recorder. Several sets of stimulus tokens, over several recording sessions, were recorded by the author. The twelve tokens ultimately chosen were free from noticeable diphthongization and were spoken with very similar peak fundamental frequency and intonation curves. They were also as close in duration as possible to 400 msec. As to intensity, it was attempted to hold vocal effort

constant. This naturally resulted in some vowels having a higher measurable intensity than others.

Once the test words were recorded on audio tape, they were sampled and stored on digital tape for eventual presentation to the listeners by a PDP-12 computer. The intensities were adjusted by ear in the course of transferring the stimuli to digital tape. Some adjustment of the stimulus durations was also carried out after the utterances were stored on digital tape.

A certain amount of care must be exercised in converting the signals to digital form, so as to distort the utterances as little as possible. The following discussion shows to what extent the goal of naturalness was achieved. (See Bogert, 1973, for an introductory account of the problem of digitizing and storing signals.)

In digitally storing a continuously varying signal, one must sample the signal at regular intervals and record the voltage measured at that time, the voltage being proportional to the air pressure variation when the sound was originally produced. The two relevant parameters are (1) sample rate (number of samples per second) and (2) what will be called precision, the number of binary digits used to store each sample. The frequency response is determined by the sample rate. The number of possible gradations in intensity, or amplitude, is determined by the precision.

Sample rate sets a limit of frequency response in the following way: the maximum frequency at which there may be energy present above the noise level is one-half the sample rate (one-half the sample rate is known as the Nyquist rate). The stimuli must therefore be sampled through a low-pass filter that effectively suppresses all the energy present in the signal above the Nyquist frequency. But the Nyquist rate cannot be the desired frequency response, since filters cannot exist that offer an instantaneous drop-off in intensity at the desired frequency. Consequently, the desired frequency response and the sample rate must be chosen such that some filter is capable of attenuating the signal from maximum intensity to the noise level between the desired frequency response and the Nyquist rate.

The filter used in the present study was a UTC FLL-3500. This filter has a response of -3 dB at 3500 Hz and -40 dB by 4500 Hz. The sample rate therefore had to be at least 9000 Hz. This allowed a perfectly satisfactory frequency response. There was no attenuation of the third formant of any vowel; in fact, the fourth formants of the stimulus tokens were present with only slight attenuation.

Now consider the problem of precision. As mentioned above, the more bits per sample, the finer the intensity gradations. Obviously, the greater the precision, the more closely the stored samples approximate a continuous curve. Flanagan et al (1970) state that speech is "typically" stored at 7 bits per sample (p. 22). We will take this number as a reasonable lower limit for the quantization. This gives each sample a range of 2^7-1 , or 127 possible non-zero values.

It is convenient to discuss intensity in terms of decibels. The difference in decibels for any two signals A and B is expressed by

$$\text{dB} = 20 \log \frac{\text{pressure of A}}{\text{pressure of B}},$$

where pressure refers either to actual sound pressure or voltage. For a ratio of 127 to 1,

$$\begin{aligned} \text{dB} &= 20 \log \frac{127}{1} \\ &= 20 \log 127 \\ &= 20 \times 2.104 \\ \text{dB} &= 42. \end{aligned}$$

For a 7-bit signal, then, there is a difference of 42 dB between the loudest possible signal and the softest one. (For 4 bits and above, $\text{dB} = 6n$, where n = number of bits.)

Given the two minimum requirements of a 9000 Hz sample rate and a precision of 7 bits per sample, the practical problem of storing the samples in the available space can be discussed. The computer that controlled the experiment was a PDP-12 with 12 k of core memory (1 k = 1024 words). In order for the listener to have immediate access to all three members of a stimulus triad, the control program and three stimuli must reside in core simultaneously. The control program occupies 2 k of storage, leaving 10 k for the stimuli. It was convenient to limit each stimulus to 3 k, or 3072 words of memory.

Each stimulus token is approximately 400 milliseconds long. To store 400 msec. at 9000 samples per second, 0.4×9000 , or 3600 samples are required. It is immediately clear that 3 k of storage is insufficient, if the simplest storing procedure of one sample per one computer word is used. A common solution is the use of one-half word, or 6-bit samples, stored 2-to-a-word. But preliminary data stored at 6 bits was informally judged to be unsatisfactory. This finding is supported by Flanagan's et al recommendation of 7 bits per sample.

The ultimate solution was to pack the stimuli with every three samples compressed into every two words, that is, eight bits per sample, using every available bit. This allowed storage of 4608 samples for each stimulus, and the precision is better than the minimum required. Notice that the ratio of most-to-least intense sounds is now at $6 \times n = 6 \times 8 = 48$ dB, based on a precision of 8 bits rather than 7. Since the filter was capable of reducing signal strength only by 43 dB, some low-level noise originally present above 4500 Hz could theoretically have been folded down into the audible frequency range of the stimulus. If this sort of effect did in fact occur, it had no effect on the naturalness of the stimuli and caused the listeners no distraction.

The sample rate that eventually was used was 11,430 Hz. This is the fastest rate possible that allowed a 400 msec. signal to be stored in the available space. Once the test words were successfully sampled and packed onto a digital tape ready for playback, some physical properties of the stimuli were measured. The measurements were based not on the original test words recorded on audio tape, but on the versions stored on digital tape, as the listeners would hear them.

A reference list of the measurements made is shown in Table 3.3. The formant frequencies and test vowel durations were measured from wide band spectrograms made on a Kay 6061A Sonagraph. The formant amplitudes were obtained from a Fast Fourier Transform procedure written for the PDP-12 computer by Lloyd Rice. Overall amplitude values are RMS values read from a B & K 2409 voltmeter, normalized so that the least intense stimulus, [æ] as it happens, was given a value of zero. Finally, peak fundamental frequency was computed by measurements taken from a static waveform display generated by the PDP-12. The duration of the first six periods of the test vowel, excluding the first two periods after the release of the [b] closure, was taken as a reasonable basis for computing the peak pitch. Since the vowels were spoken with a falling intonation, this estimate is slightly low, but the relationship between vowel tokens is unaffected.

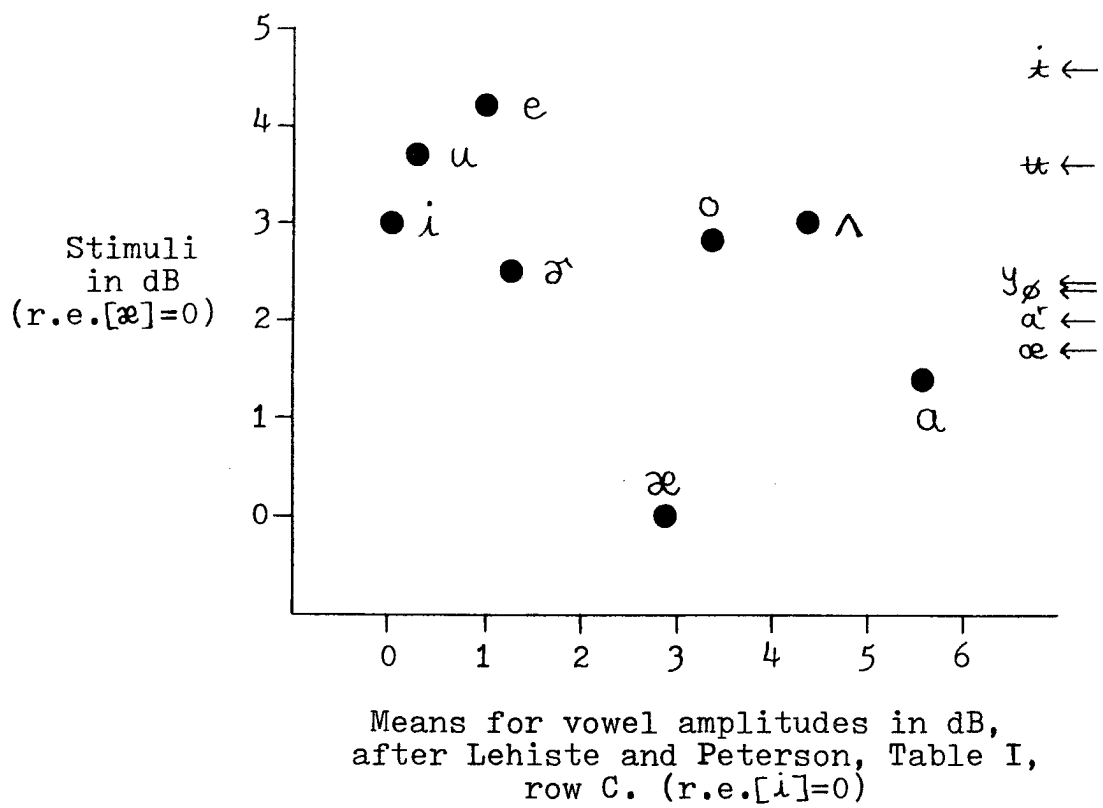
The acoustic variables will be referred to extensively in Chapters 5 and 6. Linear regression analyses will be used to help determine what aspects of the signal were significant in shaping the listeners' judgments. Some preliminary comments are appropriate here.

The overall amplitude (A0), fundamental frequency (F0), and duration (D) values require some comment in this regard.

Table 3.3: Acoustic measurements made on stimulus vowels.

	F0	F1	F2	F1+F2	F2-F1	F3	F4	A0	A1	A2	A3	A4	Duration (msec)
ɑ	125	610	1215	1825	605	2175	3530	2.0	33	31	17	11	270
ʌ	125	480	1215	1695	735	2480	3620	3.0	37	37	28	13	240
ɪ	124	305	1695	2000	1390	2520	3820	4.6	37	27	22	8	270
ɥ	128	215	1870	2085	1655	2175	3190	2.3	33	22	18	15	225
ø	125	305	1610	1915	1305	2175	3390	2.4	38	23	18	16	240
æ	124	350	1305	1655	955	1825	3500	2.5	39	27	23	14	255
æ	124	520	1740	2260	1220	2390	3420	0.0	36	25	19	8	265
u	126	260	695	955	435	2305	3400	3.7	36	18	9		255
o	130	305	740	1045	435	2391	3500	2.8	36	30	9	7	225
ɑ	126	610	1130	1740	520	2565	3500	1.4	37	31	21	11	270
e	129	350	2175	2525	1825	2565	3640	4.2	34	25	20	13	235
ɪ	122	260	2215	2475	1955	2825	3500	3.0	34	20	17	18	225
œ	122	405	1375	1780	970	2220	3500	1.7	36	22	9	9	240
æ	126	285	1415	1700	1130	2180	3440	3.6	37	22	15	6	235

All frequency (F) values are in Herz. A0 values are dB relative to [æ] = 0.
All other levels are referred to A4 of [u] = 0.

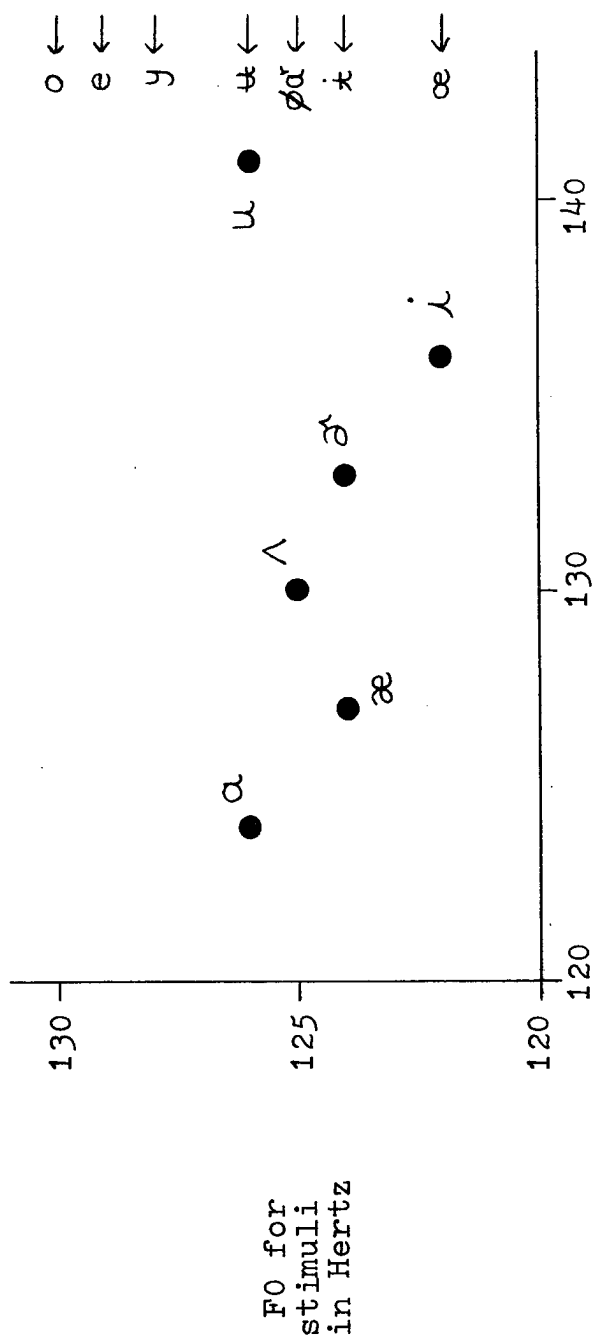


While it is obvious that any vowel can be spoken over the same range of amplitude, fundamental frequency, and duration, it is nevertheless true that under controlled conditions vowel quality is systematically related to these parameters. It is of some interest to notice how the A0, F0 and D values of Table 3.3 are related to the expected values of different vowels.

It has been observed (Lehiste and Peterson, 1959; Ladefoged, 1967) that the overall amplitude of vowels generally decreases as vowel height increases and as rounding increases. Figure 3.4 compares the amplitude of the stimulus tokens with measurements of somewhat similar American vowels made by Lehiste and Peterson. Vowels without any near American equivalents are represented by arrows at appropriate vertical locations on the figure, as they cannot be plotted. Notice first that the range of relative amplitudes is about the same for both stimuli and reference: 4.6 dB and 5.6 dB, respectively. The stimulus vowels occupy a smaller range (3.2 dB) if the outlying [æ] is not counted. But the normalization of vowel amplitude has overcompensated to some extent, giving higher or more rounded vowels a greater amplitude than the low unrounded vowels. If A0 is shown to underlie any results of the scaling analyses, its effects should be reversed.

F0 is likewise linked to vowel quality, in that high vowels tend to be spoken at a higher pitch than low vowels (Peterson and Barney, 1952). F0 for 6 stimulus vowels (Figure 3.5) is plotted against Peterson and Barney's means for male speakers for similar American vowels. The range of reference vowels is 17 Hz, while the range of stimuli is only 8 Hz. The latter is considerably less than for the reference means, but clearly audible. The patterns of variation, however, comparing stimuli and reference, are largely unrelated. Any experimental results that can be shown to depend upon F0 variation are not extendable beyond this particular set of vowel tokens.

The last of these three parameters, duration, decreases roughly with vowel height (Peterson and Lehiste, 1960). The plot of stimulus vowels against similar American vowels (Figure 3.6) demonstrates that the stimulus vowels had duration relationships very similar to those for the reference vowels. Of the stimuli with no counterparts for reference, only [ɛ̃] violates the informal prediction as to where it should belong: it has too long a duration for a high vowel. If any aspects of the scaling solutions are difficult to interpret because of the vowel [ɛ̃], and if they are traceable to the stimulus durations, then the difficulty is purely an artefact of the stimuli.



Means for F0, after Peterson and
Barney, Table II, Row M.

Figure 3.5: Fundamental frequency of stimulus vowels compared with reference data.

The remaining measures all concern formant values. The first three formant frequencies (F1, F2, F3) require no comment. The most satisfying perceptual dimensions will presumably be based on these three parameters. F4 and A4 are included in Table 3.3 to demonstrate the naturalness of the sounds, but are not used in any regression analyses.

Formant amplitudes were included in many regression analyses. This may be seen as a questionable effort, since Fant (1956) has shown that formant amplitudes are predictable if formant frequencies are known. Thus, formant amplitude appears to be redundant information, and need not be entered as regression variables. However, this criticism is not valid, since formant levels are not linearly related to formant frequency. If in fact listeners were incorporating information about formant levels in their responses, this could be discovered most directly by allowing formant amplitudes as independent regression variables.

But by no means are A1, A2, or A3 expected to emerge as significant influences. Flanagan (1957) estimates that the minimal audible difference in amplitude for formant 2 (and presumably formant 3 as well) is on the order of 5 dB, allowing only a gross quantization of the range presented to the listeners.

3.5 Selection of the listeners

The main source of subjects was a list of foreign students enrolled at UCLA, provided by the UCLA Foreign Student Office. By means of that list and various other contacts it was possible to engage eight native speakers of English, seven of German, six of Thai, seven of Turkish, and seven of Swedish. Listeners were paid for their services. It was hoped that foreign subjects could be selected who had spent no more than one to two years in the United States, but this goal was found to be unrealistic. None of the listeners had any extensive training in linguistics or phonetics, though several had been acquainted with the fundamentals of phonetic theory in the course of studying their own language, or in learning a foreign language. Most listeners professed no knowledge of phonetics whatever.

It would have been ideal to use subjects who were monolingual and totally naive as to any aspect of phonetics, linguistics, or conscious thought about language. However, since the experiment was necessarily carried out in the United States and within a relatively short time span (approximately six months), it was not possible to obtain foreign speakers who knew only one language. Only two English listeners, in fact, out of all 35 listeners, had

no knowledge of a foreign language. The existence of such linguistic contamination undoubtedly affected the listeners' behavior, but there is no a priori way of telling how much. Presumably, a listener's knowledge of foreign languages, namely English, would tend to level out whatever language-specific perceptual differences that may exist. Therefore, if any language-specific differences appear in the results, it seems fair to conclude that these differences are deep-seated and do not necessarily disappear if the listener undertakes to learn a second language. There may be other, more elusive properties that are obscured by the speakers' contact with English; these properties, if they exist, are inaccessible in the present experiment.

3.6 The experimental task

For reasons discussed in Chapter 2, listeners were asked to make triadic comparison judgments. Judgments of "most similar" and "most different" were accumulated into a vowel-by-vowel response matrix in the following way.

Stimuli were presented in groups of three, each such group being called a triad. The listener is required to determine which two stimuli in a triad appear to be the most similar, and which two appear to be the most different. Three judgments are being made, in effect. Assume, for example, that the triad being considered is $[i\ e\ a]$, more correctly $[b\acute{a}bi\ b\acute{a}be\ b\acute{a}b\alpha]$. Assume that the listener decides that $[i]$ and $[a]$ are the most different, and that $[i]$ and $[e]$ are the most similar. He has therefore established the following three inequalities:

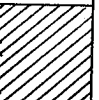
$$(1) [i\ a] > [i\ e]$$

$$(2) [i\ a] > [a\ e]$$

$$(3) [a\ e] > [i\ e]$$

Now for each inequality, one entry is made in a vowel-by-vowel matrix in the cell appropriate to the left side of the inequality. An entry can be thought of as a tally mark, or a point score of one; it makes no difference. For the example, the relevant portion of the vowel matrix is Table 3.4. Inequality number (1) directs that an entry be made in the cell $[i]-[a]$. Inequality number (2) requires a second entry in the same cell, and inequality number (3) places a mark in cell $[a]-[e]$. Cell $[i]-[e]$ receives no mark, since the difference between $[i]-[e]$ was not greater than any other difference. Notice that the matrix is symmetrical about the main diagonal, and scores are redundantly accumulated into both halves.

Table 3.4: Sample submatrix illustrating the accumulation of scores from vowel dissimilarities.

	i	e	a
i			
e			
a			

These marks are entered cumulatively for each of the 220 possible triads ($12!/3!(12-3)! = 12 \times 11 \times 10 / 6 = 220$). A score of 0 in a complete matrix means that the vowel pair concerned was thought of as the most similar pair every time it appeared. Conversely, an eventual score of 20 (each cell was referred to 10 times) showed that the listener found that particular vowel pair to be the most different every time it occurred. Values of 1-19 represented pairs of vowels that were judged differently, depending upon the third member of the triad. Large values are taken to indicate relatively large distances in the perceptual space, and smaller values to indicate smaller distances. The cells on the main diagonal (top left to bottom right) are zero throughout, by definition. No triads were presented in which two or more vowels were identical.

To make simpler the presentation of stimuli, the recording of listeners' responses, and the computation and storage of vowel matrices, a PDP-12 computer was programmed to perform all these tasks. This required that each listener take part in the experiment one at a time.

Listeners were provided with a small metal box, containing five pushbuttons and two lights, and a sixth pushbutton, held separately, with which they communicated remotely with the computer, a room away.

The three pushbuttons, arranged in an equilateral triangle, were used to play the three stimuli in a triad. One complete triad at a time resided in core memory, and pressing a specific button caused one member of the triad to be played to the listener over a loudspeaker. Judgments were recorded in the following way. Once the listener had decided which two stimuli were "most different," he would

press the button labelled D, turning on the adjoining red light. The last two stimuli played before pressing the D button a second time, turning the light off, were entered in the computer memory as the "most different" pair. The button labelled S was used in exactly the same way to record "most similar" responses (except that the light was a green one). If both lights were inadvertantly turned on at once, the program played no stimuli, allowing no responses to be recorded until one of the lights was turned off. When both responses to a triad had been made, the listener pressed the separate, sixth button, causing the computer to read the next triad from digital tape into memory, and the cycle would begin again.

The system was designed to be as flexible as possible. Stimuli could be played in any order and repeated at will. Also, either judgment (similar or different) could be made before the other, and a judgment could be corrected by simply entering the new response as though the previous, wrong response had not been made at all. Since the buttons were inoperative while the tape was being read, the listener could furthermore use one of the lights to inform him when the next triad was ready: he need only press button six to advance to the new triad, immediately press either the S or D button, and wait patiently until the light came on. Or, of course, he could alternatively wait patiently without the light, and test the playback buttons now and again.

This flexibility, however, did not extend to the entering of points in the response matrix. Each triad must be given equal weight, which means that each triad must result in a total of three-point scores added into the matrix. Therefore, it cannot be allowed to enter the same vowel pair as both similar and different, since the inequalities upon which the point scores are based would not be well-defined. It is also not allowable to make no judgments at all for a triad, skipping directly to the next one.

To prevent these and other unallowable responses from occurring, whether by accident or listener's intent, several safeguards were added. When the advance button is pressed, the program makes the following sequence of checks:

- (1) Have all stimuli been played at least once?
- (2) Has a response been entered as "most similar" during the course of the present triad?
- (3) Has a "most different" response likewise been added?
- (4) Are the two responses different?

(5) Are both lights off?

Only if all questions can be answered "yes" does the program record the responses and begin setting up for the next triad. If any question receives a "no," it is up to the listener to make whatever correction is necessary.

The stimulus triads were arranged in a pseudo-random order such that no consecutive triads contained any of the same vowel sounds. When a vowel did reappear several triads later, it was not played by the same button as before.

It was originally decided that each listener was to perform the entire task in three sessions, presumably 45 minutes to an hour in duration. The list was therefore split into three parts. At the beginning of list 1, a set of 20 practice triads was added, not to be scored, taken from list 3. At the head of lists 2 and 3 were added 10 practice triads, taken from lists 1 and 2, respectively. The listeners were therefore to respond to 260 triads, according to this schedule:

Session 1	20 initial practice triads
	67 scored triads
Session 2	10 refresher triads
	77 scored triads
Session 3	10 refresher triads
	<u>76</u> scored triads
	260

They were told that the first set contained 10 practice triads, and they were not told that there were any practice triads at later sessions. Subjects were encouraged, but not required, to take 5-minute breaks at 20-minute intervals, in order to come back to the task refreshed.

Listeners were seated in a small room; the loudspeaker over which the stimuli were heard was seated on a shelf at eye (or ear) level with the subject, about three feet away. The stimuli were heard at a comfortable level, approximately 60 dB spl. After giving the listener the instructions as to the task, the experimenter (the author) monitored the vowel sounds being played, and the responses, at the computer console in an adjoining room. At the end of each session (or list, c.f. Section 3.7), the listeners' judgments were stored on digital magnetic tape. After a listener was through with the entire task, a master list containing all his triadic judgments was compiled and stored. The matrices of vowel-by-vowel dissimilarities could then be assembled later, at the experimenter's convenience.

3.7 Notes on the listeners' behavior

The listeners had no difficulty at all operating the control pushbuttons. They understood the instructions after three or four triads, with the experimenter present to demonstrate the use of the controls; all subjects were fluent with them by 10 triads, usually sooner.

Listeners were told to work as fast as they could in a comfortable way. A few listeners began by pondering long before making their responses, and after a few minutes were encouraged to rely more on first impressions, rather than to deliberate painfully over each judgment. This suggestion made them work not only faster, but with less frustration as well. Many listeners settled into a fairly regular pattern, listening to each vowel once, responding quickly, and going ahead. Most, however, played each member of a triad at least two or three times before responding.

Occasionally, a listener would become stuck on one triad and not be able to advance to the next. On these occasions the experimenter would interrupt and suggest that the current triad be listened to again, as though it had not yet appeared. This was always sufficient to correct whatever error the listener had made in his responses that was keeping the program from continuing.

A major discrepancy between plan and actuality had to do with the speed with which many listeners performed the task. Some subjects worked very fast, and finished the first third well under an hour. These were allowed to begin the second set of stimuli at that session. Others could not finish 2/3 of the first set in an hour, and on their return had to complete set 1 before going on to set 2. It was obviously necessary to abandon the fixed scheme of three sessions described in the previous section. The program was modified to allow the experimenter to stop anywhere within a list and store the partial response series on tape; it was similarly modified to start at any desired place within a list.

It was still thought desirable to require that each new session begin with a set of 10 practice triads. If a listener began a session somewhere within a list, rather than at the beginning of it, the 10 preceding triads on the list were used as the practice triads. And if he began a new list within an experimental session, rather than at its start, the practice triads at the head of that list were deleted. In this way, different listeners came to have different practice triads from others, and different numbers of practice triads as well.

It further occurred that several Thai listeners, for whom vowel duration and lexical tone are distinctive, made responses to the first several triads with the experimenter present that showed they were attending to relative duration, or pitch, or both. They were asked to attempt to ignore such distinctions and instead focus on whatever else made the vowels seem similar or different. Subsequent monitoring showed that those subjects were responding in a plausible way to the concept of vowel quality as intended by the experimenter. No such additional clarifying instructions appeared to be necessary for the other listeners, and none were given.

Chapter 4

Preliminary Survey of the Collected Data

4.0 The present chapter describes some preliminary investigations of the listeners' data. The data are first examined for reliability (4.1) and validity (4.2). It has not been customary in MD scaling of speech sounds to explicitly study the subjects' ability to behave in a consistent way. But if the third mode of a three-mode scaling analysis is comprised of individual listeners' data, it is imperative to know to what extent each listener's data is reliable and valid. By reliability is meant conformance with a priori expectations of the data.

This is followed by hierarchical clustering analyses for each language (4.3), which provide a more thorough overview of the data as well as an initial look at the kinds of language-specific phenomena that will be found in the MD scaling analyses. Section 4.4 describes the conversion of the listeners' raw data matrices into "absolute distance" matrices, preparatory to the PARAFAC analyses reported in Chapters 5 and 6.

The data matrices for each listener are presented in Appendix I. The English subjects are labelled E1-8, the German G1-7, the Thai T1-6, the Turkish K1-7, and the Swedish S1-7. Appendix I also includes matrices summed across persons for each language, respectively, and the 10x10 vowel matrix for all listeners taken together. This matrix was computed by summing point scores for triads involving only the ten vowels heard by all subjects: [i e a u o ɜ æ ø y ɤ].

4.1 Reliability of the data

It is reassuring to see some evidence that the listeners were responding in an internally consistent way. If the analyses were to conclude with two-mode MD scaling, this would be of much less import. The examination of Hanson's work in Section 2.5 clearly shows that data from unreliable subjects can be rendered highly stable by summing across subjects. But in a three-mode analysis, the data furnished by each subject remains independent of the others, and excessive noise in the data can cause difficulties in convergence. Some measure of intrasubject reliability, which is equated here with each listener's internal consistency, is necessary.

One customary method of measuring reliability is to compare data summed over one-half of an experimental sample with corresponding data from the other half: the more similar the two halves, the more reliable the whole. However, the data at hand do not lend themselves to this treatment. There are not enough listeners for a meaningful split-half comparison for any one language; furthermore, it is not assumed that the entire group of 35 listeners is homogenous, a requirement for doing a split-half test on the entire sample. The good interpretability of the results discussed in this and the succeeding chapter may be taken as a priori evidence for the overall reliability of the listeners. But these methods, even if applicable, would still tell us little about the listeners' internal consistency.

4.11 Observations of discrepancies for repeated triads

Each listener heard from 42 to 80 triads twice, as described in Chapter 3. Only one presentation of each triad contributed to the values in the vowel matrices, but the existence of repeated triads allows the assessment of internal consistency. It was decided to discard the first ten triads, even for reliability testing, since those triads are the ones with which listeners learned the task.

Since the listeners did not associate magnitudes with vowel pairs directly, it was impossible to simply correlate one set of pairwise judgments with another. Instead, a test was performed based on observed discrepancies in responses to identical triads.

There are six different ways to judge a triad, as outlined in Table 4.1. In this example, the possible judgments are being compared to a reference judgment of ab = most similar, ac = most different, and bc = "middle" (unspecified). If on the second hearing of a triad the listener responds exactly as he did the first time, he complies with row A in the table. If only one of his two judgments is the same as for the first presentation, then the situation corresponds to rows S or D. The two rows labelled N reflect the responses in which none of the judgments is repeated, including the tacit "middle" distance. Row R represents the least internally consistent behavior, in which the overt judgments of similarity and difference are reversed.

An overview of the discrepancies, by language, is found in Figure 4.1. The majority of repeated triads was judged exactly the same as the first presentation of those triads, as seen in column A for each language on the figure. The next most frequent repetition type was the identical judgment of similar only, followed by different only. For

English, German and Turkish, the number of N repetitions is lower than D repetitions. For Thai and Swedish, where the number of N and D repetitions is about equal, there are still many more D repetitions than expected by chance. A Chi-square test for the relations between columns N and D gives $\chi^2 = 8.59$, $p < .005$ for Swedish, and $\chi^2 = 6.97$, $p < .01$ for Thai. In any comparison involving frequency of "bad" responses R and N, and "good" responses D, S, and A, the significance values for the two tests just made are the worst possible values that can appear. It may therefore be inferred that for each language group, all the R and N repetitions are significantly fewer than chance would predict, and all the D, S, and A repetitions are significantly more frequent. It is extremely likely that a listener will repeat at least one, if not both judgments, when asked to judge a triad a second time.

Table 4.1: Six ways to judge a triad abc compared to a reference judgment: ab = most similar, bc = "middle" or unspecified, ac = most different.

probabi- lity of each type occurring randomly	MOST SIMILAR	"MIDDLE"	MOST DIFFERENT	Label	Description
1/6	ab	bc	ac	A	<u>all</u> judgments identical
1/6	ab	ac	bc	S	only <u>similar</u> judgments identical
1/6	bc	ab	ac	D	only <u>differ-</u> ent judgments identical
1/3	ac bc	ab ac	bc ab	N	<u>no</u> judgments in common
1/6	ac	bc	ab	R	<u>reversal</u> of similar and different judgments

Within the realm of the two least frequent repetition types, R and N, the instances of R are somewhat fewer than instances of N, fairly close to the chance pattern. Chi-square tests comparing these two items within languages

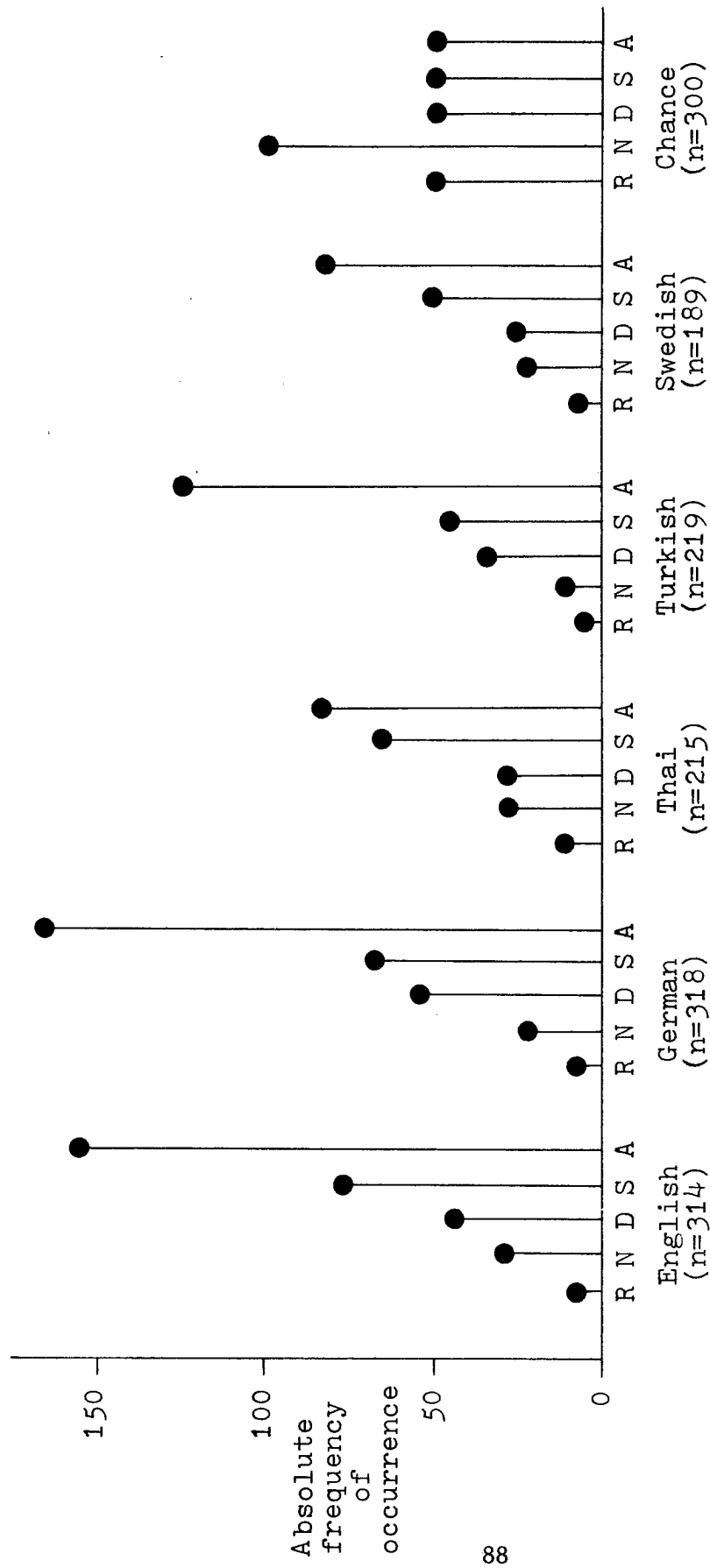


Figure 4.1: Summed intrapersonal discrepancies in judgements of identical triads.
 R = complete reversal of similar and different judgements.
 N = no pairs, including the middle pair, judged the same.
 D = judgement of most different was repeated, nothing else.
 S = judgement of most similar was repeated, nothing else.
 A = all judgements were repeated exactly.

Table 4.2: Reliability of individual subjects.

Listener	T	NBR	Repetition types					χ^2
			R	N	D	S	A	
E1	3.0	30	0	2	3	8	17	42.8
E2	1.7	30	0	3	5	10	12	24.7
E3	1.5	41	0	3	7	8	23	53.5
E4	1.1	58	1	5	10	11	31	64.0
E5	1.5	54	1	8	4	20	21	44.9
E6	2.5	30	2	6	1	7	14	23.6
E7	2.0	30	0	0	4	6	20	60.4
E8	2.1	41	4	2	11	7	17	28.2
G1	2.0	48	1	2	11	10	24	45.9
G2	1.8	70	2	3	10	18	37	82.5
G3	2.4	30	0	4	3	2	21	62.5
G4	1.9	41	1	4	2	9	25	56.7
G5	1.4	30	1	3	8	8	10	16.7 p<.003
G6	2.7	30	1	2	7	7	13	24.0
G7	1.8	69	2	5	13	13	36	74.5
T1	2.3	30	0	4	5	11	10	20.8
T2	2.8	30	2	3	1	4	20	55.1
T3	2.9	40	3	5	4	13	15	24.6
T4	1.5	30	0	6	3	10	11	14.8 p<.006
T5	1.2	46	3	4	11	16	12	24.2
T6	2.0	39	3	6	4	11	15	20.1
K1	1.8	62	0	3	8	15	37	95.3
K2	1.6	22	0	2	6	1	12	38.9
K3	2.5	30	0	0	1	9	20	66.4
K4	2.1	39	0	0	7	8	24	67.0
K5	2.3	22	3	2	0	6	11	23.6
K6	2.4	21	2	4	4	6	5	4.4 p<.5
K7	3.1	23	0	0	8	0	15	26.5
S1	2.5	24	0	6	8	7	8	13.5 p<.01
S2	2.6	30	2	2	6	11	14	34.8
S3	2.0	32	0	3	1	5	21	54.4
S4	2.5	30	0	3	7	13	9	25.9
S5	3.0	22	1	3	0	5	6	9.4 p<.06
S6	2.2	21	2	3	4	5	7	7.2 p<.14
S7	2.1	30	2	3	8	5	17	36.3

T = number of triads judged per minute.

NBR = number of triads heard twice, minus 10.

Repetition types: c.f. Table 4.1.

χ^2 = test of the null hypothesis for listeners' ability to make repeat judgements. For all χ^2 , $p < .001$ except as noted.

unanimously fail to reach even the .1 level of significance. There is no tendency to make significantly fewer R repetitions than N repetitions.

In an attempt to spot and perhaps remove unreliable listeners from further consideration, Chi-square was computed for each subject, testing the frequency of each response type against the null hypothesis. These data appear in Table 4.2. English listener #1, for example, with 30 repeated triads, has a set of observed and expected values as follows:

repetition type	observed	expected
A	17	5
S	8	5
D	3	5
N	2	10
R	0	5

$\chi^2 = 42.8$. Since $p = .001$ for $\chi^2 = 18.5$, subject E1 is behaving very reliably; in fact, all but six exceed the .001 significance level. For three of those listeners (G5, T4, and S1), $p < .01$, which is well below the criterion of $p < .05$.

The remaining three listeners deserve some comment. Two Swedish listeners, S5 and S6, show $p < .06$ and $p < .14$, respectively. Notice, though, that these two listeners heard the fewest repeated triads. Had S6, for example, heard twice as many repetitions, each repetition type also being doubled, his Chi-square would have been 14.3, with $p < .008$. The same reasoning applied to S5 gives an even greater improvement. The χ^2 values for S5 and S6 in Table 4.2 are therefore not regarded as strong evidence in favor of discarding those listeners. One Turkish subject, K6, was apparently very inconsistent ($\chi^2 = 4.4$, $p < .5$). No speculation about an improved χ^2 can help here. But since his distribution of scores in the response matrix (Section 4.2) was well within the range of the other listeners, his data were retained.

4.12 Direction of deviation from chance values

It should be pointed out that Chi-square is not sensitive to the direction of deviation from expected values. In making claims of reliability, it is crucial that the R and N repetitions be fewer than chance, and that D, S and A repetitions be more frequent. Impressively enough, the number of R and N repetitions was below the chance level for each listener. Likewise, all subjects showed more A repetitions than chance would predict, and 29 of the 35

subjects made more S repetitions than expected by chance.

D repetitions, however, showed a different pattern: 21 listeners made fewer D repetitions than expected by chance. To some extent, this can be explained according to the Weber-Fechner "law": just-noticeable differences are based on proportions, rather than the absolute magnitude of the elements being compared. For two large distances, judging which is the larger should give relatively inconsistent results. In a triangle with sides 3, 4, 5, for example, it is much easier to judge which side is longer than it is for a triangle with sides 3, 18, 19. However, the same principle may make similar judgments difficult, as well; triangle 3, 4, 5 will induce more reliable judgments of similarities than triangle 15, 16, 30. Perhaps many more vowel triads were perceived as the 3, 18, 19 type than were perceived as the 15, 16, 30 type.

Another possibility is that judging "most different" is intrinsically more difficult than judging "most similar." The concept of similarity has a limit, namely, "identical," but there is no comparable concept of "as different as possible." Judging "most different" may be a less well-defined task.

4.13 The effect of rate of judgment on reliability

As mentioned in Section 3.7, listeners were encouraged to work as fast as was comfortable, but no other restriction or requirement was placed on them in this regard. Mean rate of judgments, in triads per minute, was computed from the record of subject participation, in which the duration of listening sessions was logged to the minute. These rates are given for each listener in column T of Table 4.2. There is no apparent relationship at all between rate of judgment and reliability of responses. There is no evidence that having the listeners speed up or slow down, or making their rates more uniform, would have effected any improvements in the results.

4.2 Validity of the data

The extensive discussion of the reliability of the data in Section 4.1 must be complemented by a discussion of validity. There is no more reliable source of data than a stopped clock. But the price is zero validity, if the object is to measure time of day. There are no explicit criteria against which the validity of these data can be assessed; one purpose of the entire experiment is to discover what aspects of the stimuli are in fact relevant to listeners' judgment. As will be seen in Chapters 5 and 6, there are acoustic measures that are reflected in the data; a correspondence between acoustic and perceptual data certainly

counts as evidence for the validity of that data. But in order to provide independent evaluation of validity, and in addition to provide an overview of the contents of the data matrices, the following two sections are presented. They refer to the completed vowel-by-vowel matrices, not to any individual triadic judgments. The concern is no longer the stability of the listeners' responses; it is, rather, whether or not the vowel matrices match up in some way with a priori expectations.

4.21 The distribution of values within each individual vowel matrix

The sort of expectation discussed here is a purely statistical one: How similar is the distribution of scores in each data matrix to a random pattern? A computer was used to generate 12 matrices, size 12 x 12, by randomly assigning similarity inequalities for each triad. That is, the program simulated 12 listeners who went through the motions of performing a set of triadic comparisons, but made all their judgments without listening to the sounds. These inequalities were then assigned point scores and accumulated into matrices, as described in Section 3.6. The mean for each cell was then computed. There were few cells with extreme values, since extreme values require consistent treatment of vowel pairs. For example, the only way to accumulate the value 1 for a vowel pair is to regard that pair as "most similar" 9 out of 10 times. This should not occur often for random judgments. As it happens, for only one run out of the twelve was there produced even one cell with a value of 0, 1 or 2. Most of the values lay in the middle range.

If a frequency histogram based on the listeners' responses resembles a random pattern too closely, one of the following three possibilities must be entertained, which are subsumed under the label of Outcome 1:

a) The subject is responding randomly. However, it has already been shown that this is unlikely, since in almost every case the listeners' reliability is far better than chance (c.f. Section 4.1).

b) A subject may have some reliable but deviant way of assessing the triads. That is, he has some consistent way of responding, but it is unrelated to the similarity he perceives between sounds. For example, he might always press a particular pair of buttons for the "most similar" judgments.

c) The listener's conception of similarity happens to map into a space with few very large and few very small distances between vowels. This is not an implausible result. However, the more a listener's distribution of values resembles a random one, the more similar all distances become to each other. The extreme possibility is a space in which all vowels are simply "different" from all others. This would support a model of perception in which vowels are recognized as wholes and are not analyzed into features.

If, on the other hand, a subject's distribution of values is significantly different from the random distribution, it must be assumed that he was behaving non-randomly and, furthermore, that he perceived a wide range of possible distances between vowels. In this case, the distances are wholly compatible with a spatial model; in fact, the set of distances needs a spatial model in order to be captured adequately. This will be called Outcome 2.

Listeners' distributions are summarized by language in Figure 4.2. The vertical scale represents the number of times a given response was made. The horizontal scale groups the range of possible scores in each cell (i.e. 0 to 20) into four intervals: 0-6, 7-9, 10-12, and 13-20. The random pattern of response for each interval is indicated by the solid line; the points plotted on the figure are means for each language group. For example, the point labelled T (for Thai) in column 0-6 shows that Thai data matrices tended to have slightly fewer than 14 cells filled by values of 0, 1, 2, 3, 4, 5, or 6.

It is clear that the general pattern for the data is very different from the random one. There are many more smaller values and larger values than predicted by chance; likewise, there are fewer values in the intermediate range. This is an unquestionable case of Outcome 2.

While the overall pattern is a satisfying one, it is important to evaluate the responses one subject at a time. A χ^2 test was performed for each subject, testing the null hypothesis that the individual data matrices did not conform to Outcome 2. The scores were grouped into the same intervals that were used in Figure 4.2. According to Hays (1963), it is important that each interval has an expected value of at least 5; it is irrelevant whether the intervals are equal in size. The expected values in this test were the random frequencies; only by grouping together the scores zero through six could satisfactory expected values be obtained.

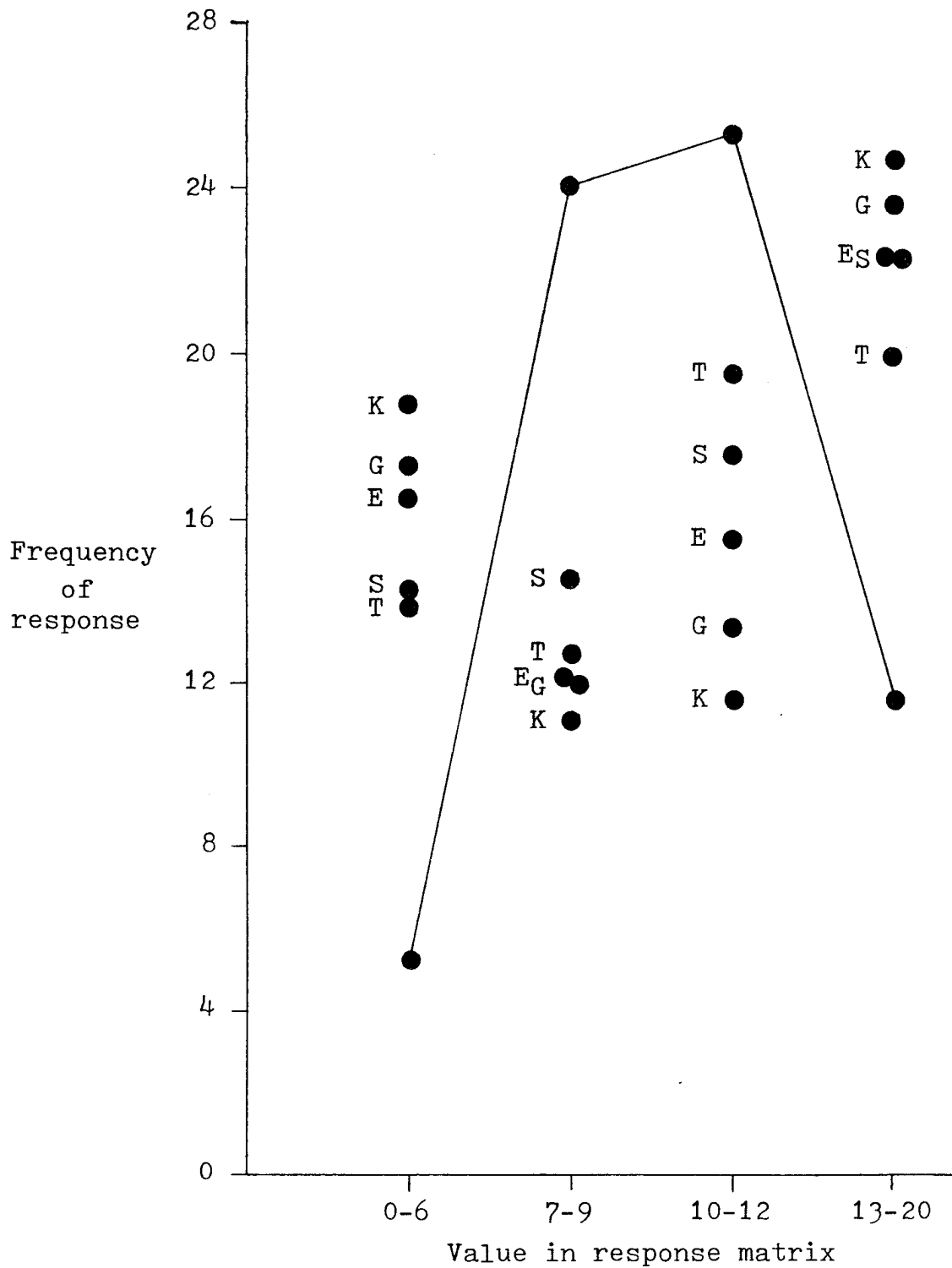


Figure 4.2: Mean frequency distributions for the five language groups compared with the random distribution (solid line). E=English, G=German, T=Thai, K=Turkish, S=Swedish.

Table 4.3: χ^2 comparing the distribution of scores in listeners' data matrices with a random distribution

	English	German	Thai	Turkish	Swedish
1	51.0	41.7	38.3	65.9	20.6
2	51.3	54.9	21.9	28.9	10.0**
3	53.9	42.3	25.0	118.7	57.7
4	32.9	82.3	21.8	86.9	31.3
5	35.5	63.5	32.8	47.0	22.7
6	13.1*	42.9	24.0	41.4	35.1
7	62.6	47.3		77.9	46.8
8	53.9				

* $p < .005$

** $p < .02$

for all other values, $p < .001$

Table 4.3 gives χ^2 for each listener. All but two listeners have values well beyond the .001 level (for $p = .001$, $\chi^2 = 16.3$). For those two subjects, E6 and S2, $p < .005$ and $p < .02$, respectively. It must be stressed that the entries in each data matrix are dependent in a complex way, as a result of the rigid requirements imposed by the process of creating the matrices from triadic judgments. The values reported are therefore inflated by some unknown amount; the significance values are not to be taken literally. But even if the usual criterion for significance is reduced 10-fold, to .005, all results except those of E6 and S2 are still highly significant, and therefore match Outcome 2.

It was decided not to discard the data from subjects E6 and S2. Although their response patterns were more like the random distribution than any other subject's data, their reliability tests (Section 4.1) were highly significant. It is very unlikely, therefore, that they were responding randomly, or making judgments based on strange criteria unrelated to the vowels. It is more reasonable to consider their distributions as instances of Outcome 1c -- their perceived distances between vowels happens to map into a space with more middle-sized distances than long or short ones.

4.22 Inspection of the content of the response matrices

This section is intended to provide an initial look at the content of the vowel matrices. It is of interest to see whether or not the matrices make sense in some intuitive way, according to expectations based on phonetic description of vowel differences. The smaller the value, the more similar the descriptions for the vowels in a pair are expected to be.

Simple examination of the data matrices (Appendix I) shows this to hold in general. For example, the distance from [ɪ] to [ɑ] is expected to be greater than that from [ɪ] to [e]. [ɑ] is a back vowel, [ɪ] and [e] are not; [ɑ] is low; [ɪ e] are not. The individual results match this prediction very well. For thirty listeners, the value for pair [ɪɑ] is twice the value for [ɪe], and no listener had an [ɪe] value that was greater than his [ɪɑ] score.

Another good example is the pair [ɑ ɑʳ]. Of the 28 listeners that heard the [ɑʳ] vowel, 22 judged the [ɑ ɑʳ] pair to be most similar, no matter what the third member of the triad, resulting in a zero in the vowel matrix. Four listeners had a 1 for the pair, and the two remaining listeners had a value of 2. Listeners' comments confirmed the experimenter's impression that the [ɑʳ] vowel was felt to be different from [ɑ] only in "voice quality," not in phonetic quality. No listeners were accustomed to using the criterion of retroflexion to keep two low vowels apart.

There were some easily observable language-specific effects, as well. For example, the eight English responses to the pair [y ø] were 0 0 0 0 2 2 5 6, where the German responses were 3 5 5 6 7 8 11. There is some overlap in the range of values, but the general trend is clear: English listeners, not being required to discriminate between [y] and [ø], or even to use them at all, find them less distant than the German group does. The existence of a contrast in German between [y] and [ø] not surprisingly causes native listeners to regard the two sounds as somewhat more distant. A similar phenomenon occurs with the Thai, Turkish and German responses to [ʌ ɑ]. This pair represents a surface contrast in Thai, and five Thai listeners rate [ʌ ɑ] with relatively large values of 3 3 9 11 12. (For listener T6 the difference is small.) Since the German and Turkish subjects do not use these two sounds contrastively, we might expect the German and Turkish listeners to judge the vowels [ʌ ɑ] as very close, as they in fact do: No German or Turkish listener had a value greater than 2 for that pair. However, the English listeners who also contrast [ʌ] and [ɑ], had judgments of 0 0 1 2 2 5 6 -- barely larger than the Turkish and German values and consider-

ably smaller than the Thai values. The discrepancy between the English and Thai results may be resolved by referring to Figure 3.2, which shows that Thai [α] and [ʌ] are farther apart on an F1-F2 plot than are English [α] and [ʌ]. It seems as though the two groups of listeners were interpreting the [ʌ] stimulus as different vowels. Subjects apparently identified the [ʌ] as a native phonemic /ʌ/, then proceeded to judge it based on an internalized concept of that phoneme.

Further inspection of the raw data would reveal additional suggestive patterns of values, but it is difficult to evaluate them. After all, the MD scaling is the tool chosen for the systematic analysis of the data matrices, and at this point it is sufficient to notice that the responses are generally interpretable. It would be of some interest, though, to explore the results summed by language (Appendix I). Instead of looking at the matrices themselves, it will be more convenient to use a hierarchical clustering analysis for that purpose. To this end the following section is included.

4.3 Hierarchical clustering analysis of vowel distances, grouped by language

Hierarchical clustering (Johnson 1967), hereafter HC, reduces a set of distances to a set of nested clusters; in the algorithm responsible for the results presented here, a cluster is defined by the maximum distance between any two objects in the cluster. This description fits the "diameter" or "maximum" clustering algorithm. Another type is the "connectedness" or "minimum" technique, in which the members of a cluster are such that no two of them are separated by more than a criterion distance. This tends to form more chain-like clusters, where the diameter method finds more tightly-knit groups. Results were obtained using both algorithms; only the "diameter" method results are shown.

Figure 4.3 is an illustration of how a set of points on a plane, labelled 1-5, are formed into groups by the "diameter" clustering process. The pairs 1, 2 and 4, 5 each form tight clusters with small diameters, and point 3 forms a larger cluster with points 4 and 5. In two dimensions, of course, it is an easy matter to encircle clusters by sight. But for data whose dimensionality is greater than 2, or not known, the discovery of clusters is best done by computer.

HC was applied to each of the matrices constituting the summed vowel distances for each language. The results are drawn as Figures 4.4 - 4.8. Each figure is essentially a mobile in structure; each crossbar can revolve freely on

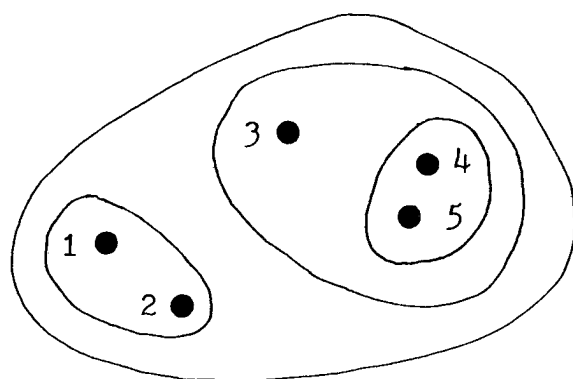


Figure 4.3: An arbitrary set of points arranged into clusters by hierarchical clustering analysis

its string, so long as the connections remain intact. A cluster can be defined as any set of vowels that ultimately descend from a single crossbar. With Figure 4.4, for example, the English results, $[u \ o]$ form a cluster, but $[u \ o \ y]$ only form a cluster insofar as $[\lambda]$ and $[\emptyset]$ are included as well. The farther one ascends up the mobile, the greater the distance between extreme points in the cluster. Double lines extend to vowels which are identical to or very similar to native vowels.

The vertical scale is a normalized expression for the diameter of the clusters formed. The diameter of the whole set, defined as the largest distance between any two vowels, was set equal to 1.0 for convenience; the smallest distance, in this case $[a \ a']$, was regarded as zero, and all other distances adjusted accordingly. The choice of zero as the smallest distance is also for convenience. Since the values in the summed matrices are ultimately based on triadic comparisons, the values are those of an additive scale; zero can be moved about freely. Actually, in the case of each language, the vowels making up the smallest cluster were in fact felt to be identical in phonetic quality, a result obviously compatible with the idea of "zero distance."

Each branch of the vowel mobiles is labelled with a distinctive feature, if possible. The clustering method itself dictates that the features be binary, since each cluster is formed by combining two smaller clusters (or individual points). The abbreviations used are as follows:

R = round
 L = low
 H = high
 B = back
 Ret = retroflex
 P = palatal

4.31 Labelling clusters according to distinctive features

It is striking that so many branches can be fit with distinctive feature labels with no difficulty. It is very gratifying to see that this is so at the lowest branches, which terminate with single vowels. With the exception of branchings involving [ɤ] and in one case [ʌ], which will be taken up shortly, each lowest branch can be labelled with a single feature. There is no a priori theoretical reason why this should be the case. Why should not [ɑ] and [e], with opposite values for both back and low, form a minimal cluster? The same possibility exists for even more remote pairs such as [i ɑ], with high, low, and back specified differently, or worse, [y ɑ], with a fourth feature, rounding, distinguishing the vowels further.

But in fact, almost all of the minimal clusters require only one distinctive feature to distinguish the member vowels. [y ø], [u o], and [i e] involve height (Figure 4.4) as the relevant feature. One pair, [i y], found in Figure 4.6, uses rounding. In the pairs [æ e] (Figure 4.5 and 4.6) and [æ ɑ] (Figure 4.8), the members are distinguished by back and low, respectively. Retroflexion is the distinguishing feature for [ɑ ɤ], [ɤ ø], and [ɤ œ]. [ɑ ɤ] is found in all figures except Figure 4.8; [ɤ ø] appears in Figure 4.7. In the case of [ɤ œ], found in Figure 4.8, Swedish [ɤ] is mislabelled as [+low]. This simply reflects the fact that Swedish subjects found non-native [ɤ] to be more similar to native low [œ] than native mid [ø].

The unlabelled [ɤ] vowel, and [ʌ] in the case of English, do not weaken the above conclusion. [ʌ] is easily labelled as -low, relative to [ɑ], which is [+low]; these labels appear on Figures 4.5, 4.6, 4.7 -- the German, Thai, and Turkish solutions. The same clustering appears in the English solution, but the label +low was already applied to the whole group [æ ʌ ɑ ɤ] in contrast with [-low] [ɤ e i]. [ʌ] is still relatively higher than [ɑ], but the feature labels are not fine enough to capture that smaller relative difference. Perhaps a label of "central" should be used for [ʌ] in this case. [ʌ] and [ɑ] are not separated by too many features; rather, they are so close acoustically that the features are too gross to distinguish them effectively.

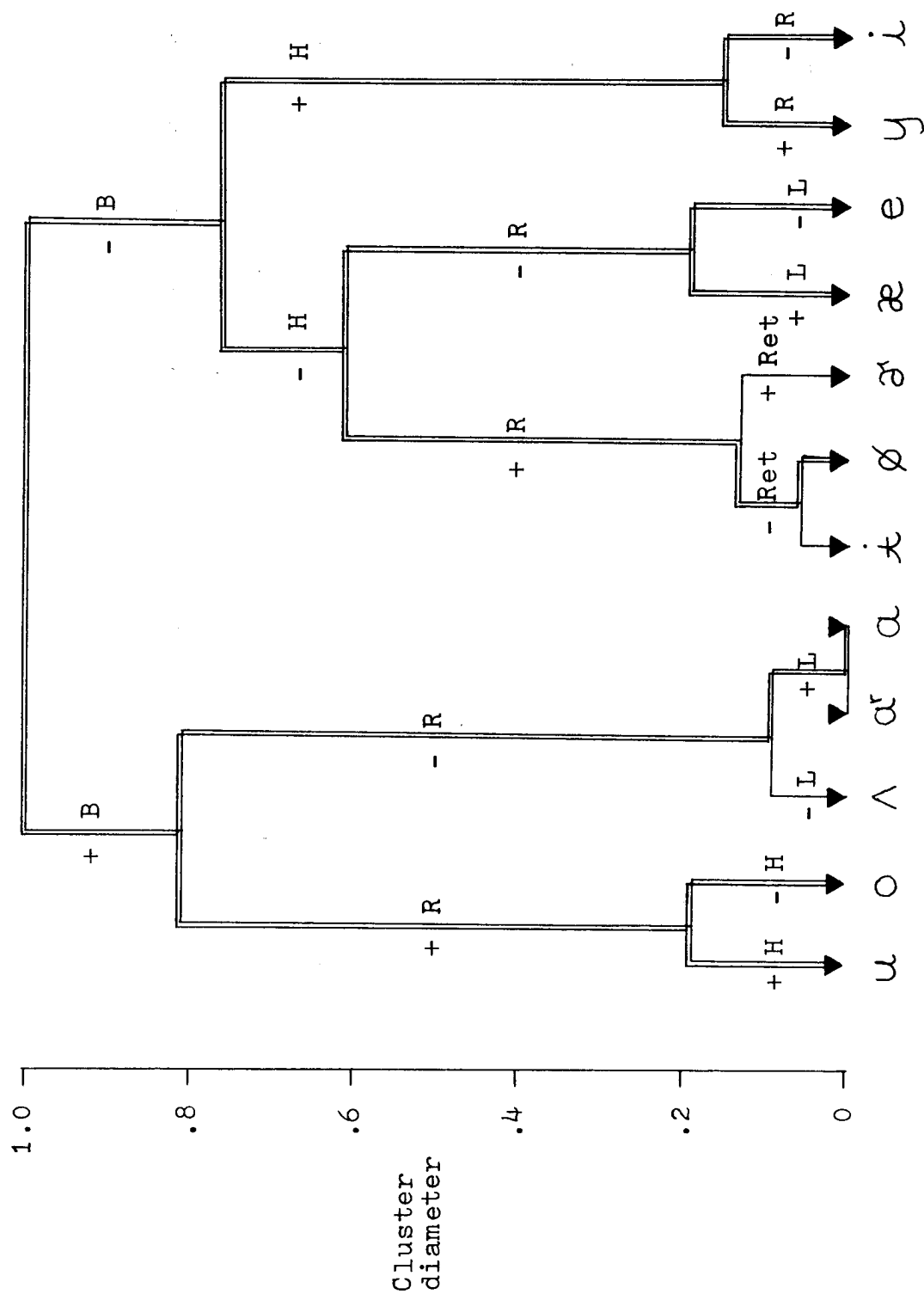


Figure 4.5: Hierarchical clustering for 7 German listeners.

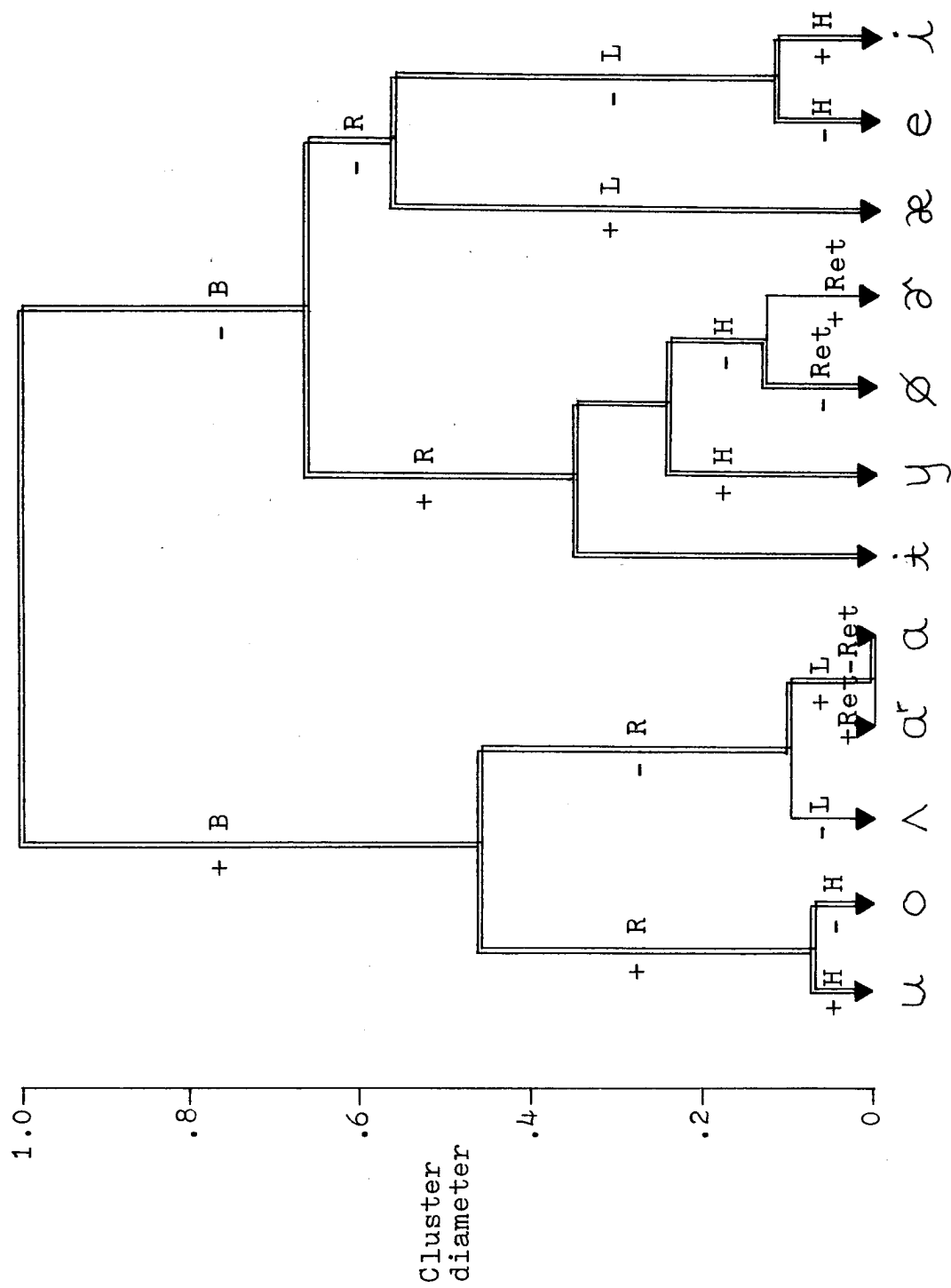


Figure 4.7: Hierarchical clustering for 7 Turkish listeners.

The vowel [ɨ̥] remains more difficult to deal with. Notice first that it is apparently mislabelled as [+round] in every diagram. All that can be said at this point is that the [ɨ̥] stimulus vowel, pronounced with no lip rounding, was grouped with the rounded vowels. This phenomenon will be discussed further when the MD-scaling analyses are presented. Also, [ɨ̥] remains unlabelled at every terminal branch. [+round] could have been used at the lowest level to distinguish [ɨ̥] from [ø] or [y] or [ɤ] as the case may be, at the expense of an obtuse description "rounding and [ɨ̥]" where the feature round now appears. Nothing would have been gained by that approach. Notice that [ɨ̥] appears as a member of a minimal cluster only in the case of Thai, German and Swedish (Figures 4.5, 4.6, 4.8), each time with [ø].

A glance at Figure 3.3 will quickly verify that [ɨ̥] and [ø] are very close in the acoustic plane of formants 1 and 2, and might well be heard as the same vowel by a listener who knew only one of them natively. This is clearly the case with Swedish and German. Although Thai has a vowel transcribed as [ɨ̥], none of the Thai listeners identified the stimulus [ɨ̥] with the Thai /ɨ̥/ vowel (hence the wistful dotted line on Figure 4.6). The former was apparently too far fronted. It should be clear that the inability to attach satisfactory labels to all the branches leading to [ɨ̥] does not vitiate the claim that the lowest branches represent very small differences in terms of distinctive features.

In the case of two languages, Thai and Swedish, some of the largest clusters cannot be described according to single features. The Thai results (Figure 4.6) show the two largest clusters to be [ʌ ɑ̃ ɑ æ e] and [ɨ̥ ø ɤ u o y i]. The feature Round almost fits, but for the inclusion of [ɨ̥] among rounded vowels (not counting the labelling of [ɨ̥] as "rounded" as an error). The two branches from the right-most group also define no feature in the Jakobsonian tradition. "Palatal," to set off [y i] from [ɨ̥ ø ɤ u o] is a convenient label, but not very satisfying. If only clearly native vowels are considered, the tree has the structure

$$((\wedge \alpha) (\ae e)) ((u o) i)$$

with branches replaced by bracketing. Now there is no need to be concerned with distinguishing [ɨ̥ y]. But the fact remains that [ɨ̥] groups with rounded vowels, and the initial division of the vowel cloud into two groups is not easy to understand.

The difficulty in labelling the topmost Thai branching is curiously symmetrical with the problem in Swedish: [y] here groups with unrounded vowels [æ ɑ e i] rather than with the

other rounded vowels (c.f. Figure 4.8). Considering only native vowels provides no solution, since the [y] stimulus was regarded by the subjects as a good example of Swedish [y]. The crux of the matter is that there is another Swedish vowel which is round, high, and non-back, namely [ʏ], and which occupies the position in the tree one would expect [y] to have. The fact that [y] is more similar to [ɪ] in tongue position ([ʏ] is more central) is some small justification for the grouping of [y] with the front vowels [i e], even though no accepted distinctive feature applies adequately to [æ a y e i].

One last labelling difficulty is the specification of rounding for the vowel [ɤ]. In the German, Thai, Turkish and Swedish results, [ɤ] groups with rounded vowels and is hence labelled [+round]. Only for English listeners did [ɤ] group with unrounded vowels. It is plausible that the retroflex quality of [ɤ] was an overriding cue for the English listeners, in spite of the very real rounding of [ɤ].

4.32 Relative length of branches and inter-language differences

As has been shown, the HC analysis finds vowels clustering in reasonable accord with distinctive feature specifications. The following discussion examines some of the differences among the results for the five languages.

The first point to be made is that if no contrast between two acoustically similar vowels exists in a language, native listeners may hear them as very similar if not identical, or, in terms of the clustering analysis, will place them in clusters of relatively small diameter. The vowels [ɑ] and [ʌ], mentioned in Section 4.22, demonstrate this. German and Turkish listeners, who have no [ɑ ʌ] contrast, found these to be clustered tightly together; Thai, however, uses both [ɑ] and [ʌ], and we find that the Thai listeners heard [ɑ] and [ʌ] to be much more dissimilar, hence the unusually large diameter for cluster [ʌ ɑ ɑʳ] on Figure 4.6. The English [ʌ ɑ ɑʳ] cluster is intermediate in size relative to the Thai and the German-Turkish extremes. This result can be accounted for by considering the difference between the English and Thai formant frequencies for [ʌ], as mentioned above (Section 4.22).

A second example is the vowel [ɨ]. German and Swedish, which have no [ɨ], show a very small distance between it and another vowel, namely [ø]. But for Turkish, [ɨ] forms only a loose cluster with the next smallest cluster [y ø ɤ]. Turkish has all three of [ɨ y ø] in the language, which can be taken as the reason for the relatively large distances among them.

The interpretation is not so simple for the remaining two languages. For Thai (Figure 4.6), if the listeners' comments are ignored and [ɛ̃] is regarded as a Thai vowel, then the closeness of [ɛ̃] and [ø] can be accounted for in that [ø] is not a native Thai vowel and was heard as similar to [ɛ̃]. [ɛ̃] is not native Thai, either, and the smallest native cluster which includes [ɛ̃] has a diameter of over .7. This interpretation makes the contrast of [ɛ̃] with other native vowels even greater than with Turkish (Figure 4.7). The alternative interpretation is that the [ɛ̃] stimulus is in fact not native-like, and the entire [ɛ̃ ø ɛ̃] branch is irrelevant to the Thais' perception of native sounds. In neither case is any violence done to the claim that a contrast implies greater perceived distances.

For the English listeners (Figure 4.4), the distance from [ɛ̃] to any other vowel clusters was greater than three pairs of native contrasts: [u o], [ɑ ʌ], and [e ɪ]. Perhaps [ɛ̃] was being perceived as a native vowel. One often hears [ɛ̃] as an allophone of [ɪ] in the environment of dentals, such as [sɛ̃stɛ̃] "sister," or as an allophone of [ʊ], as in [gɛ̃d] "good." But the vowel stimuli were long and formed open syllables, a distinctly non-English environment in which to find [ɛ̃]. In fact, the English listeners remarked informally that the vowel sounded peculiar and non-native. The length of the [ɛ̃] branch suggests that [ɛ̃], although non-native, was heard as considerably different from the other two non-native vowels [y ø], with which it eventually clustered. While it is true that no contrast exists in English between any two of [ɛ̃ y ø], [ɛ̃] at least seems to be sufficiently different acoustically from the other two to preclude a very tight cluster. English listeners must have been attentive to something other than F1 and F2; otherwise, [ø ɛ̃] would have been much more tightly grouped together. This remark goes for the Turkish listeners as well.

A final example is the vowel [ɛ̃], which for all languages but English forms tight clusters with [ɛ̃] and [ø]. [ɛ̃] for English listeners is distant from the [e ɪ] group, and clusters with it only remotely. Only in English can [ɛ̃] be regarded as a vowel phoneme.

4.33 Branch length as indication of tightness of clustering

Another kind of interlanguage comparison stems from the fact that the analyses show to what extent the vowels group into tight but widely separated clusters or are distributed evenly over the vowel space. The former case is best represented by the German results (Figure 4.5), which show five groups, namely [u o], [ʌ ɑ ɔ], [ɛ̃ ø ɛ̃], [æ e], and [y ɪ], whose diameters are about .21 or smaller. The next largest

cluster has a diameter of about .6, loosely joining [ɤ ø ʌ] and [æ e] into a larger group. The connection between [u o] and [ʌ ɑ ɔ] is even more tenuous.

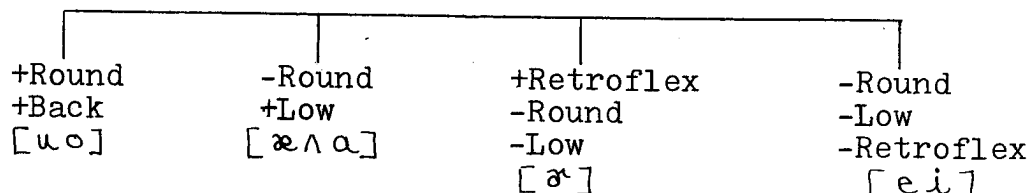
With this sort of structure, it is reasonable to try to interpret the tree as a hierarchy of feature decisions made in ultimately identifying a vowel. Long branches represent large distances between subclusters in the vowel space. These unfilled distances between clusters suggest a convenient initial perceptual classification of vowels. Considering Figure 4.5 once again, it is extremely plausible that a vowel be labelled [+back +round] before being judged according to the finer category of vowel height. On the right side of the tree, the vowels [y ɨ] are probably classified according to backness and height before being distinguished further. In this case, the "fine" distinction is made by rounding, not height.

Of the five test languages, Swedish lies at the opposite extreme. Here there is no sharp division of vowels into widely separated clusters; instead, vowels are spaced more evenly, as evidenced by the wide range of cluster diameters (c.f. Figure 4.8). With this sort of structure it is very difficult to make claims as to perceptual strategies, since there are few obvious cuts setting off one vowel group from another. The Thai results (Figure 4.6) behave in much the same way. Hypotheses about feature hierarchies for Swedish and Thai are further complicated in that the topmost branches in each case cannot be satisfactorily labelled with distinctive features.

Perhaps new, more reliable data would not show these difficulties. But six of the Swedish matrices -- including the three most "unreliable" ones (Table 4.2) -- showed a significant deviation from the random pattern (Table 4.3). They were clearly responding in a reasonable way. In addition, the Thai listeners performed perfectly adequately, based on the reliability and validity tests. The difficulty may be that the Thai and Swedish listeners behave acceptably as individuals, but not as a group. No tests have been performed to assess whether the individual results from within one language group were more similar to one another than were the results from within another language group.

The Turkish results (Figure 4.7) are much like Swedish and Thai, with respect to the differences of the clusters in the space. The remarks in the previous paragraph apply equally well to Turkish.

English (Figure 4.4) lies between the German and Swedish extremes, especially if one considers only native vowels. If the branches are terminated at a value of .15 instead of 0, then the lowest native branches are all very short, namely, [u o], [ʌ ɑ], [e i]. Also, the [+round -back] branch disappears entirely, leaving no branching at all above the cluster [u o]. This leaves the following simplified structure:



These four branches are relatively long compared with the branches of their constituent vowels. Here, as with the German results, a plausible set of hierarchical decisions can be seen which lead toward the identification of a vowel.

Actually, one can postulate a simple acoustic cue for each of the four groups (this was not as easy to do for German), as is suggested by the flow chart on Figure 4.9. The only acoustic information that needs to be extracted is the value of each of the first three formants. (Refer to Figure 3.3 for the formant frequencies of the stimuli.) The "yes" answers to the questions on the flow chart are mutually exclusive. That is, they might as well be answered by parallel processors instead of in sequence. The word "sufficiently," of course, is bound up with the larger issue of how a listener quickly adapts to a speaker's formant patterns, and this research has no light to shed on that problem.

The flow chart can only be taken as a hypothesis for future verification, because the stimuli did not include all the English phonemes, and because some of the stimuli, namely [e u o], are not characteristically monophthongs in English. The high front and back glides these three sounds exhibit are certain to be used as perceptual cues, although they cannot show up in the present analysis. Terbeek (1974) presents research in this direction.

4.34 Phonological properties

So far, the discussion has focussed on the effect on listeners' perception of the presence or absence of vowels or vowel contrasts, that is, effects based on phonemic inventory. An interesting result that cannot be accounted for in terms of the phonemic inventory is the uppermost division

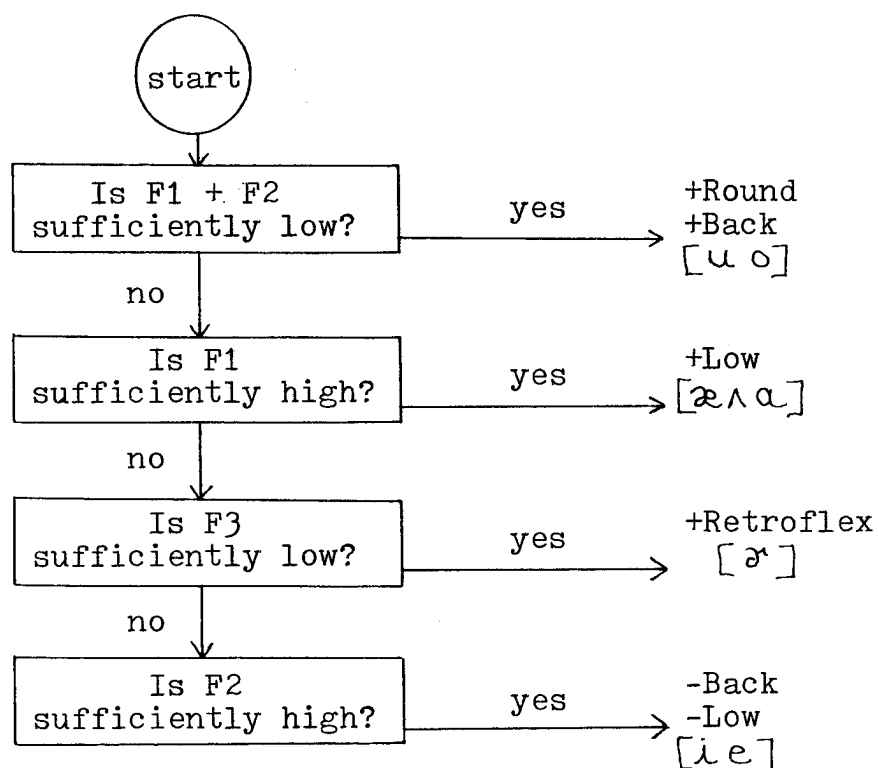


Figure 4.9: Hypothetical flowchart for perceptual sorting of vowels into major groups for English.

of the vowel space according to Backness for German and Turkish, and Rounding (with some discrepancies) for English, Thai and Swedish. Of course, it may be a chance result and require no explanation at all. But this is hardly the more interesting track to take. There is no way for a hypothesis to be defended or refuted if it has never been proposed at all.

What, then, can be hypothesized that causes the German and Turkish results to differ from the others? If Swedish were not included, an explanation based once again on the phonemic inventory might be attempted: German and Turkish have front rounded vowels; Thai and English do not. The fact that there are rounded vowels that differ in backness causes a heightened awareness of the feature Back. English and Thai listeners do not normally distinguish between [ɥ] and [ʊ], or [ø] and [o], and therefore do not have this awareness. Unfortunately for this hypothesis, the Swedish results disconfirm it directly. Swedish, too, had front rounded vowels, yet the Swedish listeners did not show this proposed special awareness.

A hypothesis worthy of attention is the following: Turkish and German show phonological relationships between pairs of sounds differing in Backness, and English, Thai, and Swedish lack them. German and Turkish listeners, it is proposed, are influenced in their judgments by the fact that back vowels are systematically related to front vowels. German possesses the umlaut relationship according to which back vowels are fronted in various situations, such as pluralization ([mʊtər] "mother"/[mʏtər] "mothers"), word derivation ([kam] "comb"/[kɛmən] "to comb"), verb inflection ([hat] "(he) has"/[hɛtə] "if (he) had"), and the formation of diminutives ([zɔn] "son"/[zʊnçən] "cute little son"). Now, only the last category is productive, but it seems very likely that German speakers have come to be aware of [aʊ] as sounds related in a functional way to [ɛ ø y], respectively. Backness, in German, does not merely distinguish phonemes; it reflects in addition a large number of sound changes that have grammatical significance.

In Turkish, bound suffixes on nouns must harmonize with the root verb according to backness, as in the following examples:

zɪl	"bell"	zɪllər	"bells"
tʃøl	"desert"	tʃøllər	"deserts"
dɪʃ	"tooth"	dɪʃlər	"teeth"
bɑʃ	"head"	bɑʃlar	"heads"
dost	"friend"	dostlar	"friends"
kuʃ	"bird"	kuʃlar	"birds"
ev	"house"	evde	"in the house"
gøz	"eye"	gøzde	"in the eye"
kɔl	"arm"	kɔlda	"in the arm"
kuʃ	"bird"	kuʃta	"in the bird"

The plural morpheme [l_r] has [e] or [a] for its vowel, depending upon whether the root vowel is front or back, respectively. The same holds for the vowel in the suffix "in." Some suffixes have high vowels, in which case the vowel must harmonize for both backness and rounding:

baş	"head"	başım	"my head"
dost	"friend"	dostum	"my friend"
diş	"tooth"	dişim	"my tooth"
göz	"eye"	gözüm	"my eye"

It is hardly arguable that these are not psychologically real relationships to Turkish speakers.

The remaining three languages simply do not make use of Back-Front relationships in this way, regardless of whether they use the dimension [$\pm$ Back] to distinguish phonemes. It is an abstract property of the morphology of German and Turkish, then, that causes natives of those languages to give backness a perceptually prominent place. This phenomenon will re-emerge later in the MD scaling analyses, and the hypothesis presented here will be defended further.

4.35 Limitations of HC analysis

HC analysis certainly affords a convenient overview of the properties of the data. Furthermore, it provides evidence that the number of features two sounds have in common is highly related to listeners' concept of similarity for those sounds. It was also possible to propose hypotheses about perceptual strategies and about the influence of abstract phonological properties on judgments. But HC analysis is totally inadequate with respect to this crucial point: it cannot reveal any details as to the structure of a cloud of points beyond reporting which points form various size groups with which other points. Three looming questions still remain: 1) How many dimensions underlie the vowel space? 2) What is the configuration of vowels in it? 3) How do the perceptual dimensions relate to linguistic or acoustic dimensions? Chapters 5 and 6 are devoted to the answers to these questions, supplied by the MD scaling analyses.

4.4 Preliminary individual scaling to obtain additive constants

The raw data matrices are not in a form acceptable to PARAFAC and PARAFAC2. Those programs require scalar products matrices as input, not distance tables. The conversion itself, described by Torgerson (1958, p. 258) is a trivial one, with the following minor difficulty: the scalar products conversion is sensitive to the location of zero on the distance scales. From the description of

the data recording procedure in Section 3.6, it should be clear that the cells in the vowel matrices obey an additive relationship, not a ratio relationship, since the location of zero on the scale is arbitrary. A good guess must be made as to where zero ought to fall.

The simplest guess would be to add or subtract a constant from every cell so that the smallest off-diagonal value is 1. It is logical enough to define the distance from a vowel to itself as zero and require all other values to be greater than zero. But perhaps the matrices as a whole are best interpreted as having no cell at all that is particularly close to zero. Perhaps the best fitting analysis could be made if the values ranged not from 1 to 21, but from 8 to 28.

The most convenient established method for estimating the additive constant is given in Torgerson (1958). He assumes that in a reasonably large cloud of points in some unknown number of dimensions, three points ought to be expected to be on a straight line. His additive constant is the smallest value which locates any three points on a straight line. That is, it is the smallest K such that $(ab + K) + (bc + K) = (ac + K)$, where ab , bc , and ac are the distances from a to b , b to c , and a to c , respectively.

This method is simple, but it is unfortunately not directly related to the properties of the data at hand. In fact, it is easy to imagine spaces in which this requirement of three points in a line should not be met. What if a 3-dimensional space were filled by eight points, corners of an imaginary cube? Torgerson's procedure would collapse the cube to some plane, distorting the true relationships. This approach is wrong to the extent that listeners make binary distinctions in vowel discriminations, each binary distinction representing a distinctive feature.

A more theoretically interesting approach is that taken by Cooper (1970) in his COSCAL program. COSCAL takes the additive constant as a parameter to be optimized during the course of an MD scaling analysis. The "correct" additive constant is one which produces the best prediction of distances from the coordinates in the space (which are derived simultaneously by the program). Here there is no a priori assumption about what effect the additive constant ought to have on distances, except that the new distances (original data plus K) lead to the best-fitting set of coordinates.

COSCAL was applied to each person's vowel matrix, and solutions were obtained in 1 through 5 dimensions. For the Swedish data, 6-dimension solutions were run as well. Later it became apparent that 6 and 7 dimension PARAFAC solutions would be necessary, and 6 and 7 dimension constants were extrapolated linearly from the previously obtained COSCAL constants. One PARAFAC analysis of Swedish ran to 8 dimensions, and appropriate constants were estimated as before.

The actual scaling solutions for each listener will not be presented. Deciding upon an appropriate number of dimensions 35 times and rotating each solution to a plausible set of axes would be far too tedious a project to justify the highly tenuous results that would obtain. The additive constants (K) and variance accounted for (FIT) for COSCAL solutions for each listener are presented as Table 4.4. Notice first that there is a different additive constant for each listener for each size solution. The more dimensions, the higher the constant. As the program is given the freedom of more and more dimensions, all distances become larger by a constant amount, decreasing the ratios between distances. The values given are on the same scale as the original input matrices, so that an additive constant of 10 will change the raw values 3 and 8 into 13 and 18, decreasing the ratio of the two distances from 2.7 to 1.3.

Negative constants appear in all 1-dimension solutions, and in some 2-dimensional spaces. For many matrix values that were small to begin with, negative constants would seem to be impossible, since they combine with the given values to produce negative numbers as distances. Although conceptually meaningless, mathematically this is no difficulty. The first step in transforming the distances to scalar products is to square each distance, making the sign irrelevant. The real problem is that if the negative constants are large enough, small values in the matrix will be reversed in rank order.

Data matrix E7 provides a good illustration. Subject E7 has an additive constant of -4.41 in one dimension, meaning that values 0, 1, 2, 3, 4 become -4.4, -3.4, -2.4, -1.4, and -0.4, respectively. When these are squared, the smallest original distance takes on the largest value, and vice versa. The vowel pair [ø y], originally associated with a zero, becomes the equivalent of 4.4, whereas [ɨ ø], at first 4, now becomes effectively 0.4. What is worse, larger distances are brought down into this confusion. [ʌ æ], for example, having an original value of 8, takes on distance 3.6. A situation in which 0 and 8 become nearly identical, and in which 4 takes on a value less than either or them, is an intolerable distortion of the data.

Table 4.4: Additive constants and variance accounted for (FIT), according to COSCAL analyses of 35 listeners arranged by native language. Where no fit value is given, K has been extrapolated from previous solutions for that person.

LISTENER	1 DIMENSION		2 DIMENSIONS		3 DIMENSIONS		4 DIMENSIONS	
	<u>K</u>	<u>Fit</u>	<u>K</u>	<u>Fit</u>	<u>K</u>	<u>Fit</u>	<u>K</u>	<u>Fit</u>
<u>English</u>								
E1	-3.40	.783	-0.48	.899	1.42	.925	3.65	.946
E2	-5.32	.470	-0.28	.863	2.73	.943	4.38	.959
E3	-4.87	.532	0.70	.867	2.70	.919	4.59	.943
E4	-4.50	.588	-0.12	.857	2.24	.912	4.55	.945
E5	-5.74	.487	-1.86	.750	2.17	.903	4.55	.931
E6	-5.99	.523	-3.16	.688	0.51	.847	3.24	.904
E7	-4.41	.439	2.49	.848	5.74	.920	7.95	.946
E8	-4.74	.467	0.96	.848	4.05	.910	6.56	.936
<u>German</u>								
G1	-3.67	.637	1.93	.891	4.94	.931	8.17	.963
G2	-3.69	.697	1.19	.856	4.48	.931	7.17	.955
G3	-4.06	.637	0.40	.915	2.90	.945	4.66	.963
G4	-3.58	.656	0.67	.787	5.70	.886	8.58	.908
G5	-4.29	.541	0.66	.795	5.40	.915	8.81	.945
G6	-4.48	.506	0.93	.869	4.07	.925	6.82	.945
G7	-3.83	.711	0.32	.912	2.00	.937	3.68	.960
<u>Thai</u>								
T1	-4.95	.564	-0.01	.872	1.68	.913	4.15	.944
T2	-5.23	.601	-2.09	.822	0.67	.886	3.59	.933
T3	-5.58	.619	-2.96	.812	-0.54	.879	2.22	.933
T4	-5.17	.535	-0.20	.859	1.84	.892	4.78	.928
T5	-4.55	.628	-1.41	.768	1.62	.848	4.76	.887
T6	-5.25	.609	-2.01	.802	0.64	.887	2.47	.917
<u>Turkish</u>								
K1	-3.29	.732	0.61	.867	3.92	.910	7.40	.934
K2	-3.96	.586	1.78	.907	4.55	.955	6.51	.972
K3	-2.48	.890	-0.16	.929	2.71	.956	5.36	.967
K4	-2.98	.662	2.93	.901	5.92	.932	9.24	.958
K5	-5.33	.462	-0.69	.852	2.39	.916	4.78	.943
K6	-4.51	.671	-0.94	.813	2.92	.910	5.78	.938
K7	-3.70	*	0.33	.960	2.52	.971	4.49	.977
<u>Swedish</u>								
S1	-4.67	.668	-1.87	.814	0.92	.904	3.22	.942
S2	-7.20	.294	-3.23	.604	0.00	.719	2.82	.795
S3	-5.80	.303	-0.14	.677	5.22	.928	8.00	.954
S4	-5.98	.502	-2.59	.747	0.43	.858	3.03	.903
S5	-5.14	.620	-1.80	.779	1.48	.867	3.96	.921
S6	-5.31	.461	0.14	.787	3.85	.903	7.44	.945
S7	-4.42	.577	-0.49	.849	2.45	.922	4.48	.946

*COSCAL could find no satisfactory solution and did not give a meaningful fit for this problem.

Table 4.4, continued

LISTENER	5 DIMENSIONS		6 DIMENSIONS		7 DIMENSIONS	8 DIMENSIONS
	<u>K</u>	<u>Fit</u>	<u>K</u>	<u>Fit</u>	<u>K</u>	<u>K</u>
<u>English</u>						
E1	6.25	.962	8.7		11.2	13.7
E2	6.26	.972	8.8		10.9	11.0
E3	6.86	.957	9.0		11.2	13.4
E4	6.86	.967	9.2		11.4	13.8
E5	6.69	.947	9.2		11.5	13.9
E6	5.76	.941	8.6		11.2	14.0
E7	10.54	.960	12.8		15.3	17.8
E8	9.80	.965	12.5		15.4	18.4
<u>German</u>						
G1	11.11	.980	14.4		8.17	20.7
G2	9.59	.969	12.0		7.17	16.6
G3	7.56	.981	10.5		4.66	16.5
G4	12.86	.947	17.1		8.56	25.4
G5	12.31	.977	15.7		8.81	22.6
G6	9.62	.966	12.3		6.82	17.8
G7	5.38	.974	7.0		3.68	10.4
<u>Thai</u>						
T1	6.68	.963	9.3		4.15	14.5
T2	6.26	.961	9.1		3.59	14.9
T3	4.64	.966	7.1		2.22	12.3
T4	7.88	.957	10.8		4.78	16.8
T5	8.27	.918	11.7		4.76	18.7
T6	4.84	.943	7.4		2.47	12.3
<u>Turkish</u>						
K1	10.73	.958	14.1		17.4	20.8
K2	8.30	.981	10.0		11.7	13.4
K3	8.05	.980	10.7		13.2	16.0
K4	11.71	.971	14.5		17.4	20.3
K5	6.99	.959	8.9		10.8	12.8
K6	8.26	.958	10.4		12.6	14.8
K7	6.83	.984	8.8		11.0	13.2
<u>Swedish</u>						
S1	5.43	.964	7.93	.978	10.2	12.7
S2	6.18	.868	8.96	.902	12.0	15.1
S3	10.46	.975	12.63	.985	14.5	16.4
S4	6.13	.947	9.19	.975	12.3	15.4
S5	6.31	.943	8.77	.963	11.1	13.6
S6	10.32	.971	12.49	.980	14.2	16.0
S7	7.06	.964	9.95	.984	12.7	15.6

All the COSCAL runs resulting in large negative constants give very low fit values, which in and of itself shows those solutions to be inadequate. As if the low fit values are not enough, they can only be obtained by distorting the observed data.

In general, the two or three dimension space for each subject has attained a good fit, which is improved upon only slightly by extracting more dimensions. A few sample curves of fit against dimensionality are drawn on Figure 4.10 to illustrate this. For Subject G1, the 2-dimension solution is not improved significantly by extracting a third dimension. Subject E5 represents listeners whose data are best accounted for in a 3-dimensional space, not a 2-dimensional one. There are several cases, though, of a gradual improvement in fit with no clear elbow in the curve. Subject S2 is the worst example of this pattern. It is worth mentioning that the patterns of fit-by-dimensionality do not group listeners according to native language.

After the additive constants were found, the next step was to add them to the data matrices. The resulting absolute distance matrices were then converted to scalar products matrices, according to Torgerson's procedure. For the analyses involving 35 listeners and 10 vowels, discussed in Chapter 5, the rows and columns containing the excluded vowels [ʊʌæɐ] were deleted from the scalar products matrices already obtained.

Since different constants have been obtained for different dimensionalities, the data submitted to PARAFAC will depend upon the number of dimensions required. This approach allows each analysis the benefit of the doubt. When a solution is requested in n dimensions, the data used to derive that solution have been adjusted to conform to the expectation that there are n dimensions that in fact underlie the responses.

This is an appropriate place to ask how many dimensions PARAFAC might be expected to derive. Perhaps this method allows too many factors to be extracted. After all, it has been shown that most listeners' individual COSCAL results support two or three dimensions, and no more. Would a larger PARAFAC solution imply that spurious dimensions have been constructed out of noise? The final decision as to the correctness of a solution must, of course, be its interpretability. But it can also be argued on formal grounds that the data are indeed sufficient to support solutions in more than three dimensions.

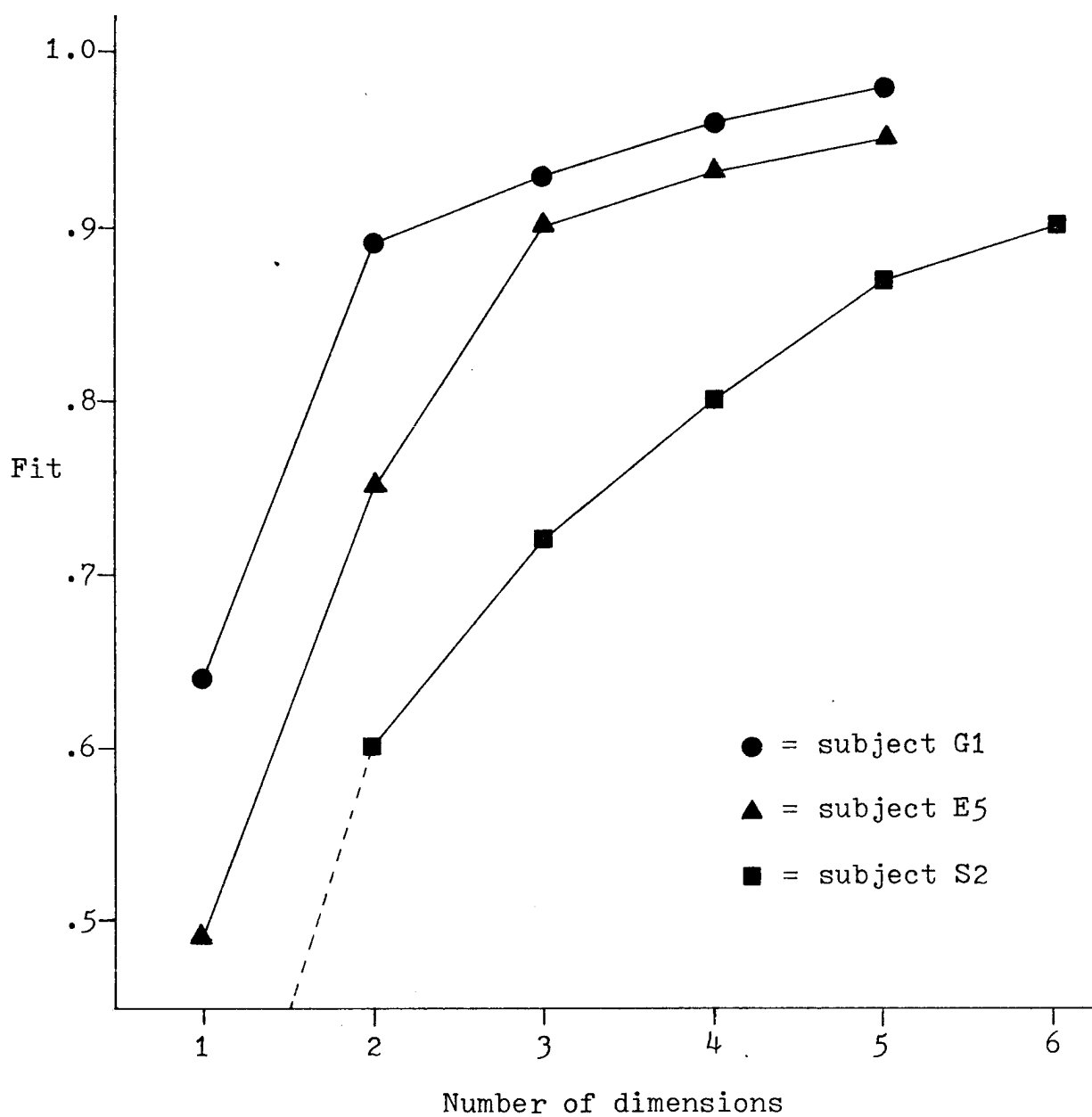


Figure 4.10: Sample curves of Fit against dimensionality from individual COSCAL analyses.

The argument begins with the admission that no one subject's data is robust enough to reveal the presence of more than a few dimensions. For a 1-dimensional COSCAL solution, there is a ratio of 5.5 active cells to each coordinate requested (66 independent cells in a 12 x 12 distance matrix, divided by 12 coordinates), which is a barely comfortable margin. For two dimensions, the ratio drops to 2.75; for 3, to 1.84. The closer the ratio gets to 1, the more the solution is buffeted about, so to speak, by noise in the data. With a ratio of 1, there is no reduction or distillation of the data, and thus no smoothing out of the noise.

But according to this criterion, the PARAFAC analyses promise to be much more robust than the COSCAL ones. A 3-dimensional PARAFAC solution for 10 vowels and 35 listeners would show a ratio of 11.7 cells (66 vowel pairs x 35 persons / 3 dimensions x 35 persons + 10 vowels) per output coordinate. A 6-dimensional PARAFAC solution for that data has a ratio of 5.8. Even the 1-dimensional individual COSCAL analyses are more affected by experimental error. When PARAFAC is applied to the data from each language separately, the worst input/output ratio for a good solution is 5.3. The PARAFAC analyses presented in the following two chapters are actually more reliable than the COSCAL ones, not less. (See appendix III for a more detailed discussion of the I/O ratio.)

Chapter 5

The 6-Dimension PARAFAC Solution for all 35 Listeners

5.0 This chapter is devoted to the PARAFAC analysis performed on the complete set of data from all 35 listeners. There are three tasks at hand: 1) to select the solution with the "correct" number of dimensions and justify this selection, 2) to interpret the vowel space of that solution, and 3) to interpret the pattern of personal weighting factors.

The solution presented as the correct one is the 6-dimensional space. Uniqueness of solution breaks down after six dimensions; the 6-dimensional solution can be interpreted, whereas the two 7-dimensional solutions in their entirety cannot be. But more detailed evidence for preferring that solution over others will be postponed. It is much easier to follow if the properties of the chosen space are described first. In Section 5.1, the six PARAFAC dimensions are discussed in terms of linguistic features. The possibility of re-interpreting the space as consisting of three curved scales is also considered. The interpretation of the space is continued in Section 5.2, where stepwise multiple linear regression analyses are used to investigate the acoustic bases for the dimensions. Special perceptual strategies are proposed to account for dimensions that do not seem to be based on linear combinations of acoustic variables. Not until Section 5.3 is the selection of the 6-D space defended. Some consequences of preferring solutions of fewer or more dimensions are discussed as well.

Section 5.4 deals with the personal weighting factors. These values bear directly on the issue of language-specific versus idiosyncratic importance attached to each of the perceptual dimensions. The pattern of weighting factors that emerged argues against the absolute universal hypothesis, in which the salience of the dimensions is the same from one language to another.

5.1 Distinctive feature interpretation of the 6-dimensional solution

The vowel space is plotted in Figures 5.1, 5.2, and 5.3. In each figure, the PARAFAC dimensions have been labelled according to the most appropriate distinctive features. But none of the three planes is completely filled, and it is possible in each case to draw a smooth curve that comes fairly close to all of the vowels on the plane. The curves themselves can be interpreted as distinctive features which are about as satisfactory as the original linear dimensions.

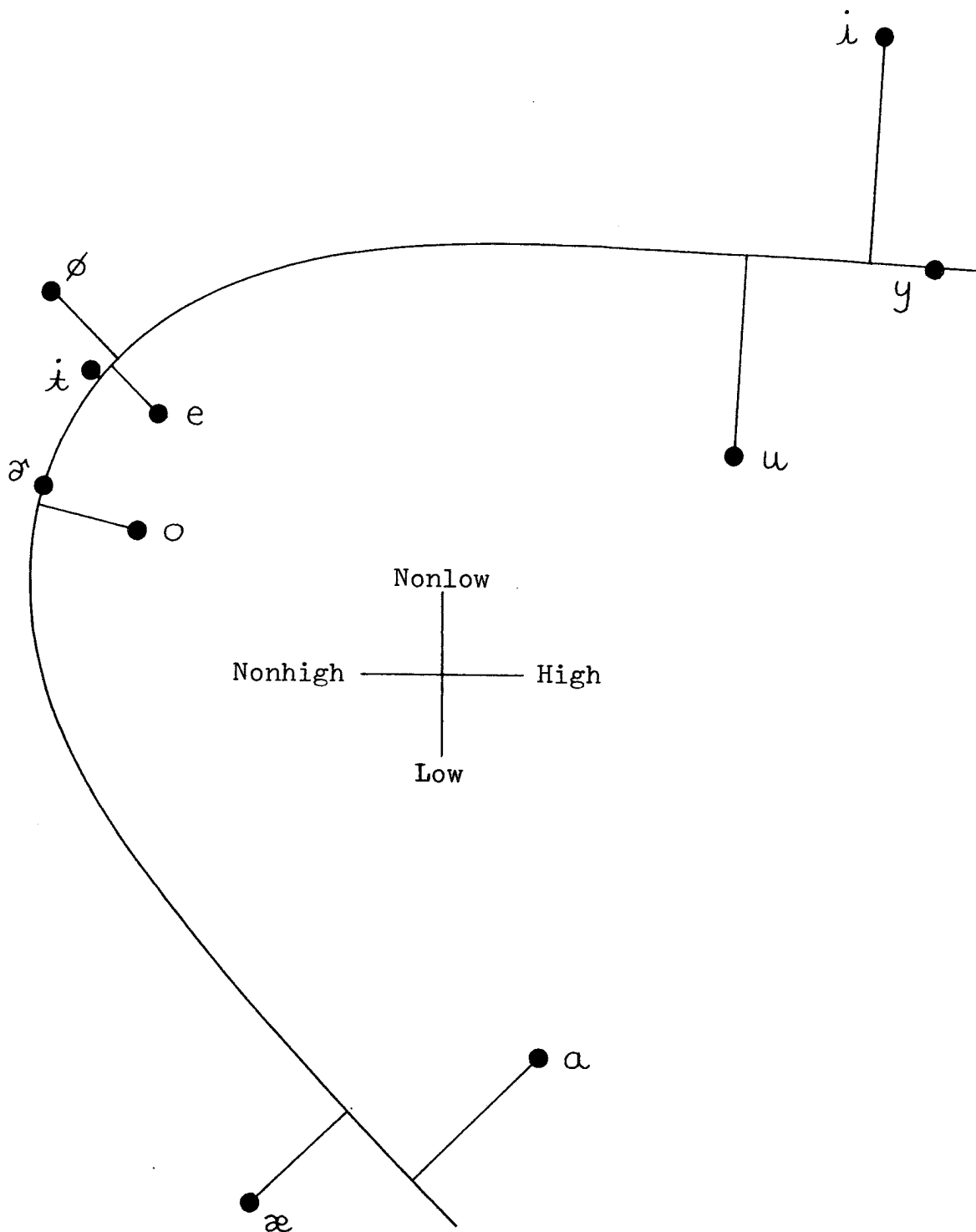


Figure 5.2: Low/Nonlow and High/Nonhigh from 6D solution, with curved line fitted by sight.

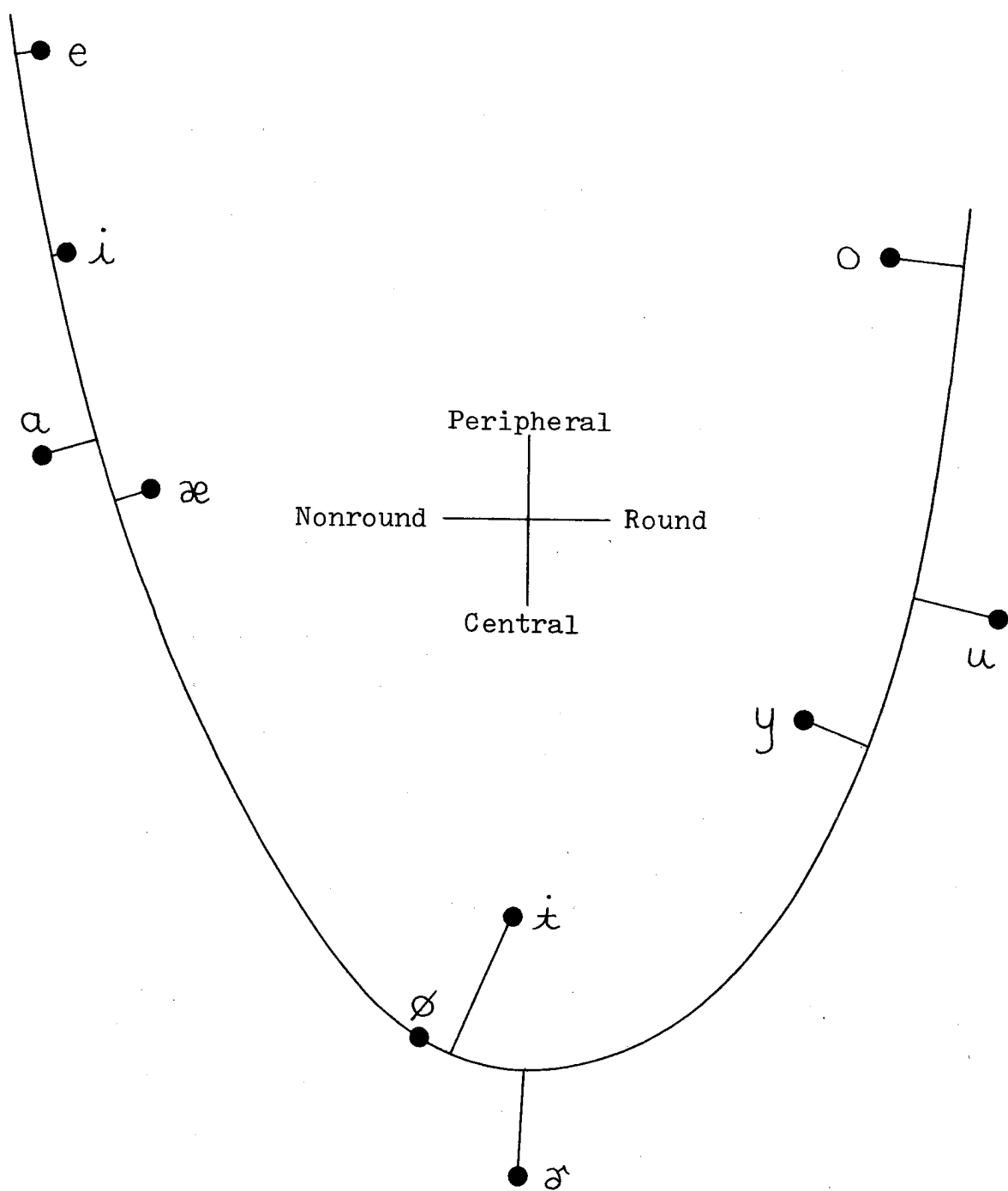


Figure 5.3: Round/Nonround and Peripheral/Central from the 6D solution, with curved line fitted by sight.

5.11 Backness

On Figure 5.1, the vertical dimension is labelled Back/Nonback1. This dimension separates the three back vowels [u o ɔ] from front [i e æ y ø]. Nonback [ɤ] groups with the front vowels. [ɤ] lies between the back and nonback clusters. The placement of [ɤ] as farther back than [ɤ] could be predicted from their F2 values (c.f. Figure 3.3). However, F2 on this scale does not account for many other vowel positions, even though it is the single acoustic variable most related to backness. The [ɑ] vowel is a good example. The [ɑ] stimulus is closer in F2 to [ɤ] than to [u] and [o], but on the Back/Nonback1 dimension, it groups with [u o], not with [ɤ].

The horizontal dimension on Figure 5.1 (Back/Nonback2) differs from the vertical one only in the placement of [ɤ] and [ɤ]. Here, [ɤ] groups tightly with the back vowels, well separated from the front cluster. [ɤ] appears at the extreme right-hand end of the scale, but it lies closer to the front vowels than does [ɤ].

The main function of this dimension -- to associate [ɤ] with back vowels -- is readily apparent, though the reason such a dimension should appear may not be. A very likely suggestion is that the Turkish listeners gave rise to Back/Nonback2. In Turkish, [ɤ] is phonologically a back vowel, since in rounding harmony it alternates with [u] and in backness harmony it alternates with [ɤ]. It seems unreasonable to predict that Back/Nonback2 would have appeared if the Turkish listeners had not been present. The discussion of personal weights in Section 5.4 and the single-language analyses in Chapter 6 support the idea that Back/Nonback2 is a Turkish dimension.

The fact that Figure 5.1 has much empty space prompts one to propose that there is a simple curved backness dimension, as drawn on the figure. The projections of the vowels onto the curve are taken to constitute the new curved scale; deviations from the curve are small and therefore disregarded. The curve comes about because the listeners' data generally reflect a distance from front vowels to back vowels which is smaller than the distance from front vowels to [ɤ] to back vowels. Unfortunately, [ɤ] appears as the "frontmost" vowel under such an interpretation. This might only be a case of apparent misnaming; the scale could be a perceptually real one for which there is no traditional label.

5.12 Height

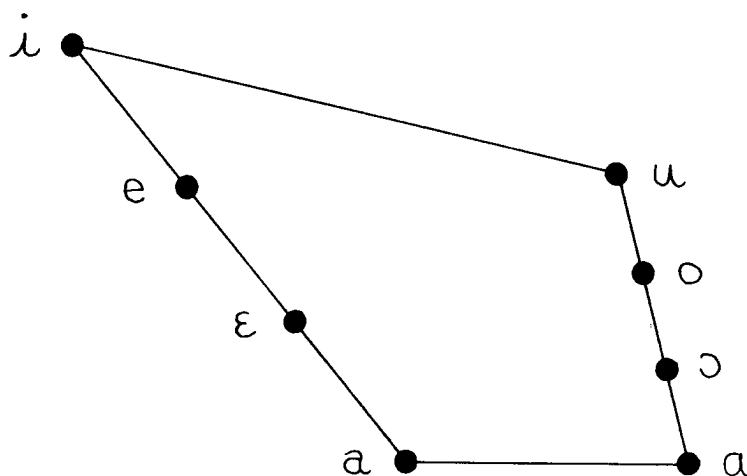
Of the two dimensions plotted on Figure 5.2, the most intuitively reasonable one is Low/Nonlow, plotted vertically.

[æ a] are separated considerably from the other eight vowels, all of which are mid or high. The high vowels [u] and [y] are mixed with the mid vowels, but that is of no importance if the purpose of the dimension is to allow a rough grouping into low and nonlow. The horizontal direction called High/Nonhigh does in fact separate the high vowels [i u y] from the others. The placement of [ɤ] among mid vowels is not regarded as a discrepancy, given that it is extremely similar to mid [ø] in its F1 and F2 values. Unfortunately, there is no large gap that can be defended as a plausible distinctive feature cut. What is worse, the two low vowels [æ a] appear in the middle of the scale, between high and mid.

This difficulty disappears if one regards the plane as a curved dimension of height, suggested by the curve drawn on Figure 5.2. High, mid, and low vowels are clearly separated from each other. However, the deviations from the curve, especially for [i], [u], and [a], are greater than those of the vowel positions on Figure 5.1 and cannot be so easily disregarded.

In accounting for the deviations from the proposed curve, it is observed that the back vowels, most notably [a] and [u], lie within the curve. The distance from [u] to [o] to [a] is therefore less than the distance from [i] to [e] to [æ], although in both cases the distance from high to low is smaller than the distance high-to-mid-to-low. There seems to be an interaction of height with backness: back vowels are not as perceptually different from each other as front vowels are.

It is interesting to compare this result with the cardinal vowel diagram preferred by Daniel Jones (1967). He comments explicitly that "it appears to be preferable, however, to adopt a form which shows that the distance [i a] is longer than the distance [a u] . . . [compared with charts that are symmetrical with respect to backness. D.T.] (footnote 9, p. 37)." The diagram itself is this:



Although Jones conceived of this as a stylized plot of the relative position of the highest point of the tongue, it appears to match perceptual data, as well (or perhaps "instead"). Ladefoged (1967) found that phoneticians tended to regard front cardinal vowels [i e ε a] as spaced farther apart than the back ones [ɑ ɔ ɒ u] (p. 99-100). The deviations from the height curve for the present data reflect this finding.

5.13 Rounding

The scale called Round/Nonround appears horizontally on Figure 5.3. The vowels form three fairly tight groups on this dimension, in spite of being spread out vertically: unrounded [e i ɛ a] to the left, [ø ɤ ɶ] in the middle, and rounded [y u ɔ] at the right. As for the middle group, [ø] is rounded, as is the [ɶ] stimulus. [ɤ] once again is found to be near [ø], presumably because it is so like [ø] in its first two formant frequencies. The middle group may be tentatively labelled "round," together with [y u ɔ].

The vertical direction on Figure 5.3 does not group vowels into obvious clusters. The fairly continuous pattern does not reflect any commonly used distinctive feature. The name Peripheral/Central comes from the observation that the three vowels at the bottom of the dimension, [ɤ ø ɶ], appear near the center of the F1-F2 space (Figure 3.3), while the remaining seven are clearly more peripheral. The label is certainly not an exact one. Why should [y] appear to be less peripheral than [i], for example? Or worse, why is [u] less peripheral than [ɔ]? The F1-F2 plane does not reflect such differences nor does it account for difference in vertical position for other pairs on Figure 5.3. No intuitively better term suggests itself, and the implication of a satisfying categorical division of the scale is wishful thinking.

Once again, it appears that one curved dimension will suffice as the interpretation of the plane of Figure 5.3. The deviations from the curve are slight and have no apparent pattern. Cutting the curve where there are gaps between vowels produces the same three clusters which appeared on the Round/Nonround scale itself: peripheral nonround [e i ɛ a], central [ø ɤ ɶ], and peripheral round [y u ɔ].

5.14 Overview of the space of three curved dimensions

In order to see how the proposed curves of height, backness and rounding behave with respect to each other, the vowel positions along each curve were measured and regarded as coordinates in an ordinary 3-dimensional space. Coordinates appear in Appendix II; 2-dimensional plots are presented in Figure 5.4. It must be remembered that the back vowel devia-

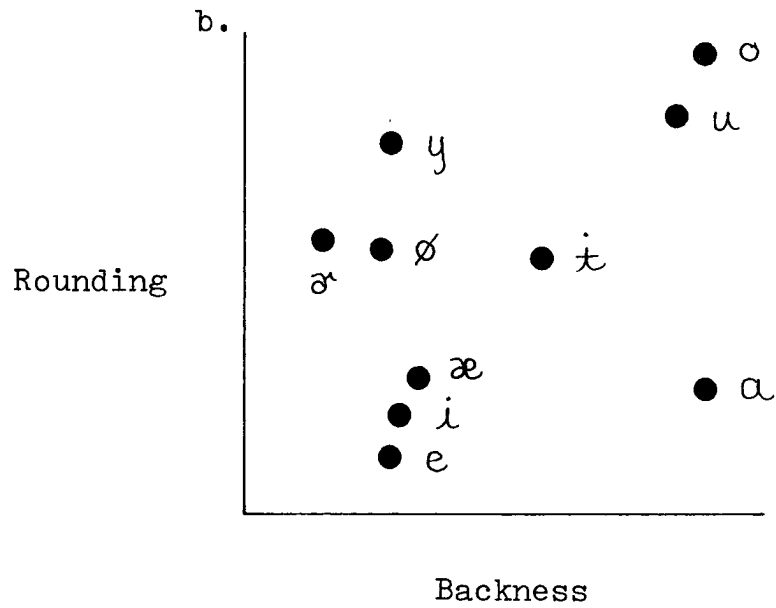
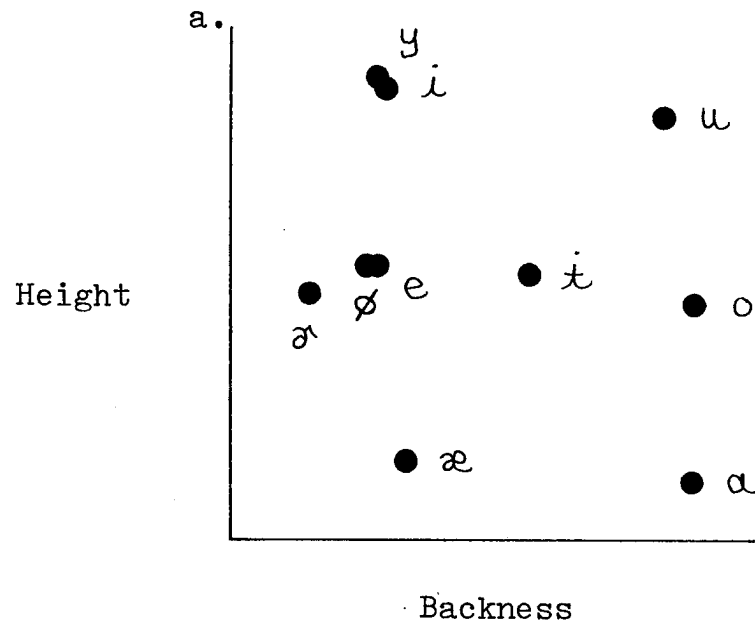


Figure 5.4: Distances along Backness, Height, and Rounding curves from Figures 5.1, 5.2, and 5.3.

tions from the height curve (Section 5.12) are ignored here, as is the curvature of each scale. These plots do not show perceived distances. Rather, they represent hypotheses about the way listeners classify vowels.

Figure 5.4a shows the (straightened) height curve plotted against the (straightened) backness curve. The pattern is striking in its simplicity; each dimension reflects a 3-valued distinctive feature. Backness might be regarded as a 4-valued scale, in order to set off [ɤ] from [e] and [ø]. But the proportions of the figure are not conducive to separating [ɤ] in this way. Furthermore, no listener had both [ɤ] and [ø] natively; no listener had any need to distinguish them.

Figure 5.4b plots distances on the backness curve against distances on the rounding curve. How many feature values are necessary to capture the significant rounding relationships? It is possible to propose at least three: [ɪ e æ ɑ] are [0 round], [y u o] can be called [2 round], and the vowels [ɜ ɔ ɘ] which lie between [0 round] vowels and [2 round] ones must therefore be [1 round]. But if the object is to propose an efficient strategy -- one that makes no more classifications than necessary -- two feature values will suffice. [ɪ e æ ɑ] are, of course, [-round], and [u o y ɘ ɜ ɔ] may all be classed as [+round]. All of those vowels are distinguished from one another by backness and height, except [ø] and [ɤ]. But distinguishing these two vowels is not a matter for the rounding scale, no matter how finely graded it is.

The reduction of six dimensions to three, as summarized here, has a certain elegance and simplicity. However, no mention has yet been made of the acoustic basis of the listeners' judgments. Whether the three curves of Height, Backness, and Rounding constitute a better interpretation than the original six dimensions as derived by PARAFAC should depend in part on the way the proposed dimensions (straight or curved) can be accounted for by the acoustic variables.

5.2 Linear regression versus special strategies

The most appropriate simple test of the relationship between acoustic measurements and perceptual scales is step-wise multiple linear regression. Basically, each PARAFAC dimension and each curved dimension is entered into a series of regressions as the dependent variable. Each successive regression includes the acoustic variable that brings about the most improvement in the prediction of the location of vowels along the perceptual scale. The overall quality of each regression is expressed by the multiple correlation coefficient.

Before any analyses were performed, two decisions were necessary. First of all, eleven independent acoustic measurements were made (F0, F1, F2, F3, F4, A0, A1, A2, A3, A4, D), and the vowels under consideration are but 10. The number of degrees of freedom in a regression analysis is equal to the number of points (vowels) minus the number of variables, the dependent variable included. There must therefore be no more than eight acoustic variables present, in order to achieve even one degree of freedom. Ten vowels minus nine independent variables minus one dependent variable results in zero degrees of freedom.

Which three variables to eliminate? The amplitude and frequency of the fourth formant seem like indisputable choices, and they are discarded with no feeling of loss. Ideally, the overall amplitude (A0) and stimulus vowel duration (D) should be irrelevant as well. But the vowels did vary somewhat in duration, in spite of attempts to produce identically long vowels. Listeners may have been attending to duration differences; duration must be included to determine whether this is so. As for overall amplitude, it is reasonable to ask whether the initial normalization of peak amplitude affected the judgments, and in what way. The third variable ultimately chosen for exclusion was A3, amplitude of the third formant. The remaining variables, then, are F0, F1, F2, F3, A0, A1, A2, and D.

The second decision is whether to use the acoustic variables in their original form or to normalize them. If the values are left unchanged, then the regression coefficients will be distorted by inherent differences in scale from one variable to another. Suppose, for example, that the overall amplitude and the first formant were each equally subjectively important in placing a vowel on a perceptual dimension. Unfortunately, F1 values range from 215 to 610 (Hz), where A0 has a much smaller range of 0.0 to 4.6 (dB). The coefficient for A0 will be around 100 times as large as that for F1, giving a markedly distorted picture of the relative salience of F1 and A0. In order for relative salience to be read directly from regression coefficients, some normalization can be accomplished to make the mean and distribution of the variables more equal. In the analyses described below, this was achieved by converting all variables, dependent and independent, to standard scores. This, too, has a disadvantage: a variable that not only has little variance, but in fact is perceptually irrelevant, may very likely show a greater regression coefficient than it deserves. In any case, however, the multiple r values are unaffected.

It is worthwhile to note that stepwise analysis is crucial for a data set as small as this one (10 points). The number of points is uncomfortably close to the number of independent variables available; in fact, 9 lists of random numbers would result in a perfect fit for any perceptual scale; 8 would give an extremely good fit. Conversely, the 8 chosen lists of acoustic variables could predict a list of random numbers rather well. One random list, regressed against all 8 acoustic measurements, resulted in a multiple $\bar{r}$ of .99.

In evaluating the regression analyses, the following guidelines were observed. First, solutions that require more than four variables to achieve an $\bar{r}$ of about .9 were regarded with some suspicion. Second, attention was also paid to the way the predicted scales group the vowels into clusters. It is possible to have a high correlation associated with a prediction of vowel positions that obscures gaps that appear on the perceptual scale at hand. This sort of result is less preferred than one which matches the basic perceptual clustering of vowels, even though $\bar{r}$ may not be as good.

It often occurs that no linear combination of variables is a satisfactory reflection of a perceptual dimension. This is a perfectly plausible result, since there is no reason why human beings, in judging vowel sounds, should be restricted to computing weighted sums or differences of acoustic properties as the source of their distance judgments. Linear regressions also impose the unjustified condition that a listener must fix a vowel's location on each dimension without knowing where it lies on any other dimension. By postulating that the position of a vowel on some scale(s) is ascertained only after determining its location on other scales, it is often possible to reduce the number of acoustic variables thought to underlie those scales.

The regression analyses are summarized in Figure 5.5, which plots stepwise improvement in $\bar{r}$ vertically. Each vertical family of dots refers to a different perceptual scale. This figure will be referred to often. Figure 3.3 and Table 3.3 will be useful references in regard to the proposed special strategies.

5.21 Backness

It comes as no surprise to find that F2 alone correlates well with the Back/Nonback1 dimension (Figure 5.5). But a good correlation coefficient per se is actually not the most sought-for result, as mentioned above. Rather, it is hoped that the acoustic variables can be combined to produce

scales which show gaps where the perceptual scales do. Figure 5.6 illustrates the problem. In the Back/Nonback1 scale, included in the figure for reference, [u o ɑ] are set well off from the other vowels. But in the prediction of that scale made by F2 alone, there is no satisfactorily large distance between [ɜ] and the back vowels. The addition of a second variable, F1, improves the pattern, even though r improves insignificantly. A ternary division into [u o ɑ] [ɜ] [ɔ̃ ɛ̃ æ̃ ỹ ẽ ɪ̃] is more plausible with two variables. Including any more variables does not allow the prediction to achieve a better front-back relationship; in fact, the position of [ɔ̃] becomes harder to account for, as seen in the top two rows of predictions.

The expression for the 2-variable regression, ignoring the unessential constant term and scaling the coefficients for convenience, is $F2 - .16F1$. The influence of F1 appears to be slight, but this is only an artefact of the use of standard scores. Regressions using raw formant frequencies would show F1 to be much more important.

More appropriate coefficients can be found by attempting to draw a line through the F1-F2 plane such that the projections of the vowels onto that line provide maximal separation of the three groups [u o ɑ], [ɜ] and [ɪ e æ ɪ y ɔ̃]. It may come as somewhat of an anticlimax to find that the coefficients for such a projection are about equal; that is, Back/Nonback1 is reflected reasonably well by the raw difference between the first two formants, as plotted in Figure 5.7. The words "about equal" are chosen as a reminder that this finding is an informal one, based on inspection of Figure 3.3. An "exact" pair of coefficients could only be proposed after establishing mathematical conventions to maximize the distance between groups and minimize the within-group distances. This is not a goal of the present work.

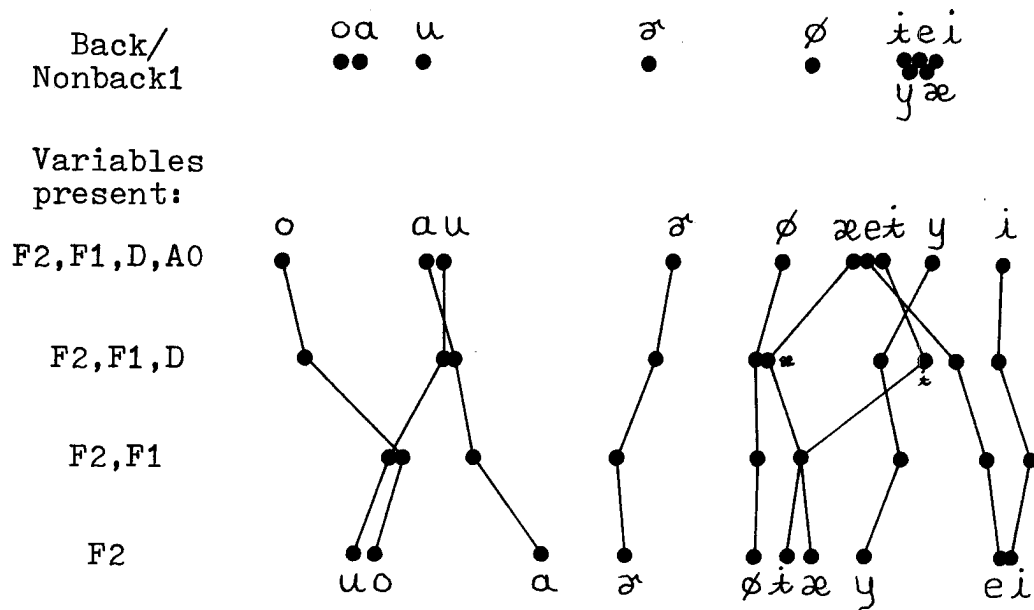


Figure 5.6: Back/Nonback1 and its prediction by some combinations of acoustic variables.

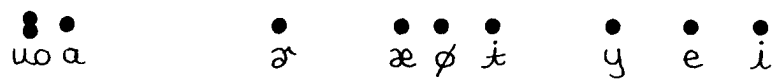


Figure 5.7: F2-F1 as a plausible source for the perceptual scale of Back/Nonback1.

The main difference between F2-F1 and Back/Nonback1 is that the F2-F1 values for [i e y ɤ ø æ] are too spread out. It seems that the perceptual dimension is the result of a gross quantization of the F2-F1 values that tends to pull the vowels into clusters. This effect is precisely what is to be expected if the vowels are categorized according to distinctive features upon which the listeners' difference judgments are based.

Back/Nonback2, which associates [ɤ] with back vowels, is badly predicted by F2 alone (Figure 5.5), and the multiple r only reaches .85 for two variables. But the real problem is that the crucial vowel [ɤ] is clearly grouped with front vowels in the 2-variable case, as is shown in Figure 5.8 (bottom row). Four variables are needed to place [ɤ] with [u o ɑ], but these include duration. As was pointed out in Section 3.4, the duration of [ɤ] was somewhat anomalous. If this perceptual scale is to retain its credibility, a more satisfactory explanation must be found.

An interesting solution is to abandon the search for a satisfactory linear regression and propose instead a strategy that depends upon the previous assignment of values on other dimensions. F3 alone is then sufficient to locate vowels on Back/Nonback2, given the following sequence of decisions.

1. Label the vowel back or front, according to Back/Nonback1. [u o ɑ] are the only back vowels.
2. Label the vowel high, mid, or low. This is assumed to be possible, whether or not the Height curve is preferred over the PARAFAC dimensions relevant to vowel height.
3. Label the vowel round or nonround (according to either the PARAFAC rounding dimension or the rounding curve).
4. If the vowel is mid, nonback, and rounded, examine F3.

The vowels meeting this description, according to the feature descriptions made in Section 5.1, are [ɤ], [ø] and [œ], which have widely divergent F3 values: 2520, 2175, and 1825, respectively. [ɤ], with its high F3, is set at one end of the scale; the low F3 of [œ] places it at the opposite end, and [ø] lies in between. If it is thought necessary to assign Back/Nonback2 values to every vowel, all others are assigned values that match their positions on Back/Nonback1.

One difficulty with this approach is finding a reason why a nonback round mid vowel with high F3 should associate with backness, and why one with low F3 should associate with fronting. If this dimension is only used by Turkish listeners, then the grouping of [ɤ] with back vowels is a phonological effect, related to the position of the tongue for [ɤ]. But then the positioning of [æ] as the extreme "front" vowel is puzzling, since Turkish has no [æ]. It seems that Back/Nonback2 is somewhat difficult to explain in terms of acoustic properties, regardless of whether those properties are entered into linear regressions or are manipulated more freely.

What improvement can be expected in relating acoustic variables to the Backness curve of Figure 5.1? Positions along the curve are predicted extremely well ($r = .96$, Figure 5.5) by only two measures, F2 and F3. But Figure 5.9 shows this high correlation to be deceptive, because the clear separation of [ɤ] from the front vowels is not present. [æ] is also not set off as desired, and the front vowels are not clustered together tightly enough. If more variables are added, the positions of [ɤ] and [æ] are improved, but the next two variables are fundamental frequency and duration, which are only indirectly related to vowel quality (Section 3.4). The clustering of front vowels is not improved at all.

An account of vowel locations on the backness scale requires a strategy such as the following sequence.

1. F2-F1 is attended to, and provisional labels of front and back are applied, which effects a quantization into two tightly grouped clusters: [u o ɔ] and [i e æ y ø ɐ ɤ].
2. Vowels are further categorized according to height and rounding.
3. If a vowel has been provisionally classed as [ø], its third formant is assessed. A high value gives a relabelling as [ɤ], and a reassignment of centered on the backness scale. A low value prompts a relabelling as [æ].

This interpretation can be simplified by taking the individual languages more into account. Only Turkish distinguishes [ɤ] from [ø], requiring the decision process just mentioned. German and Swedish do not have [ɤ] or [æ]; the backness curve for them collapses to one dimension, probably one like Back/Nonback1, which groups vowels into two clusters according to formants 1 and 2. English has [æ] but not [ɤ]; Back/Nonback1 is better able to set off [æ] than is the back-

ness curve. Thai, although it has [ɤ̃], does not contrast it with [ø] or [ɤ̃], making step #3 above unnecessary. These considerations lead to the prediction that when the languages are analyzed separately, only Turkish will show the backness curve, where the other languages will make use of one simple backness dimension. This will be borne out in Chapter 6.

5.22 Height

The most striking feature of the Low/Nonlow dimension, namely, the gap between [æ a] and the other vowels, is well predicted by F1 alone, even though the multiple $\bar{r}$ barely reaches .90. Discrepancies in ordering within the Low and Nonlow groups account for the relatively low $\bar{r}$ (see Figure 5.10). Including more variables does not improve the predicted grouping of vowels.

As for High/Nonhigh, which places [ɑ] and [æ] between high and mid vowels, the attempt to find acoustic correlates is largely unsuccessful. Figure 5.11 illustrates the problem. Three variables are required to bring the correlation coefficient to .89: A1, F0, and F2. This is the most unlikely collection of predictors found for any of the dimensions. F0 and A1 values for the stimuli have extremely narrow ranges -- 8 Herz (about a semitone) for the former and 6dB for the latter. No justification exists for claiming any generality for the distribution of these values. F2 can only account for the difference in placement of pairs such as [e ø] or [i y], not for the general appearance of the scale. Even the inclusion of up to seven variables cannot account for the position of [ɑ]. The regression analyses show that if [i y u] are grouped together, no combination of variables can be found that will include [ɑ] among them.

In an earlier discussion of this problem (Terbeek, 1973), it was proposed that [ɑ u i y] be seen as one of the two subgroups of vowels. The explanation was that the sounds were classified according to a particular articulatory scheme. Vowels formed at corners of the vowel triangle, that is, the lowest, most fronted, and the most backed vowels, were perceived as being somewhat different from vowels formed in any other way. This dimension, therefore, could be used to directly evaluate the markedness of vowels: [ɤ̃ ø ɤ̃ o e æ], by virtue of their specification on this dimension, were more highly marked than [ɑ i y u]. The latter, all formed at articulatory "corners," are more predictably present in vowel systems. The fact that "marked" [e] is, according to this scale, less expectable than "unmarked" [y] is no difficulty; a second marking dimension dealing with the relationship between rounding and backness restores the balance. There is



Figure 5.10: F1 as a plausible source for the perceptual scale of Low/Nonlow.

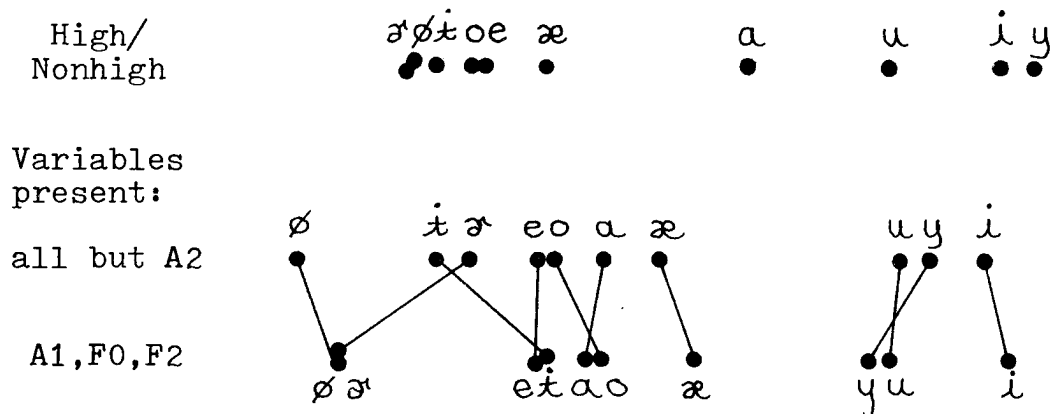


Figure 5.11: High/Nonhigh and its prediction by some combinations of acoustic variables.

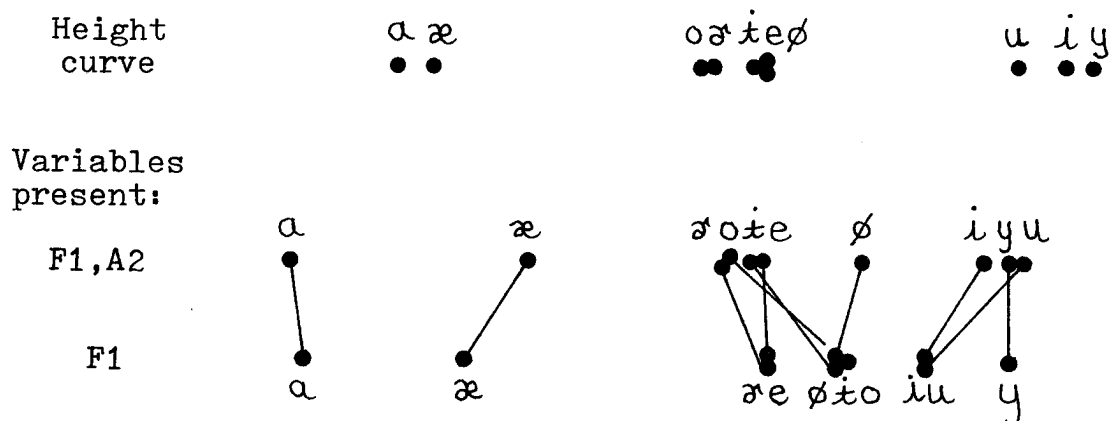


Figure 5.12: Height curve and its prediction by some combinations of acoustic variables.

no new evidence that shows this approach to be wrong. But it is at least equalled in plausibility by an appeal to a curved height dimension which is based more solidly on reasonable acoustic parameters.

If the two height dimensions are re-interpreted as a single curved scale, as suggested in Figure 5.2, F1 appears as the single best predictor ($r = .92$, Figure 5.5). The addition of a second variable (A2), which increases r to .95, slightly improves the prediction of the gap between mid and high vowels, but degrades the gap between low and mid vowels (c.f. Figure 5.12). Including more variables merely increases the size of the High--Mid gap.

The 3-way quantization of the height curve suggests once more that the difference judgments may be based upon a feature extraction, not raw acoustic information. The closeness of [α] and [æ] on the perceptual scale is easiest to account for by postulating that vowels are labelled [+low] or [-low] on the basis of F1.

It is not at all clear whether F1 alone is the best source for the mid--high distinction. The fact that A2 is the second variable to emerge in the regression analyses seems to be an accident, although no better solution presents itself. But however that distinction is made, it appears to be just as significant a perceptual difference as the difference between low and mid.

It may seem that the three distinct clusters making up the height curve support a 3-way linguistic description of vowel height, as opposed to a description using four levels, as found in Jones (1967) and Ladefoged (1972). Unfortunately, the main vowels which would occupy the extra level, [ɛ] and [ɔ], are not present in the experiment. Whether those vowels would group with low or mid vowels, or form a separate group between low and mid, is unknown.

5.23 Rounding

The three clusters of vowels on the rounding dimension can be predicted reasonably well by a weighted sum of F1 and F2 ($r = .93$, Figure 5.5). The result of the linear regression using these two variables is seen in Figure 5.13. A minor problem is the appearance of [ɪ] ambiguously between [æɑe] and [ɛøyɔ]. But linear regressions try to minimize all deviations from the pattern of the dependent variable; in this case the "best" overall fit does not best fit the three perceptual clusters. A better fit can be found by inspection of Figure 3.3; the regression is $F1 + .29F2$. The prediction is plotted in Figure 5.14. [æeɑɪ] are still spread out, but the separation between that group and the middle one is much clearer.

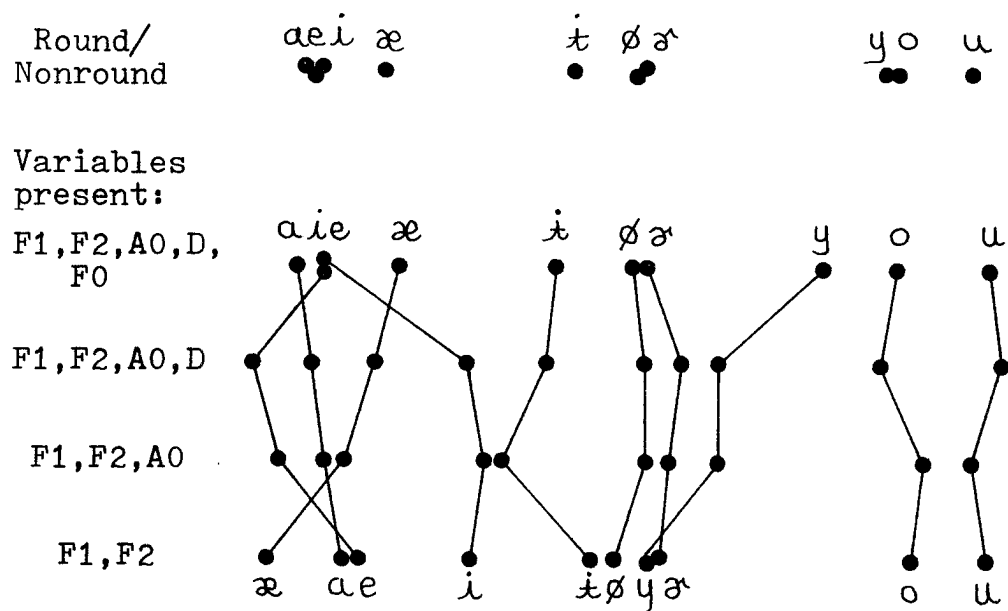


Figure 5.13: Round/Nonround and its prediction by some combinations of acoustic variables.

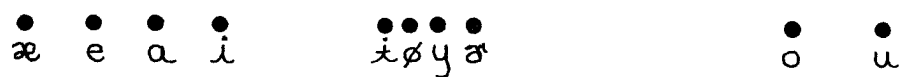


Figure 5.14: $F1+.29F2$ as a plausible source for the perceptual scale of Round/Nonround.

A more serious problem is that the 2-variable predictions place [y] within the middle [ɪ ʌ ø] group. It is possible to resolve this discrepancy by adding more variables, namely, A0, D and F0. (see Figure 5.13). This amounts to the claim that listeners used relative stress to place [y] near [u o] on the rounding scale, which in turn implies that this result cannot be generalized any farther than the stimulus tokens themselves. However, it is not necessary to view the rounding scale in this way.

An alternate view is that listeners apply a simple non-linear strategy to F1 and F2 alone, which operates as follows:

1. Does the vowel have as high an F2 as possible, given its F1? If so, it is [0 Round]. This step is represented by the upward concave curve on Figure 5.15.
2. If the vowel is not [0 Round], it is either [1 Round] or [2 Round], depending upon whether its first formant is sufficiently low. This step is represented by the downward concave curve.

Granted that Figure 5.15 presents a reasonable case as to how [y] might be grouped with [u] and [o], it is worth briefly considering why. [y] certainly lies closer to [ɪ ø ʌ] on the F1-F2 plane than it does to [u o]. It is distinctly possible that speakers of different languages categorized [y] in different ways, and that the placement of [y] on the Rounding dimension is biased toward those listeners who heard it as some kind of [u]. Step #2 of the above strategy is therefore not necessarily applicable to each language group, as will be seen in Chapter 6.

The last of the six dimensions is Peripheral/Central. The linear regressions, summarized in Figure 5.5, are not at all satisfactory. The single most salient variable is F3, which does no more than isolate [ɔ] at one end of the scale. The next two most significant variables, duration and F0, effect some slight improvement in ordering [a ɪ o e] at the opposite end of the scale. Not until five variables are included does [ɪ] appear near [ɔ]. Figure 5.16 gives the 3- and 5-variable predictions. As far as special strategies are concerned, it is possible to use the F1-F2 plane to isolate [ɪ ø ʌ], which appear at one end of this dimension. The above strategy proposed for labelling [ɪ ø ʌ] as "1 Round" will suffice to label them "+Central" as well. But the lack of a convincing gap separating those three vowels from the others -- or any other suggestive gap -- makes any proposal a fragile, tenuous one.

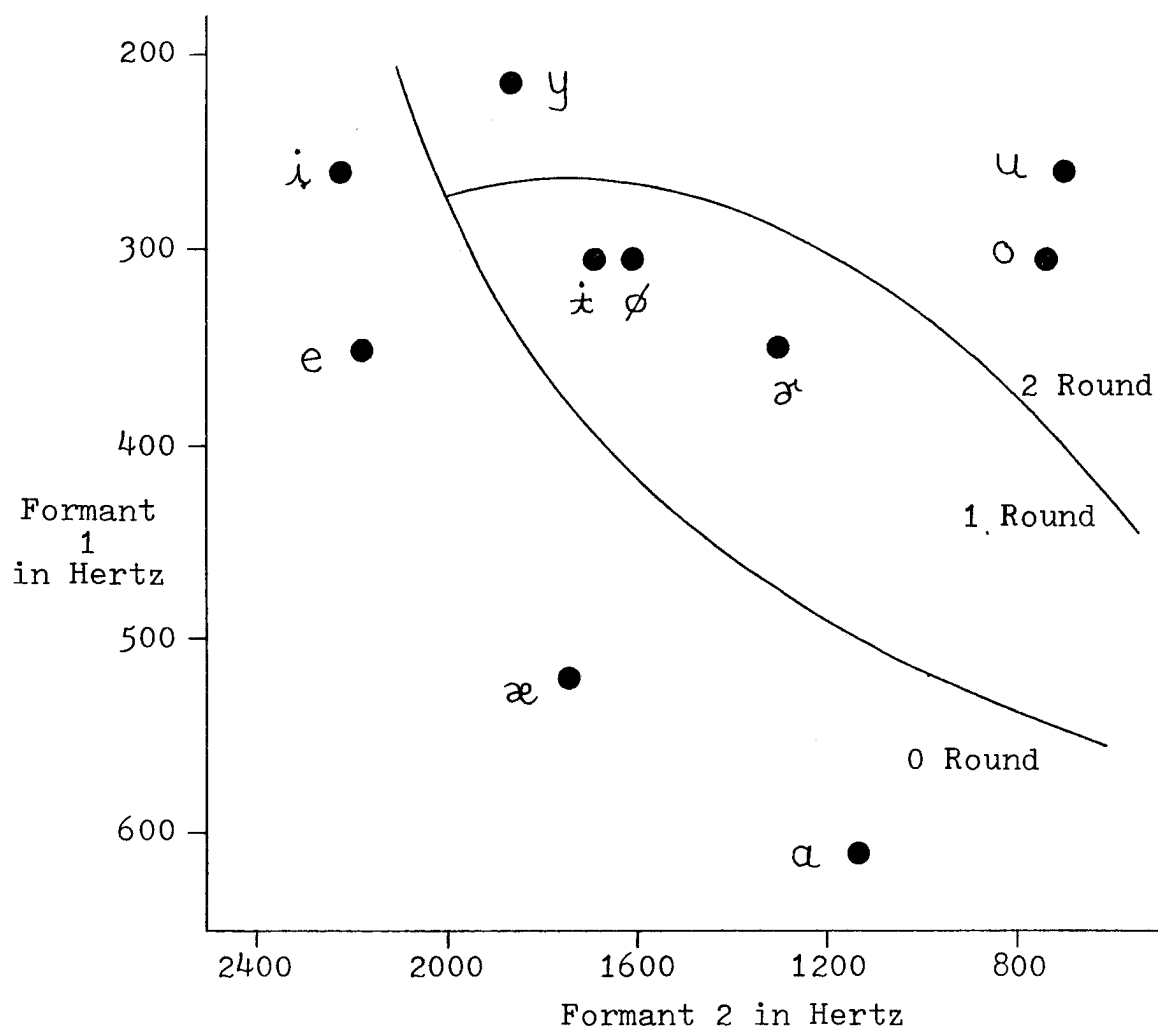


Figure 5.15: Partitioning of the F1/F2 plane to account for vowel clusters found on the perceptual dimension of rounding.

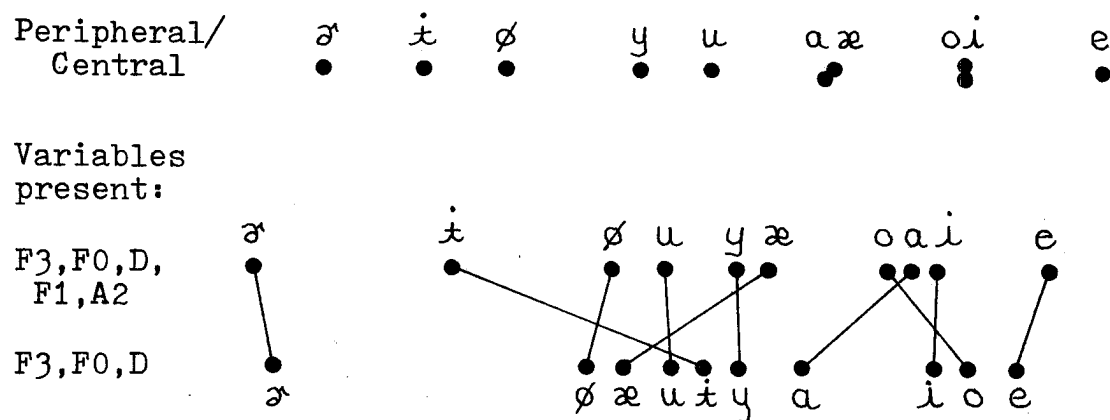


Figure 5.16: Peripheral/Central and its prediction by some combinations of acoustic variables.

The lack of clustering invites the speculation that several perceptual scales are present here, scales that are similar enough to collapse together in a confusing way. For example, assume that English listeners listened to F3 as a cue for retroflexion, and that subjects from a different language group categorized vowels according to whether they appear on the periphery of the F1-F2 plane or more toward the center. The former dimension would separate [ɤ] from the remaining vowels, while the latter would group [ɨ ø ɤ] together. A composite dimension could emerge, diluting the crucial properties of both "real" dimensions. This possibility will be discussed further when the individual language solutions are presented in Chapter 6; the example anticipates the analysis of English and Thai.

A final attempt at interpreting Central/Peripheral is to regard it as a component of a curved dimension of Rounding, as drawn in Figure 5.3. The Rounding curve can be cut to yield the same three-vowel clusters found in Round/Nonround alone. In this view, the position of vowels along the curve requires basically the same acoustic information as for the Round/Nonround dimension itself (Figure 5.5). The inability to correctly place [y] using only F1 and F2 reappears, and the strategy proposed to overcome this problem is appropriate here, too. It would seem that Central/Peripheral can be treated as an artefact: listeners' perception of distances along the dimension of Rounding cannot be captured on a straight line.

Unfortunately, a closer inspection of the vowel positions along the Rounding curve casts doubt on this otherwise satisfying proposal. The difficulty is that the clustering of vowels into the three groups [aeɨɤ], [ɨ ø ɤ], and [u o y] is weaker for the curved scale than for the original Round/Nonround scale. The continuous nature of the Central/Peripheral dimension has not disappeared; it blends into the curved scale, blurring the boundaries of the clusters. Central/Peripheral contributes nothing to the interpretation of Rounding.

This is not the case with the Height and Backness curves. The curved Height scale reflects a much more intuitively satisfying and more discrete pattern of vowels than either of its component dimensions. The Backness curve provides a different but no less interpretable scale than either of its two PARAFAC dimensions. But with rounding, the curved scale adds nothing. It seems more prudent to regard Round/Nonround as a dimension distinct from Central/Peripheral, and to interpret the latter only in the context of individual languages (Chapter 6).

5.3 In defense of six dimensions

Now that the 6-dimension vowel space has been discussed in detail, it is appropriate to give the evidence for preferring it to a solution with more or fewer dimensions. Three kinds of evidence are formal and easy to describe quickly: 1) the plot of fit against dimensions extracted, 2) the appearance of unacceptable personal weighting factors, and 3) the appearance of multiple solutions. A fourth kind of evidence is the interpretability of the dimensions. Therefore it would have been somewhat difficult to deal with, if it had preceded Section 5.2. The contents of the 1- through 7-dimensional solutions will be summarized to show how the interpretation of the results as a whole would be altered if other solutions were preferred.

5.31 Formal evidence for six dimensions

All three formal indicators are given in Figure 5.17. The first of these, the plot of fit vs. dimensions extracted, is a relatively uninformative, gradual curve. The most prominent elbow is at two dimensions, but the choice of a 2-dimensional space as the final solution is excessively conservative. None of the studies discussed in Chapter 2 (with the exception of the 6-vowel experiment by Mohr and Wang) had satisfactory solutions of fewer than three dimensions; the large number of vowels and listeners used in the present analysis indicates a 3-dimensional solution at the very least. There is another elbow in the fit curve at three dimensions, but it is so slight as to be barely credible. The evidence provided by the fit curve is therefore worth little.

The bottom curve on Figure 5.17 shows the extent to which the personal weighting factors deviate from an acceptable MD-scaling interpretation. The measure used is the sum of the squared unacceptable personal weighting factors. As was pointed out in Chapter 2, there is no mathematical constraint that the weights must be positive. But PARAFAC prints the square of the personal weights, and a negative value represents a breakdown of the model. Strictly speaking, only the 1-, 2- and 3-dimensional spaces are real MD-scaling solutions, because there are violations of this constraint in all the others.

As a practical matter, these violations represent a very small fraction of the personal weights. In six dimensions, for example, the sum of the squared personal weights is 635, and the lone violation is only .0003 of that amount. For the 4-, 5- and 7-dimensional spaces, the proportion of violations to summed squared weights is somewhat larger, but

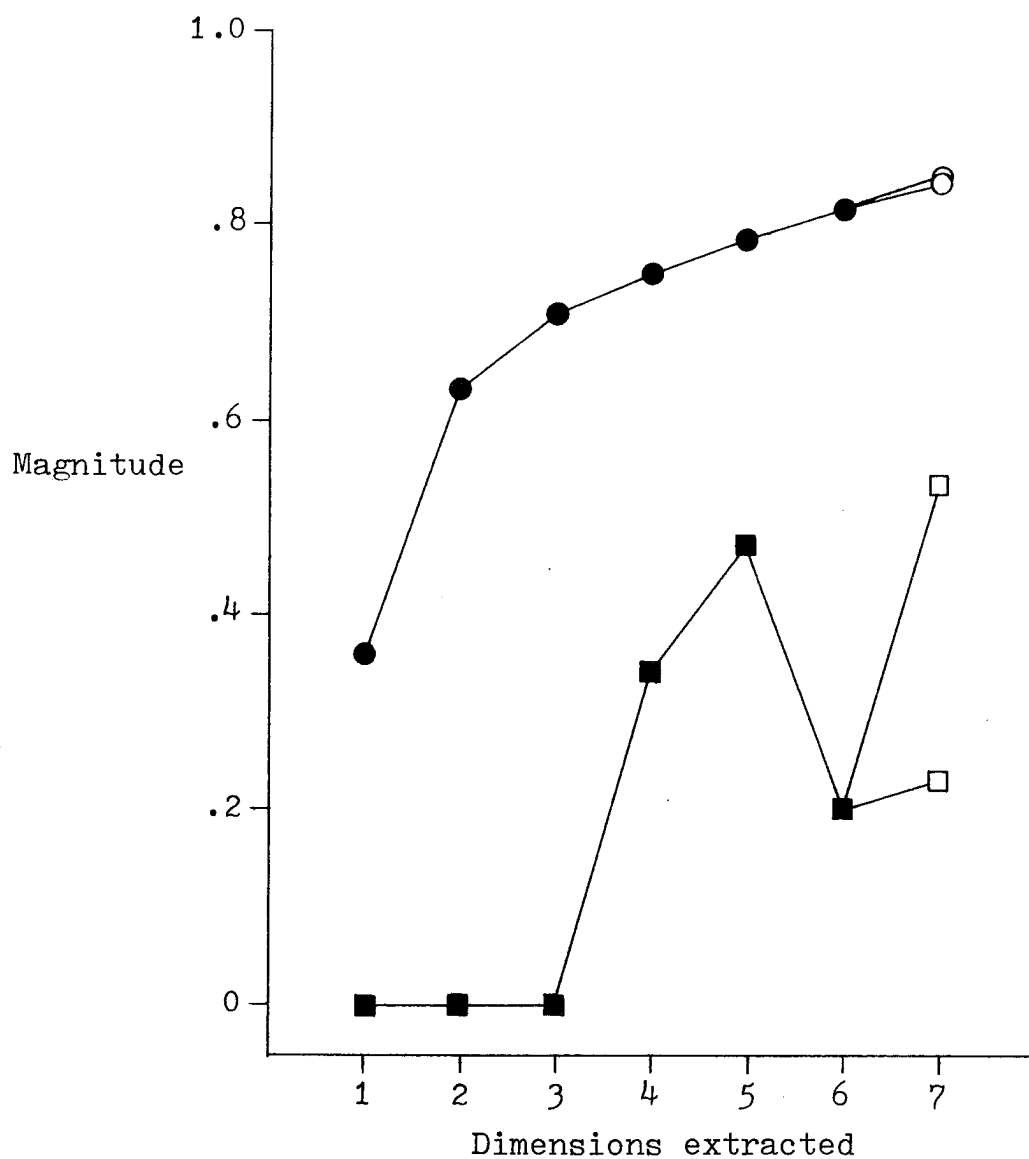


Figure 5.17: 3 kinds of formal evidence for determining the correct number of dimensions, for the PARAFAC analysis of all 35 listeners.

- = Fit
- = Sum of squared unacceptable personal weighting factors
- , □ = Appearance of multiple solutions

still small. Also, there are only two violations in each of these three spaces. The view adopted here is this: to discard an entire solution because one or two out of 35 subjects showed small interpretable values on only one dimension places a mistaken trust in blind rigor. Imaginary personal weighting factors, small in absolute value, are regarded as zero.

The plot of dimensions vs. the sum of squared violations therefore appears to be no sure guide. The 3-dimensional space, the largest one to have no violations at all, is to be barely preferred over the solution to its right. Among those larger solutions, the 6-dimensional space has an equally slight advantage.

The criterion that most influenced the choice of six dimensions was the uniqueness-of-solution property. For each case of n dimensions, four solutions were obtained, each one with a different starting pattern of random coordinates. The four solutions were identical for the 1- through 6-dimension cases. But with 7 dimensions, two different solutions resulted. As shown on Figure 5.17, they had the same fit value, but very different personal weighting factor violations. According to the theory underlying PARAFAC, the correct number of dimensions is the number of dimensions that appears immediately before multiple solutions occur. The 6-dimensional solution, then, is the appropriate one to focus on.

5.32 The interpretation of alternate solutions

In any use of MD-scaling, the ultimate evidence for any solution is its interpretability. This section defends the interpretability of the 6-dimension solution against the other solutions obtained. Rather than regarding the other solutions as whole units, one at a time, the whole family of solutions will be considered in "cross-section": each dimension of the 6-D space will be compared with its counterparts in the 1- through 5- and 7-D spaces. The relative salience of a dimension is to be seen in the order of appearance of that dimension in progressing from one size space to the next.

It is convenient to refer to "counterpart" dimensions from one space to another, and the "evolution" of the 1-D space into the 6-D space, and the presence of "a dimension" in several solutions. But it must be emphasized that all the solutions discussed are independent. All the dimensions present in any one solution are derived simultaneously from a pattern of random starting coordinates, a pattern different from that of the other solutions. To say that Back/Nonback1, for example, "appears in" both the 2- and 3-D spaces merely

means that in each of those independently-derived spaces there appears a dimension which shares an overall resemblance to the dimension in the 6-D space called Back/Nonback1.

Figure 5.18 shows the "history" of Back/Nonback1. Each row on the figure represents a dimension from a different solution. This figure shows the distinction between back [ʊ ɔ ɑ] and all other vowels to be highly salient indeed, since it appears in every solution obtained. In rows 1 through 4 (solutions with 1, 2, 3, and 4 dimensions), the front vowels are more spread out than in the remaining rows, and the 6, 7A and 7B solutions have [æ] displaced towards the center. As it happens, the backness scale of the 5-dimensional space is the one most easily cut into two closely-grouped clusters (which point will be taken up again shortly). But a dimension that separates back vowels from others is an undeniable component of the listener's responses and does not depend upon which solution is preferred.

In looking for the next-most salient dimension, one might hope to find a scale which appeared in the 2-dimensional space and in all subsequent ones, essentially unchanged. But this hope is unfulfilled. The second dimension of the 2-D space cannot be traced directly to any one dimension of the 6-D space; rather, it is related to two of them. One of the dimensions with an indirect ancestor in the 2-D space is Round/Nonround.

Figure 5.19 traces the evolution of the rounding dimension from 2 through 7 dimensions. It is first rewarding to notice that the 5- and 6-dimensional spaces are the easiest to cut according to distinctive features, with clear gaps between the three clusters. These two dimensions are intrinsically more interpretable than the others. Both 7-dimensional solutions spread the vowels out considerably. In the 2-, 3- and 4-D spaces, the vowels are also more spread out; worse yet, the gap between highly rounded [ʊ ɔ ʏ] and the [ɪ ʊ ɐ] group is less clear than for the 5-, 6- and 7-D solutions. Still, the vowels are at least ordered according to Rounding.

The other scale which can be traced to the second dimension of the 2-D space is Low/Nonlow (c.f. Figure 5.20). The bottom three rows of this figure are identical to those of Figure 5.19. In the solutions with five or more dimensions, the separation of [æ ɑ] from nonlow vowels is strikingly apparent. In the 2-, 3- and 4-D spaces, the label is hardly appropriate, since [e] and [ɪ] are placed considerably closer to [æ ɑ], vitiating a plausible low-nonlow cut. Yet many vowels are ordered according to height. [æ e ɪ] form a series of increasing height, left to right, and so do [ɑ ɔ ʊ]. [ø] and [y] are displaced farther to the right than their unrounded

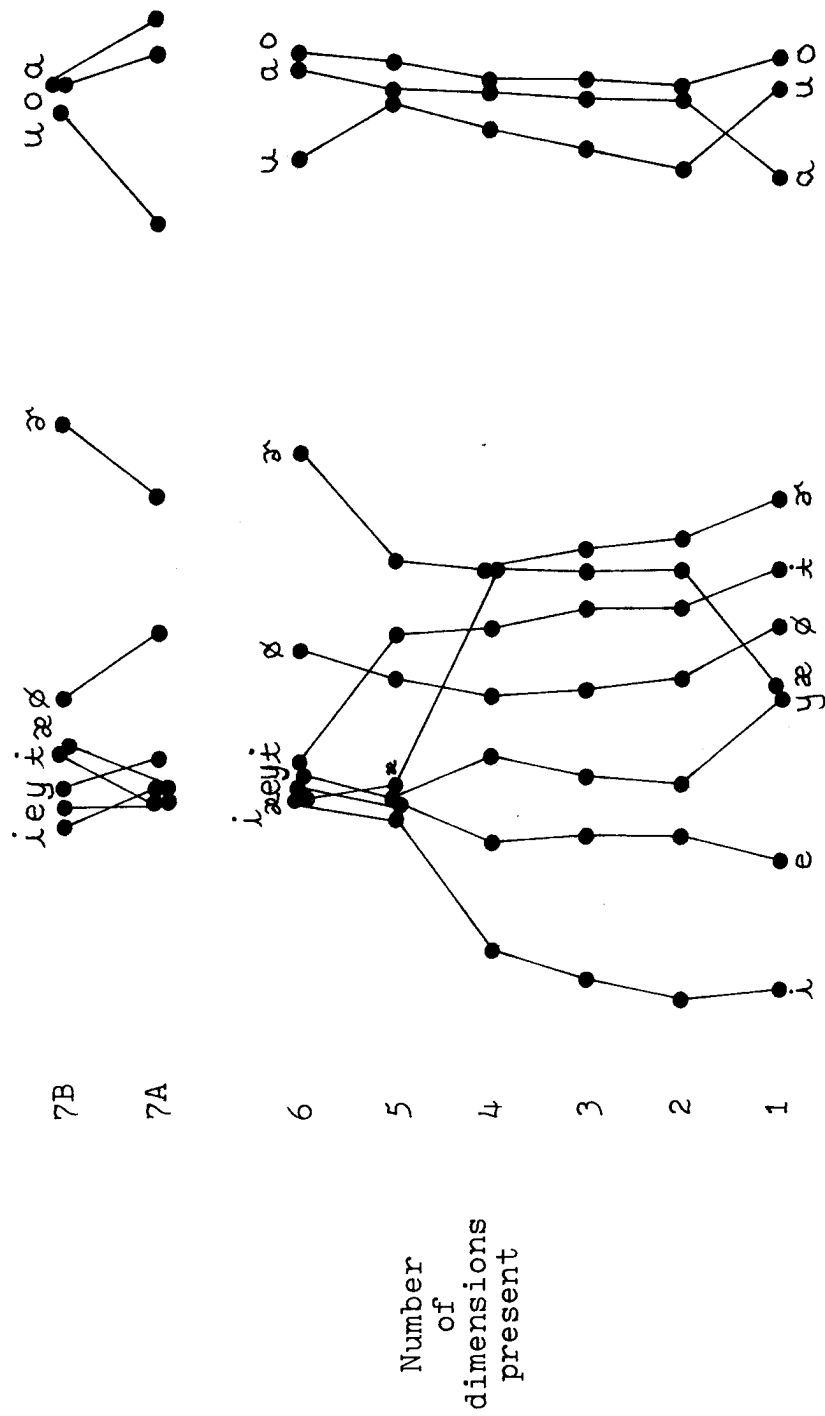


Figure 5.18: Evolution of the Back/Nonback1 dimension.

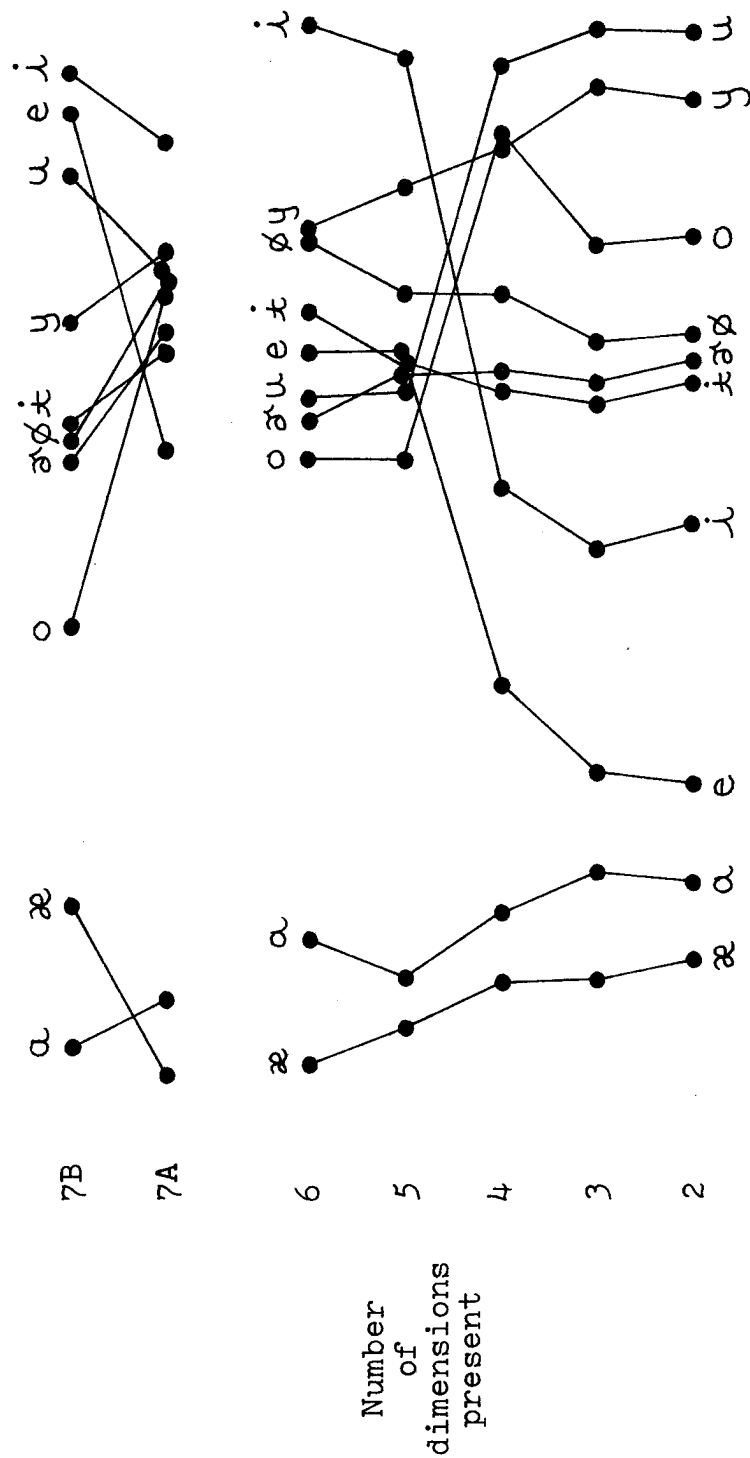


Figure 5.20: Evolution of the Low/Nonlow dimension.

counterparts [e] and [i] because the dimension partially reflects rounding as well.

If Figures 5.19 and 5.20 are compared, it is clear that the Low/Nonlow and Round/Nonround scales do not emerge as separate dimensions until five dimensions are obtained. In the smaller spaces, they are collapsed into one dimension that has properties of both rounding and height, and is therefore difficult to interpret. The "collapsing" referred to has a simple mathematical meaning. If Round/Nonround and Low/Nonlow values from the 5-D space are plotted against one another, the projections of the vowels along the line $X=Y$ correlate well with the positions of the vowels on one dimension of the 4-D space ($r = .991$).

Although the scales of Back/Nonback1, Round/Nonround, and Low/Nonlow appear to be the three most salient dimensions, they do not appear independently until the 5-D solution. What, then, constitutes the third dimension of the 3-D space? Again, it does not match any one of the patterns in the 6-D space. Rather, two dimensions are collapsed into one: Peripheral/Central, and High/Nonhigh.

Figure 5.21 shows the dimensions in the 3-, 4-, 5- and 7-dimensional spaces that correspond best with Peripheral/Central of the 6-D space. All of the dimensions on the figure show [ɤ], [ɤ̃], and [ø], in that order, at the left end of the scale, and none shows any convincing gaps that suggest a distinctive feature cut. No appeal to a different solution can mollify the difficulty experienced earlier in interpreting Peripheral/Central. The shift in position of the seven rightmost vowels between the 3- and 4-D solutions neither aids nor hinders the interpretation.

The development of Peripheral/Central is paralleled by the development of High/Nonhigh, shown in Figure 5.22. First, solutions of four dimensions or more all have a dimension which corresponds very closely to the High/Nonhigh scale in the 6-D space: [y i u] lie at one end, [a] lies toward the middle, and the remaining six vowels cluster at the other. The unsatisfying presence of [a] is a property of all these solutions. Second, High/Nonhigh does not appear in the 3-D space. The dimension shown at the bottom of Figure 5.22 is the same as the one drawn in Figure 5.21.

This dimension from the 3-D space is certainly the "source" for both Peripheral/Central and High/NH. If the Peripheral/Central and High/NH scales from the 4-D space are plotted against one another, the projection onto a 45° line correlates highly with the 3-D scale in these two figures ($r = .991$). Peripheral/Central and High/NH are approximately equal in

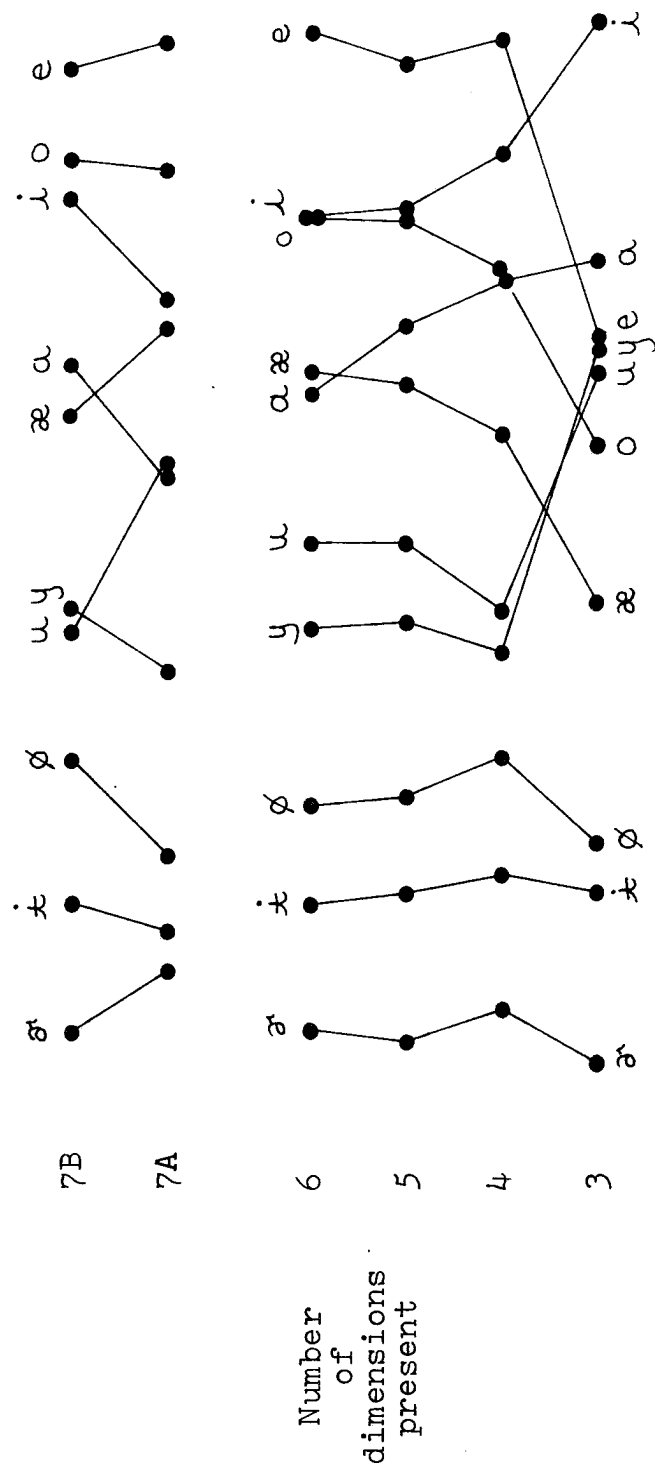


Figure 5.21: Evolution of the Peripheral/Central dimension.

overall salience and cannot be distinguished by PARAFAC until four dimensions are obtained.

Only Back/Nonback2 remains to be discussed. This dimension emerges at six dimensions and remains unchanged in both of the 7-D spaces (Figure 5.23). The Back/Nonback1 scale from the 5-D solution is provided for comparison only. The late appearance of Back/Nonback2 is consistent with the proposal made earlier that it is relevant to Turkish listeners only. High salience for only a fraction of the listeners could logically underlie a dimension's apparent overall low salience.

The evolution of the 6-dimensional solution as a whole is sketched in Figure 5.24. Each dot represents a single dimension. Horizontal dashed lines connect dimensions which together constitute a particular solution; vertical solid lines connect dimensions which re-appear in successive solutions. Where two solid lines descend from one point, the upper point represents a projection of two scales which appear separately in the next-largest solution.

This figure allows the interpretability of each solution to be summarized. Which space, then, is the "correct" solution, based on interpretability? 1-D is hardly a candidate; it is interpretable but undeniably incomplete. 2-, 3- and 4-D include dimensions which collapse two interpretable scales into a single unsatisfactory scale. Not until 5-D are all collapsed dimensions resolved into component dimensions for the first time.

The 5-D space has one slight advantage over the 6-D solution. Back/Nonback1 is more easily interpreted as a binary scale; [æ] lies much closer to the [-Back] group in the 5-D version of Back/Nonback1 (c.f. Figure 5.18). All the other dimensions which appear in both 5-D and 6-D are essentially identical (c.f. Figures 5.19-22). 5-D is certainly a good solution, yet it has the obvious disadvantage of excluding an interpretable dimension, namely Back/Nonback2.

The two 7-D spaces do not appear on Figure 5.24, because the interpretable dimensions of both of them are the same as for the 6-D solution (c.f. Figures 5.18-23). The remaining dimension in each 7-D space is not interpretable, either within or outside of the realm of established distinctive features or acoustic measures. Nothing is lost by conforming to the theoretical limits imposed by PARAFAC: finding multiple solutions means that too many factors have been extracted. It is possible, of course, that with a greater number of vowels, or more listeners, or more languages, more dimensions



Figure 5.23: Evolution of the Back/Nonback2 dimension.

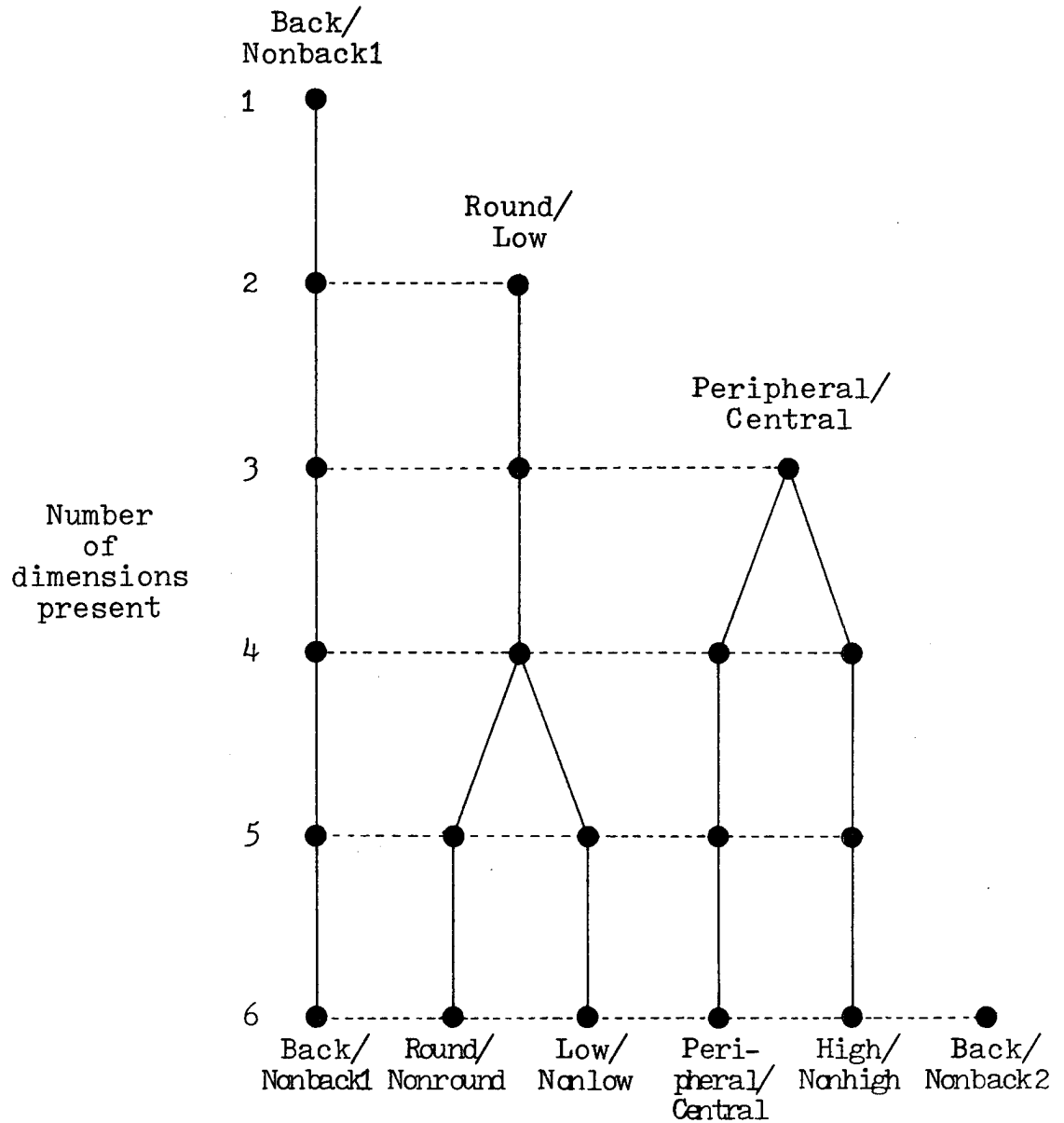


Figure 5.24: Schematic relationship of dimensions found in PARAFAC solutions for all 35 listeners, in 1 through 6 dimensions.

might have emerged. But such dimensions, if they exist, are not accessible by the data at hand.

5.4 Personal variation within the 6-D PARAFAC solution

The relationship between vowel dimensions and listeners can now be examined by referring to the weighting factors which give the relative salience of each dimension for each person. Interpreting the patterns of personal weights requires relating those patterns to known information about the listeners, namely, what native languages they speak. While little can be said about the patterns representing each individual, it is still possible to discover to what extent listeners' weights pattern according to native language.

The personal weights (Appendix II) have been normalized. If the original PARAFAC value on the j th dimension for the i th person is d_{ij} , then the normalized weight w_{ij} equals $(d_{ij}/\sum d_j)^{\frac{1}{2}}$. This effects two changes. For one, the square roots of the PARAFAC values are interpreted as the true weights, not the values themselves. Second, the weights for each person are reduced to the same scale size (sum of the squared weights = 1). The fact that a difference in overall magnitude of the weights did occur from one listener to the next is directly attributable to the additive constants used to prepare the data for PARAFAC. The 6-D additive constants correlate very highly ($r = .967$) with $(\sum d_j)^{\frac{1}{2}}$, the square root of each listener's sum of squared weights. It is difficult to attach any further significance to the absolute magnitude differences. It certainly makes no sense to say that one listener has a perceptual space of a larger magnitude than another.

It is no accident that the following discussion is devoid of statistical tests. The weighting factors are formally amenable to two-way analyses of variance, but they cannot meet the assumptions for a reliable F test as described by Hays (1963), Chapter 13. For one thing, the values are not independent across rows, due to the normalization. Another problem is that the fewer values in each group, the more crucial becomes the assumption of a normal distribution for the variance of each group. These two objections are not serious, since the unnormalized data could be used, and there is no evidence to show that the variances are not normally distributed. But for small numbers of observations (fewer than ten per group) it is very important that all groups are of equal size, which is not the case with the present data. Hays does not even give the computations necessary to handle unequal size groups of small n .

Figure 5.25 plots the normalized weighting factors, each circle representing one listener's weight on a particular dimension. Reading horizontally, one can compare the relative salience of the six dimensions within one language.

The English listeners show clear preferences for some scales and not others. The most salient dimension is rounding, with Back/Nonback1 somewhat less significant. A redundancy rule for English which predicts the rounding of [ʊ] and [ɔ] from the fact that they are Nonlow and Back is not a plausible reflection of a listener's perceptual strategy, no matter how satisfying the rule is on phonological grounds. The heavy weighting of rounding itself suggests the prediction of backness, given rounding.

Back/Nonback2 is somewhat less salient than Back/Nonback1. It is curious that English listeners should use Back/Nonback2 at all. Perhaps, since English subjects do use F3 to identify [ɹ], they may be able to use it in distinguishing [ɹ] from [ø] and [æ], even though they declared the stimulus [ɹ] to be definitely non-native. This would show up in the use of the Back/Nonback2 weights.

Of the remaining three dimensions, Low/Nonlow has the highest weights, with High/Nonhigh values considerably less. This may be weakly related to some facts about English phonetics. Considering only native vowels which have counterparts in the stimuli, namely [e i o u ɔ], at least two features set off all low vowels from non-low vowels. One is lowness itself, based on the first formant. Beyond that, [i e o u] often have high glides, and [ɔ] has retroflexion as well. There is no extra criterion that augments the separation of high and mid vowels. On the other hand, the "mid" stimuli [e o] are relatively high, and the large acoustic difference between low and mid stimuli may well have biased to listeners towards Low/Nonlow. The medium-sized weights on Peripheral/Central do not lend themselves to any interpretation.

The German listeners value the scales of Back/Nonback1, Round/Nonround and Low/Nonlow most highly. High/Nonhigh is slightly less salient. But Back/Nonback2 and Peripheral/Central are clearly the least salient scales. All values on both these dimensions are below .33, to which level only three points on the other four scales descend.

Back/Nonback2 is predictably unimportant as a dimension; Germans have no [ɹ] and therefore no reason to distinguish it from [ø], which this scale aids in doing. It is difficult to give a direct reason for the low value on Peripheral/Central, because it is not known what causes high values. If it is correctly interpreted as a component of a curved

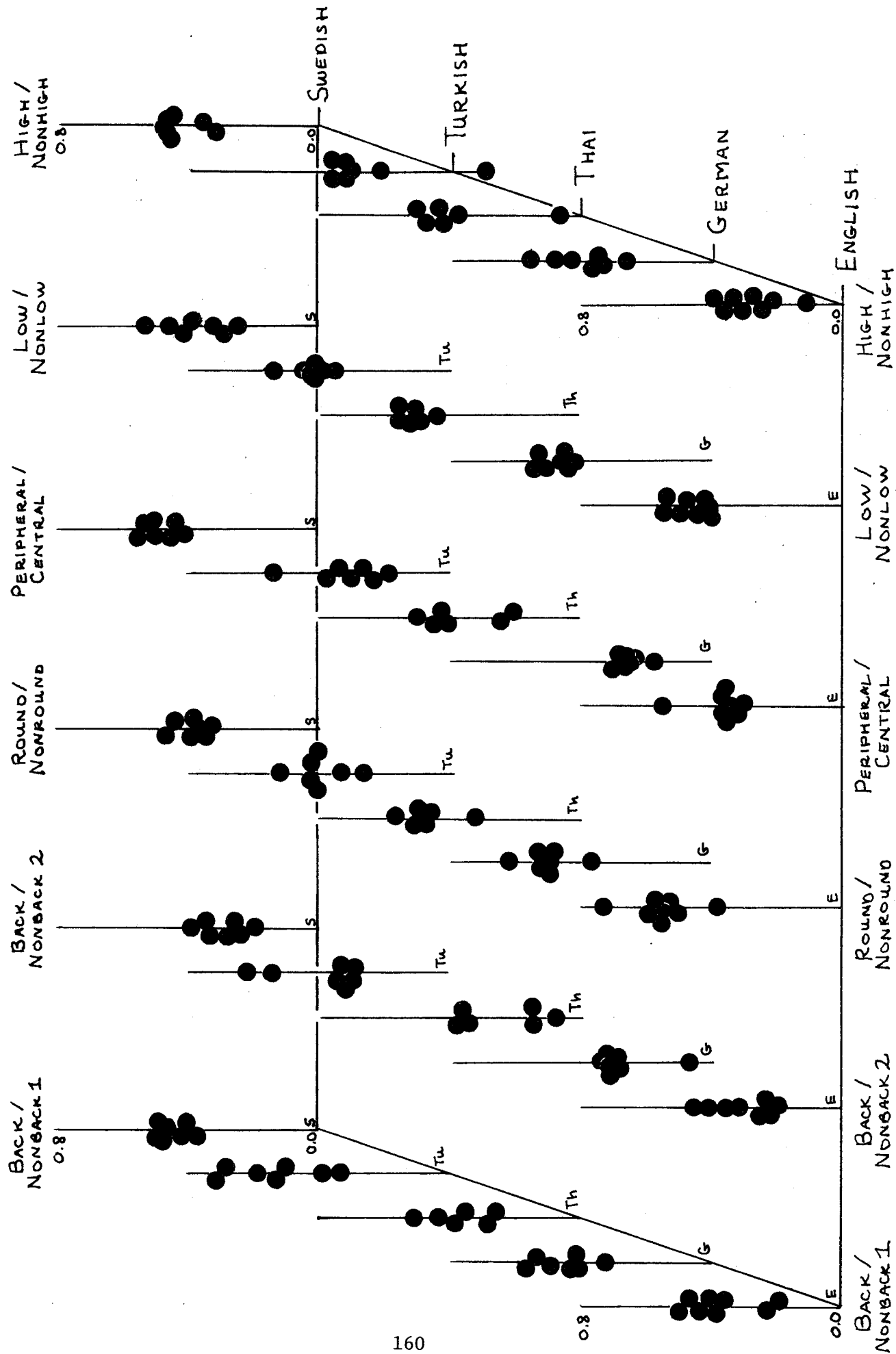


FIGURE 5.25 PERSONAL WEIGHTING FACTORS FOR THE 35 SUBJECT SOLUTION

Rounding dimension, then the Germans' evaluation of distances from - to + round is not subject to much underestimation; rounding is closer to a linear scale for German listeners. If Peripheral/Nonperipheral is a separate dimension, the conclusion is that Germans pay relatively little attention to whether a vowel is on the periphery of the F1-F2 plane or not.

For the Turkish subjects, the obvious topic of interest is the relation between the two Backness dimensions. Disappointingly, only two listeners have very high loadings on Back/Nonback2. However, they were the only two listeners who identified the stimulus [ɤ] as a Turkish [ɤ], so the connection between this scale and this one vowel remains valid. Furthermore, subject K6 has a much higher weight on Back/Nonback2 than on Back/Nonback1 (.62 and .34, respectively), while the other subject with a high Back/Nonback2 value (.54) has a Back/Nonback1 value almost as great (.50). All other subjects show greater use of the first backness scale than the second. Personal strategies obviously play a large part in vowel identification.

Rounding, for Turkish listeners, is slightly more salient than Peripheral/Central, and Low/Nonlow is comfortably more salient than High/Nonhigh. Phonologically, Turkish has but two vowel heights, separating [ɪy ɤ u] from [e ø ɑ ɔ]. There is simply no need for Turkish listeners to be able to make judgments on two binary height dimensions. The one salient dimension is expectably Low/Nonlow, being more simply related to F1, the main acoustic correlate of vowel height.

One listener (K6) showed a negative weight in the original PARAFAC output, expressed in the normalized form as an imaginary number. This point appears below the base line on Figure 5.25. The fact that only one such value appeared in the entire 6-D solution, together with the fact that the absolute value of the imaginary weight is small, suggests that the incompatibility of model and data is truly minor.

The patterns of variation for Thai and Swedish are not as clearly defined. One very likely reason is the low reliability of the Thai and Swedish subject groups. The mean χ^2 for subject reliability (c.f. Table 4.2) is 27 and 26 for Thai and Swedish, respectively, where the means for English, German and Turkish are 43, 52, and 48, respectively.

With the Thais, Round/Nonround, Low/Nonlow, and High/Nonhigh have very similar values, ignoring one listener's extremely low weight on High/Nonhigh. Back/Nonback has about the same range of values as Peripheral/Central. The differences among these dimensions are not interpretable. The only

interesting deviation is found in the Back/Nonback2 scale, which has no values above .4, and three values below .15. It is tempting to claim that this is the least salient of all the six scales for the Thais, but that claim does not hold when the subjects' weights are examined individually (c.f. Appendix II). For subjects T1, T4 and T5, Back/Nonback2 does indeed show the smallest weight. But T2, T3 and T6 show this scale to be one of the more salient ones. In fact, for T6 and T3, Back/Nonback2 is at least as salient as Back/Nonback1. It would be interesting to see, with more subjects and a different set of stimuli, whether or not there really is a bimodal distribution on this scale, and whether or not Thais in fact use this dimension to identify their [ɛ] vowel.

Swedish listeners' weighting factors show even less variation than the Thais'. One curious fact is that the Swedish subjects emphasize Peripheral/Central to a slightly greater extent than they do rounding. We have no explanation for this. The only other noteworthy property of the Swedish weights is that Back/Nonback2 weights are smaller than those for Back/Nonback1 for every subject. Swedish, having no [ɛ], has no need to rely on Back/Nonback2.

The most immediate question to be asked of these results is whether they support Hypothesis 1 or 2 as proposed in Chapter 2. According to Hypothesis 1, all languages use the same dimensions, and native language does not influence the importance a listener attaches to a dimension. This latter claim appears to be true of some dimensions and not others. Low/Nonlow, for example, shows very little difference in personal weightings from one language to another (see Figure 5.25). The values for Round/Nonround are also reasonably uniform across languages. But this is not true of Peripheral/Central. All Swedish values on this scale are greater than the German values; the English and German distributions overlap by only one point. With the exception of one Turkish listener's outlying value, the Swedish and Turkish distributions do not overlap, either. High/Nonhigh also has some unevenness in its distribution of points; English values are generally lower than those for the other languages.

Although the distribution of weights on Peripheral/Central and High/Nonhigh are difficult to interpret, other differences are not. The greatest weights on Back/Nonback1 are recorded by the Turkish and the German listeners; this is in keeping with the hypothesis proposed in Chapter 4, that backness harmony and umlaut, respectively, can cause a heightened awareness of backness relationships. Finally, the salience of Back/Nonback2 is clearly influenced by native language: the two listeners who heard [ɛ] as native placed more weight on this dimension than did any other listener.

Hypothesis 2, which claimed that native language does influence the weighting factors, matches the results more closely than the first hypothesis. Yet both H1 and H2 assume a common set of dimensions for all languages, an assumption implicit in this chapter's analysis. A third hypothesis, H3, holds that there may be dimensions specific to particular languages. H3 is examined directly in Chapter 6.

Chapter 6

PARAFAC and PARAFAC2 Solutions for Individual Languages

6.0 The analysis in Chapter 5 required the assumption that all 35 listeners used the same set of dimensions, which varied in relative importance from one listener to the next. But there might be certain dimensions, or certain rotations of planes, that are peculiar to specific languages, as proposed by Hypothesis 3 (Section 2.4). To investigate this possibility, PARAFAC analyses were run, taking the languages one at a time. This also afforded the opportunity to see how Swedish listeners treated [æ] and [œ], and how all the others treated [ʌ] and [ɑ]. These solutions are discussed in Sections 6.1 through 6.5.

For the German, Turkish, Thai and English data, solutions will be presented for 4, 4, 3, and 5 dimensions, respectively. The Swedish results were very difficult to interpret at first, no matter how many dimensions were chosen. Some earlier work with synthetic data (Terbeek and Harshman, 1972) suggested the use of PARAFAC2, which not only solves for the vowel co-ordinates but finds the best-fitting oblique rotation of axes. Not only was the PARAFAC2 solution in three dimensions easy to interpret, but it showed how one Swedish subject was the source of the difficulty in interpreting the original PARAFAC solutions. Data for other languages were also submitted to PARAFAC2, but the results were no more insightful than the PARAFAC solutions discussed below. Appendix II contains the co-ordinates for the solutions discussed.

Regression analyses are used, as before, to relate the perceptual dimensions to the acoustic properties of the stimuli. The following strategy has been followed. Regressions are first attempted using F1, F2 and F3 only. If the first three formants cannot predict the scale at hand with an r of .90 or better, then F0, A0, D, A1, A2, A3 are included as well. Since 12 vowels are being considered, there is no difficulty in performing regressions with 9 variables. Naturally, the more variables used, the better the prediction. It is prudent to be skeptical of predictions involving more than 2 or 3 variables.

In Chapter 5, the standard scores for the frequencies and levels were used as the regression data. But it was found that the regression coefficients were more interpretable if they were expressed in terms of more familiar absolute values, not standard scores. Therefore, absolute values of frequencies and levels are used here (Table 3.3).

6.1 German

The most satisfying analysis of the German listeners' responses is the 4-D solution (c.f. Figure 6.1). The two 5-D solutions each obtained patterns similar to the 4-D scales, plus one uninterpretable dimension. The 1-D solution combined both backness and height. In two dimensions, backness emerged independently, but the second scale was a mixture of height and rounding. The High/Nonhigh dimension appeared only in the 3- and 4-D spaces. The remaining two dimensions, Low/Nonlow and Round/Nonround, did not emerge independently until four dimensions were extracted. The input/output ratio of the 4-D solution is 6.1, slightly better than that for the combined language space in six dimensions.

The High/Nonhigh and Low/Nonlow dimensions can be interpreted as a plane containing one curved dimension of vowel height, as shown in Figure 6.2. This curve (like the ones in Chapter 5 and those to appear later in this chapter) was fitted by sight using a flexible edge, which was then marked and straightened in order to measure the vowel distances. Notice that it is possible to find a curve (i.e., Figure 6.2) which fits the vowel positions much better than was possible for all languages taken together. For the German listeners, the amount of curvature is the same for front and back vowels. [u] and [ʊ] are not displaced towards the interior of the space.

The space can therefore be thought of as having three perceptual dimensions. For two of them, backness and rounding, vowel distances can be accounted for by simple, Euclidean co-ordinates; for the third, height, perceived distances are based on a more complex pattern -- a curve.

Figure 6.3 shows the stimulus vowels plotted on these three dimensions, height having been straightened for simplicity. Native-like vowels are circles, non-native vowels are squares. It is not difficult to assign binary feature values to the native vowels. Figure 6.3a, height vs. backness, shows a clear distinction between high [i y u] and all other vowels. (This is also apparent from Figure 6.2.) Backness, ignoring for the moment [æ], shows an obvious cut between front [i y e ø] and back [u o ɔ]. If [æ] is a native vowel, then the backness scale must be regarded as ternary. But the stimulus [æ], which listeners took to be an example of [ɛ:], is too low and too far back for a good native [ɛ:]. [æ] does no serious damage to the claim of a binary backness scale.

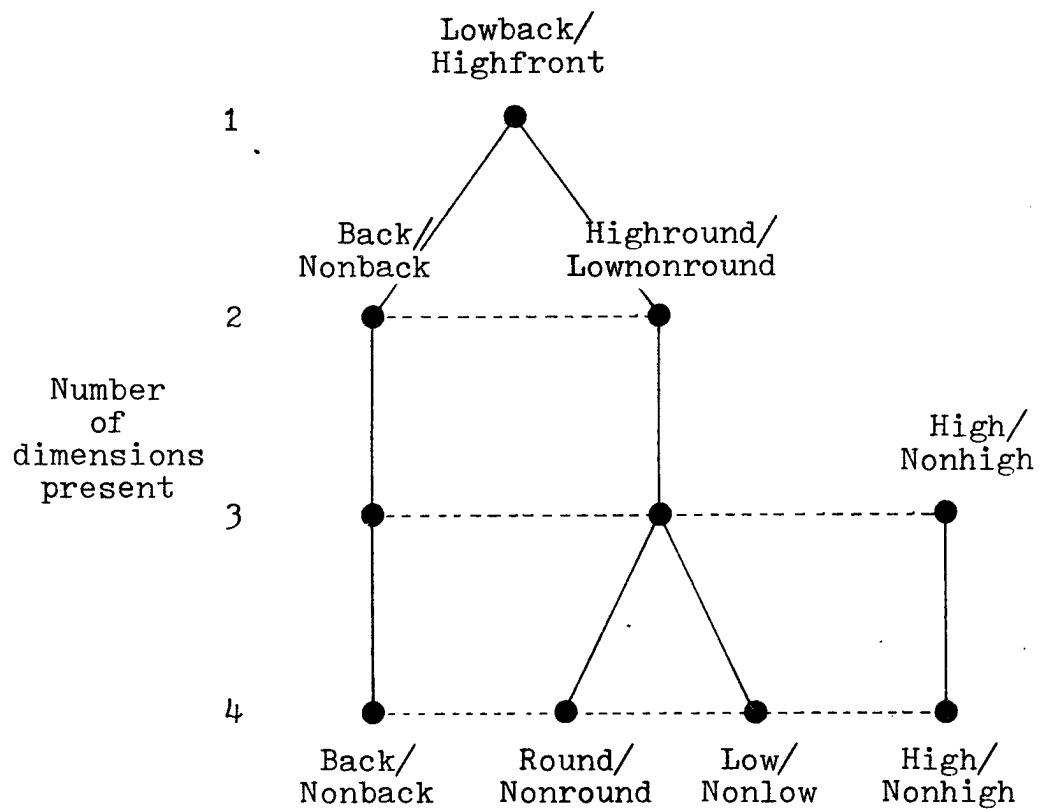


Figure 6.1: Summary of the PARAFAC solutions for German in 1-4 dimensions.

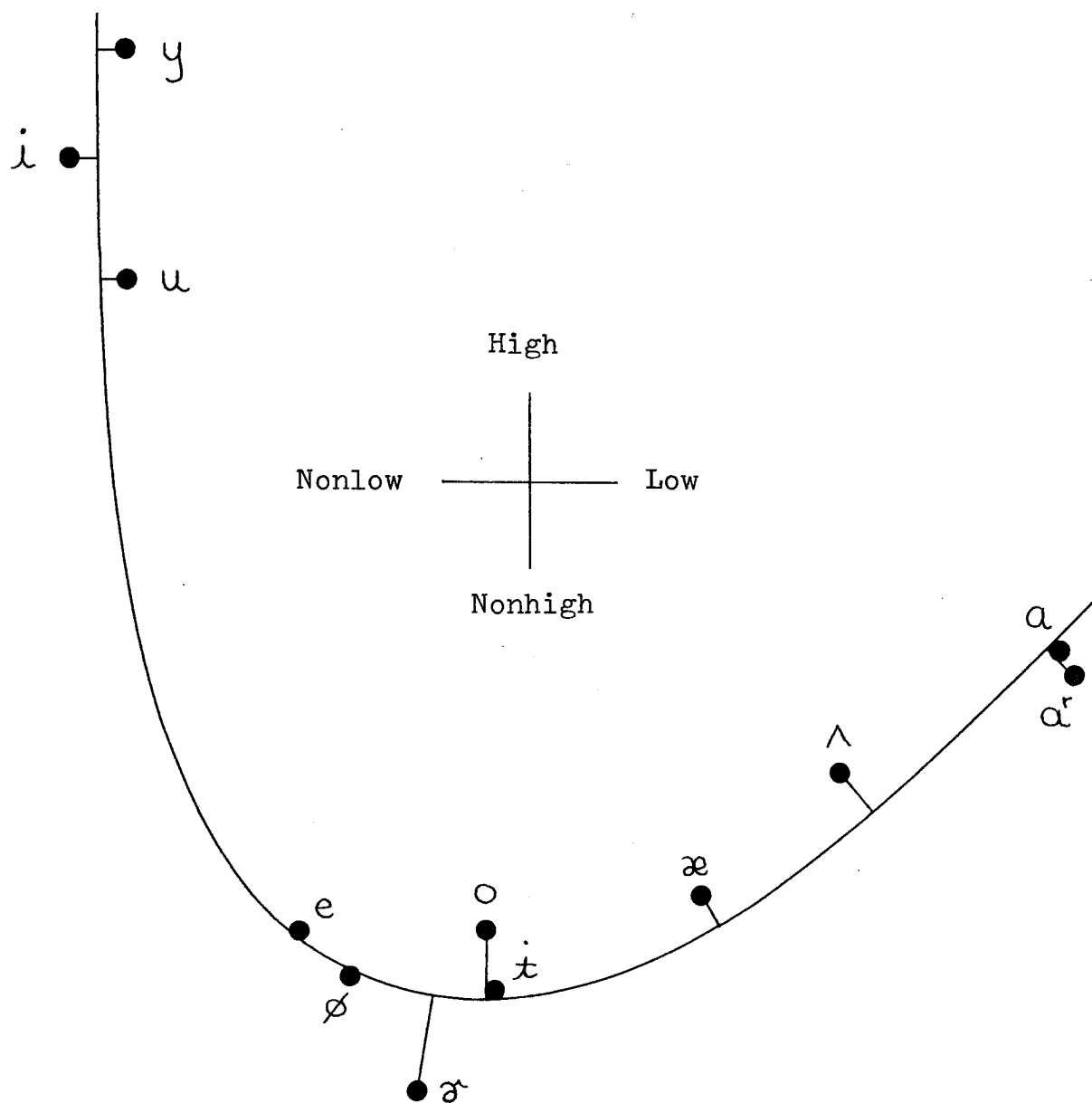


Figure 6.2: High/Nonhigh and Low/Nonlow from the 4D solution for German, with curved line fitted by sight.

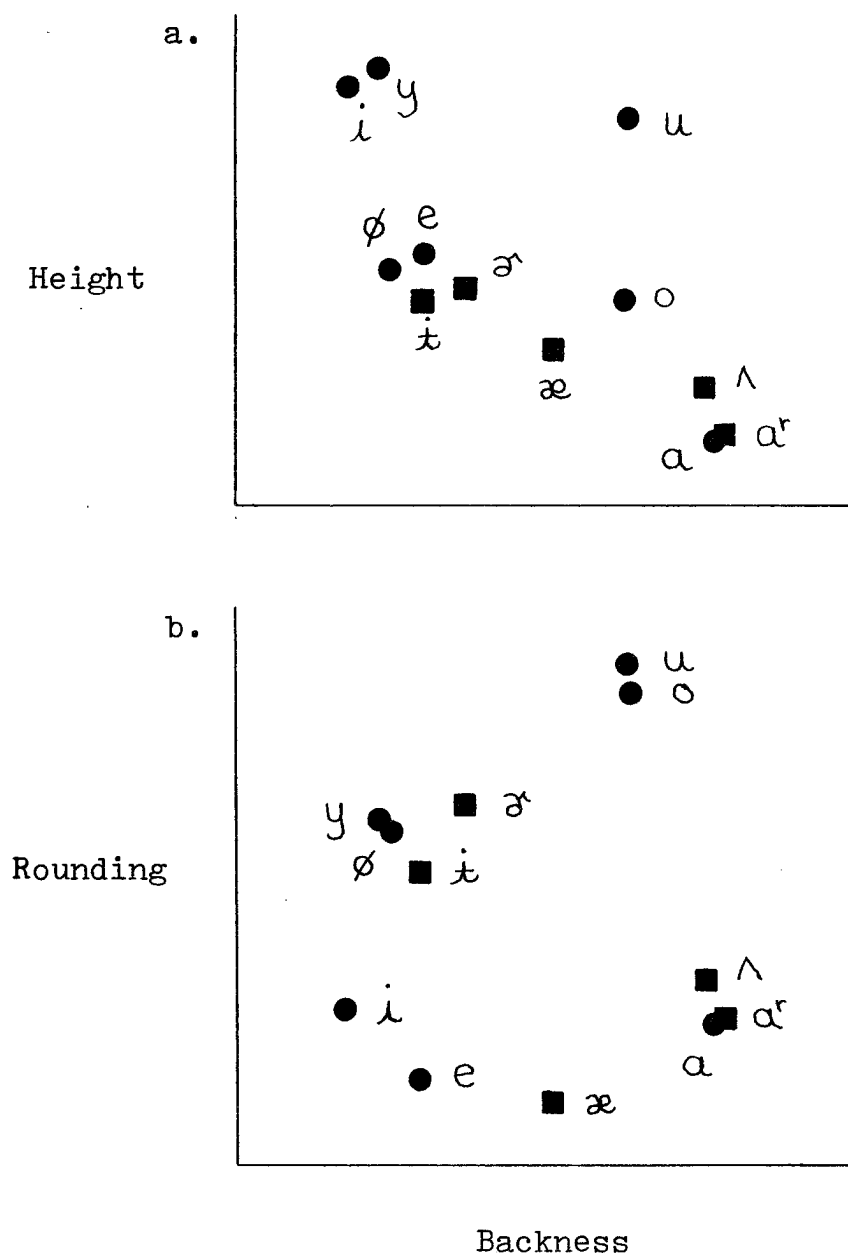


Figure 6.3: Back/Nonback, Round/Nonround, and straightened Height curve from the 4D PARAFAC solution for German.
 ● = native vowels; ■ = non-native.

Rounding, plotted in Figure 6.3b, is a binary scale, too. The best cut between + and - round is made at one value for front vowels, and at a higher value for back vowels. This suggests that rounding decisions will be more reliable if backness is assigned first.

In order to discover the acoustic basis for these dimensions (or features), stepwise regressions of acoustic measures against the perceptual scales were performed. The results are summarized in Figure 6.4. Each successively higher dot shows the value of the multiple correlation coefficient obtained when the next best predictor is included among the variables used. Some predictions of perceptual scales are shown in more detail in Figure 6.5. The labels at the left of each graph are relative coefficients. The constant (y-intercept) term is ignored, and the coefficients are scaled so that one of them is 1. Coefficients for regressions using many variables are not given.

Back/Nonback is based mainly on F2 ($r = .77$), but the contribution of F1 brings the correlation up to .95, which is too striking an improvement to ignore. Back/Nonback, for the German listeners, is a weighted difference between the first and second formants (Figure 6.5a). The coefficient of .31 for F2 is very similar to the ratio of the frequency ranges for F1 and F2. For the stimulus vowels, the range of F1 divided by the range of F2 is 1520/395, or .26. Perceptually, the two formant frequency ranges are regarded as though they both were drawn to the same scale.

Two vowels are somewhat out of place on this dimension: [ʌ] and [ɤ], neither of them native German vowels. [ʌ] is still grouped with back vowels and is perceptually closer to [ɑ] than [u], presumably because of its higher intrinsic amplitude. [ɤ] is shifted towards the front group. It might well be due to the listeners' identification of [ɤ] as [ø], placing it more towards [ø] on the Backness scale.

Round/Nonround is clearly seen to be a function of F1 and F2 ($r = .97$). In this case, the function is a weighted sum, not a difference (Figure 6.5b). As with Back/Nonback, the F2 values are scaled down so that the range of F2 variation is perceptually equivalent to that of F1. The most interesting difference between this scale and Round/Nonround dimension in Chapter 5 is that [y] is grouped with [ɤ & ø], not with [u & o] (Figure 6.5b). Here there is no need to postulate any special processing for the [y] vowel; its position is based on the weighted sum of its first and second formant frequencies, nothing more.

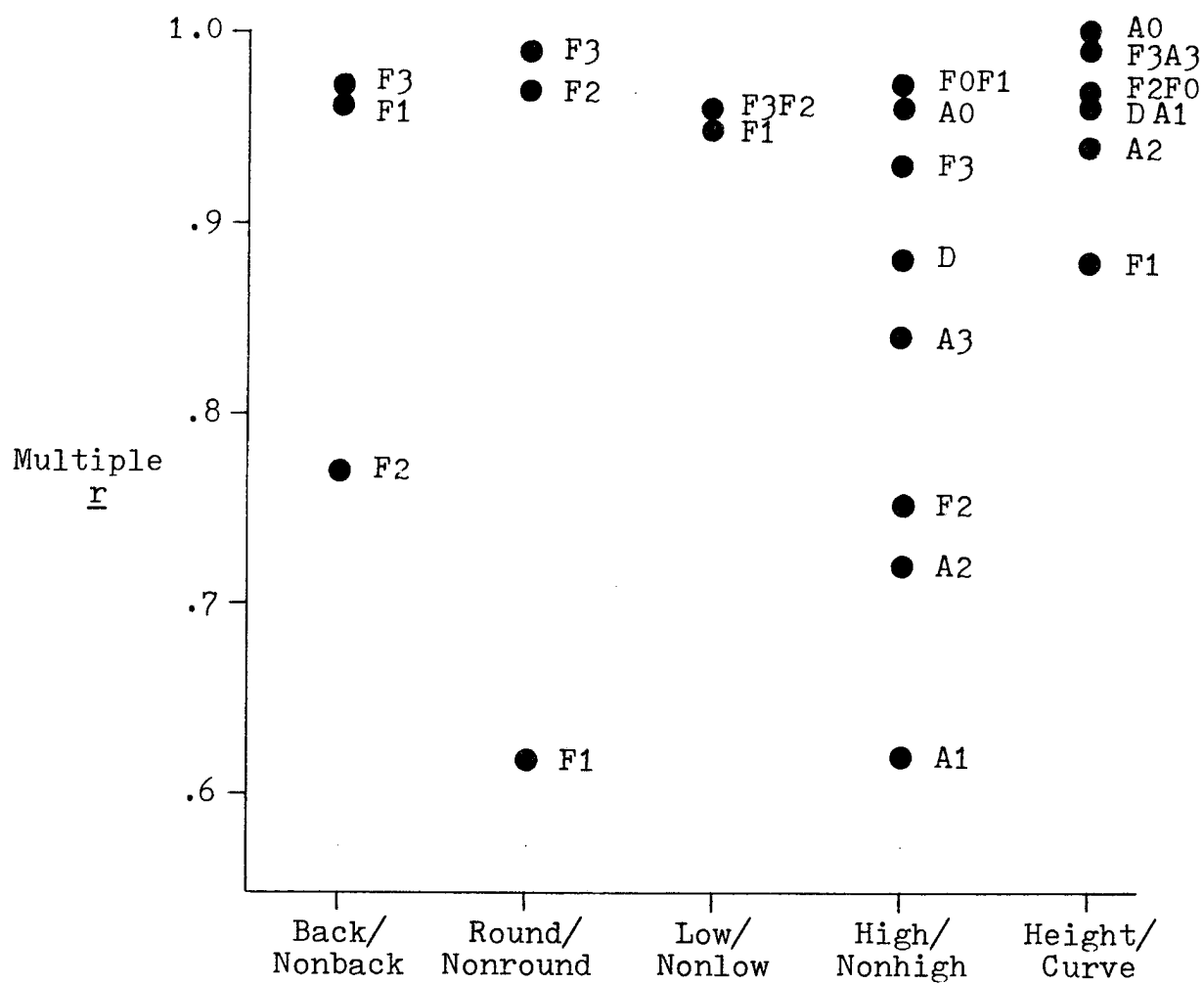


Figure 6.4: Summary of regression analyses, 4D PARAFAC solution for 7 German listeners.

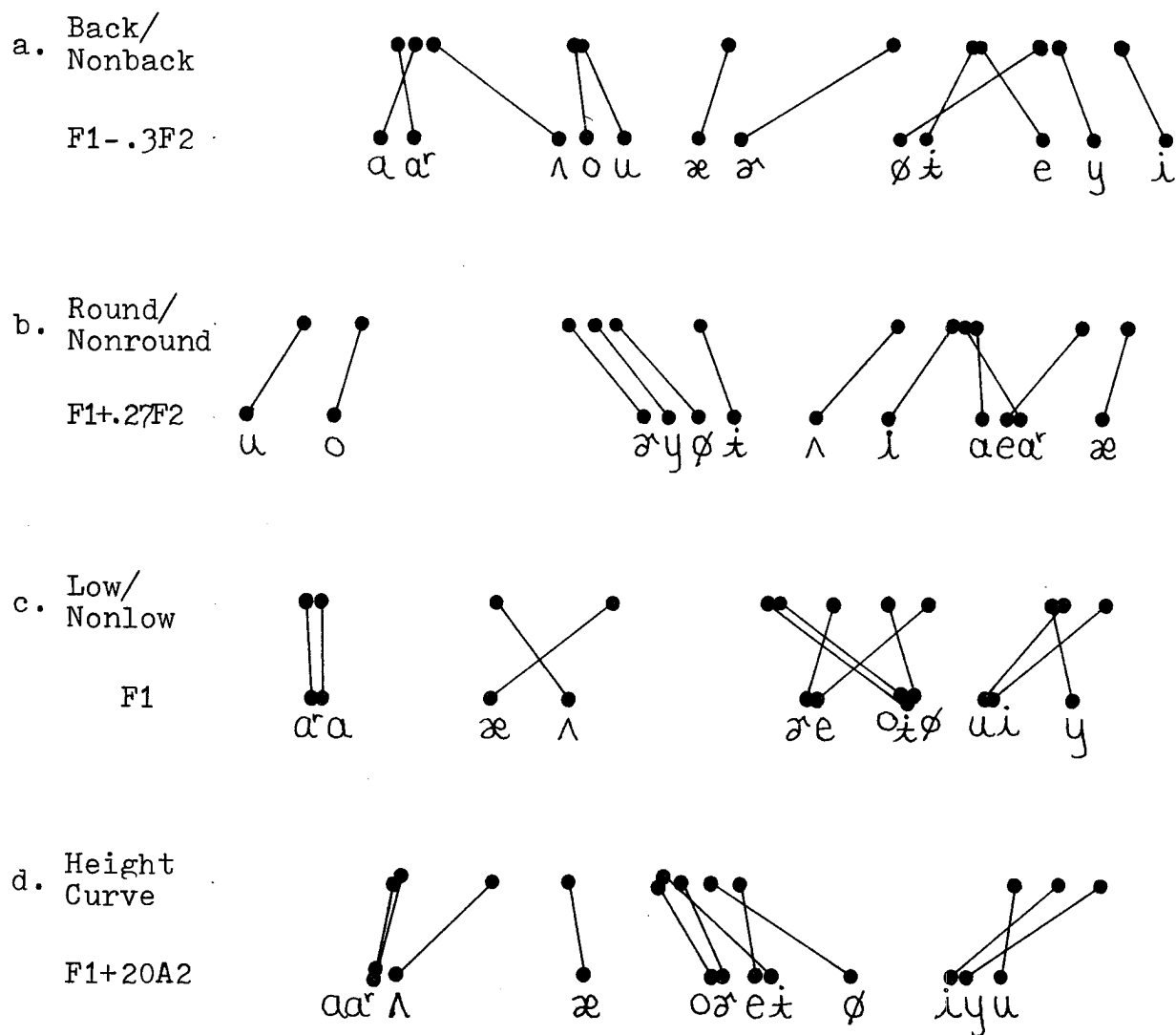


Figure 6.5: Selected comparisons of German PARAFAC dimensions with their predictions by acoustic variables.

Perhaps it should be mentioned that the label "round" is not correct from the viewpoint of articulation. [ʌ] was not rounded at all, and [ɤ] is only incidentally rounded, as has been pointed out earlier. The perceptual dimension is clearly an acoustic one, whose interpretation is not enhanced by any appeal to articulatory activity. Nevertheless, the label "round" is a convenient one. It is, after all, a correct description for the native vowels.

Low/Nonlow (Figure 6.5c) correlates very highly with F1 alone ($r = .95$). The major gaps separating [ɑ ɔ], [ʌ ɤ], [ɛ ɔ ɛ e o] and [i u y] are present in F1 as well. High/Nonhigh, however, cannot reach an r of .9 even with five variables considered. And only at six variables does the prediction begin to account for the main separation of the vowels [i u y] from all the others. It is unlikely that High/Nonhigh values are actually computed in this way by the listeners.

The Height curve is accounted for mainly by F1 ($r = .88$), as expected, with some improvement brought about by A2 ($r = .94$). It happens that A2 values for [i u y] are lower than for the other 10 vowels, but only by a slim margin. A2 is not the source for the large gap setting off [i u y], nor is any other acoustic measure. Perhaps that gap is a result of some strategy applied to F1 itself: vowels with extremely low F1 are classed as "extreme" in addition to [-Low]. At any rate, the main conclusion is that F1 is a perceptual scale for these listeners, with high vowels being perceived as even farther from mid and low vowels than they are according to F1.

There is one final point to be made about the 4-D PARAFAC solution for German. Two person weights are imaginary; that is, the squares of those weights, printed by PARAFAC, are negative. These two values are not small (c.f. Appendix II) and cannot be ignored. Four courses are open.

1) The loadings can be interpreted. The interpretation of an imaginary weight is easy to describe: vowels which lie close to one another on the relevant perceptual dimension (in this case Low/Nonlow) are judged by the subject as far apart; distant vowels are judged to be similar. But this is really an admission that there is no interpretation, that the person's data do not conform to the distance model upon which the experiment is based. This leads to the second possibility.

2) The subjects' data can be discarded as anomalous. This was done, and the remaining five subjects' data were analyzed by PARAFAC. The 3-D solution, the last before

multiple solutions occurred, consisted of dimensions essentially identical to the Round/Nonround, High/Nonhigh, and Back/Nonback scales of the present 4-D space. All person loadings were acceptable. But this 3-D solution fails to capture an interesting fact: 5 of the 7 listeners do assess F1, that is Low/Nonlow, in a way perfectly compatible with the three-mode scaling model.

3) The third possibility is to retain 7 subjects, but to discard the 4-D solution in favor of the 3-D solution. There the person weights are all real. But it has already been shown that the 3-D space is difficult to interpret, confusing as it does Rounding and Lowness.

4) The course actually taken has been to recognize two subjects' behavior on one dimension as anomalous and inexplicable, but to retain the solution itself as the best reflection of the remaining five listeners' behavior.

At this point it looks as though the analysis of the German data has confirmed the predictions made by the 35 person study of Chapter 5. The four dimensions of the 6-D solution which the Germans weighted most heavily have emerged as the only four interpretable scales. The Turkish results, discussed immediately below, will behave the same way. But the results from Thai, English and Swedish will not, making a stronger case for the language-specific Hypothesis 3.

6.2 Turkish

No more than four dimensions can be reasonably extracted for the Turkish data; there are multiple solutions in five dimensions. Again, as with the German analysis, the input/output ratio (6.1) is no worse than for the 6-D space of Chapter 5. Unlike the case of German, all person weights are acceptable. Considering the vowel space, Figure 6.6 shows that Back/Nonback1 emerges in every solution, while Back/Nonback2 emerges in the 3- and 4-D spaces. High/Low and Round/Nonround appear as one collapsed dimension in the 2- and 3-D solutions, emerging independently only in the 4-D space.

The two backness dimensions define a curved backness dimension, shown in Figure 6.7. A curve passing between [æ] and [ɨ], rather than through [ɨ] would be better, in retrospect, than the curve given. This has no effect on the arguments referring to the backness curve. On Back/Nonback1, [æ] and [ɨ] appear with front vowels; on Back/Nonback2, they group with back vowels. Along the curve, [æ] and [ɨ] form a third, middle group. [æ] is not a native Turkish vowel, although some listeners heard it as a very open (or "nasal") [ɛ].

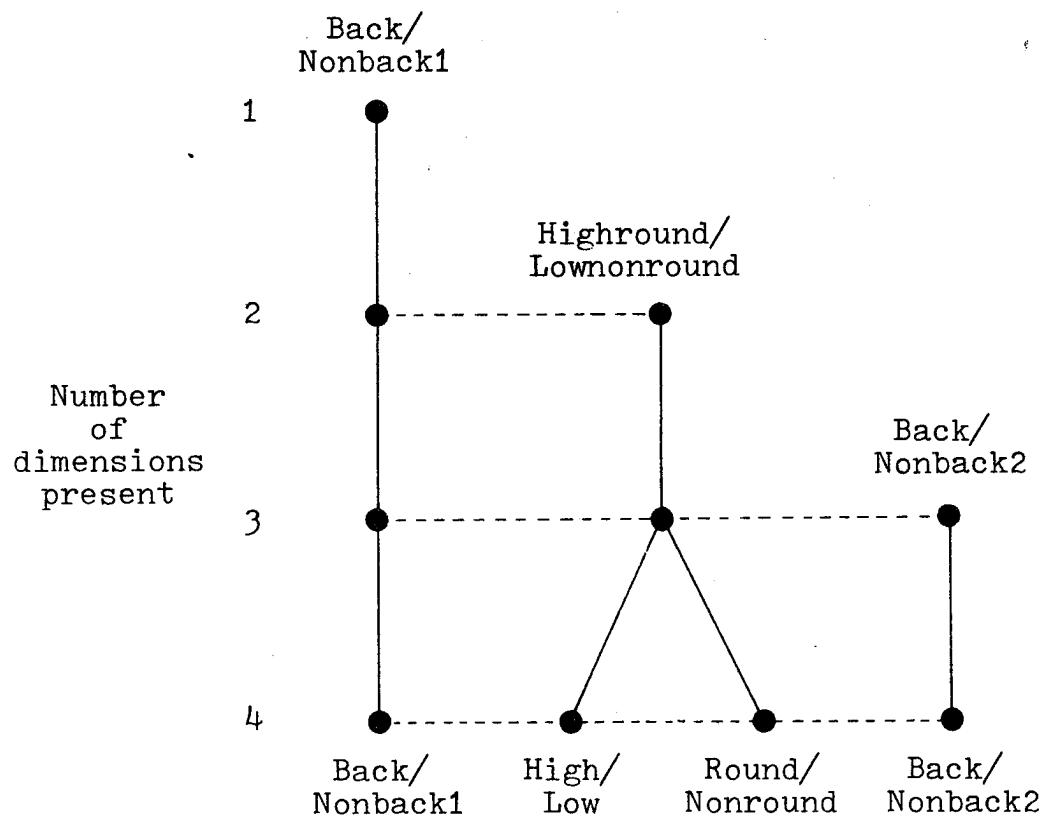


Figure 6.6: Summary of the PARAFAC solutions for Turkish in 1-4 dimensions.

Figure 6.8 shows Height and Rounding plotted against the straightened Backness curve. The Rounding scale (Figure 6.8b) affords a clear binary grouping of vowels; [ɔ̞ ɤ̞ y̞ ɔ̞ ʊ̞] are [+Round], and [e̞ i̞ æ̞ ʌ̞ ɑ̞ ɔ̞] are [-Round]. [ɤ̞] groups with rounded vowels. This problem was discussed earlier in Section 4.3.

The Height dimension (Figure 6.8a) requires three feature values for all the native-like stimuli. [ʊ̞ ɔ̞ ɑ̞] must be high, mid, and low, respectively. For the Front vowels, only two heights are needed to distinguish high [i̞ y̞] from non-high [e̞ ɤ̞]. An appropriate Height value for [ɤ̞] might be either mid, relating its actual perceived height to that of [e̞, ɔ̞] and [ɤ̞], or high, reflecting its status as the highest central vowel.

The regression analyses for Turkish are not straightforward. None of the dimensions are satisfactorily predicted by a simple combination of acoustic measures, as summarized in Figure 6.9. Back/Nonback1, an obviously binary scale, correlates reasonably well ($r = .94$) with a linear combination of F1 and F2, but those measures do not account for the large gap between the perceptually back and front groups (Figure 6.10a). F2 alone does order all the vowels so that a division into Back and Nonback could be made (by a cut between [ɑ̞] and [ɔ̞]), but the large gap found in the perceptual scale is not present. The addition of F1 to the regression equation serves to join the back vowels more tightly together and move [ɔ̞] somewhat towards the front vowels. But the resulting small distance between [ʌ̞] and [ɔ̞] is a tenuous source of a large perceptual gap.

It is possible to resolve this difficulty by proposing the following line of reasoning concerning [ɔ̞]. Since [ɔ̞] is non-native, the listeners are unused to hearing its low third formant (1825 Hz), invading as it does the expected range of F2. The apparent F2 might therefore lie between the true F2 (1305 Hz) and the true F3. If the apparent F2 is estimated as the mean of F2 and F3, or 1565 Hz, then [ɔ̞] would in fact group with front vowels. Open circles mark its proposed place on Figure 6.10a. With this solution, $F2 - 1.5F1$ is an adequate source for Back/Nonback1.

Back/Nonback2 is even more difficult to account for in terms of linear combinations of acoustic variables. The best two-variable prediction, $F2 - 1.3F3$ (Figure 6.10b), finds [æ̞], [ɑ̞], and [ɤ̞] out of place. Adding a third variable, duration, brings the correlation to .91 (Figure 6.9), and improves the relative spacing of the vowels considerably. The price paid is the unfortunate claim that listeners were assessing vowel duration as part of their backness estimates. One problem is

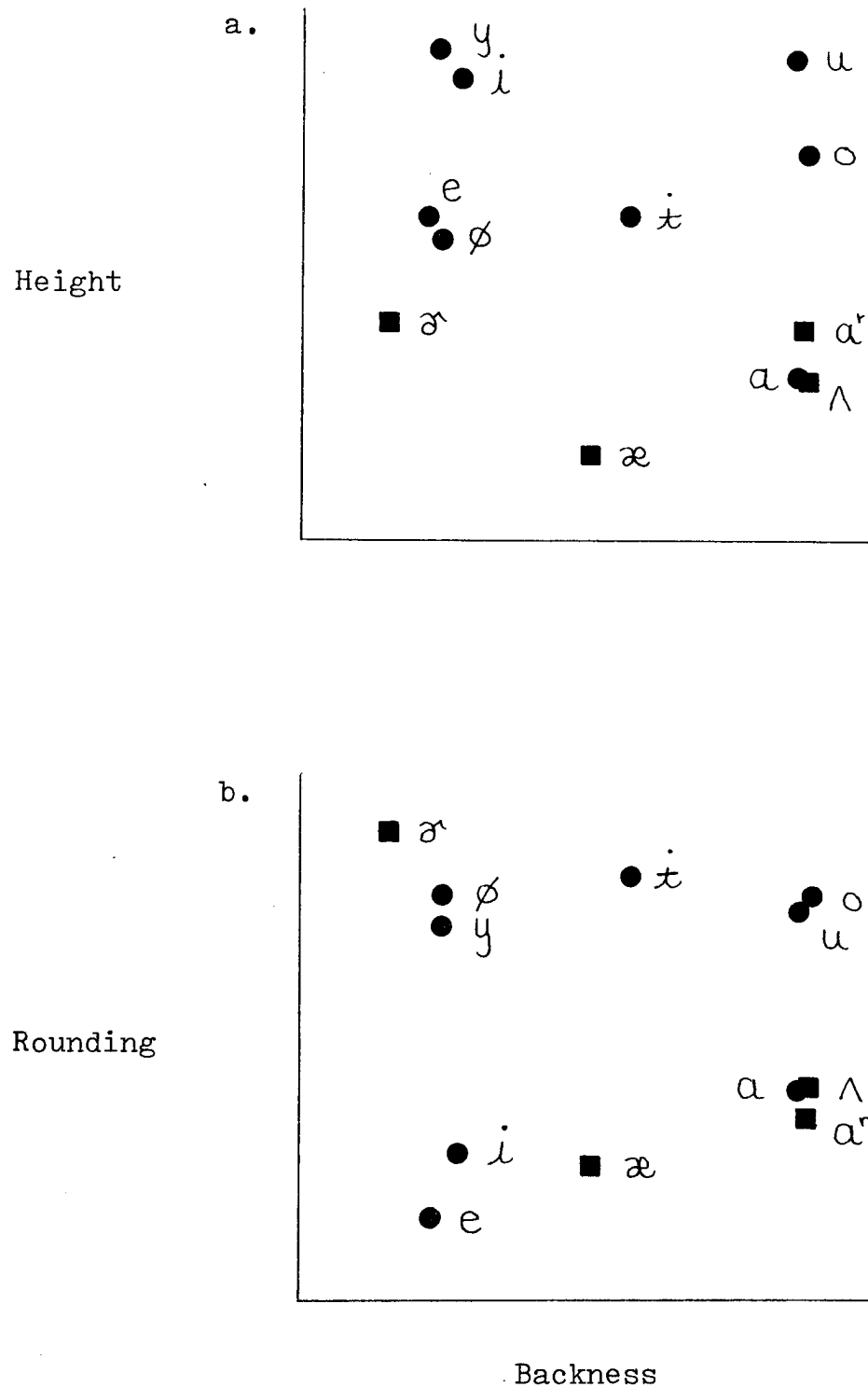


Figure 6.8: Round/Nonround, High/Low, and straightened Backness curve from the 4D PARAFAC solution for Turkish. ● = native vowels; ■ = non-native.

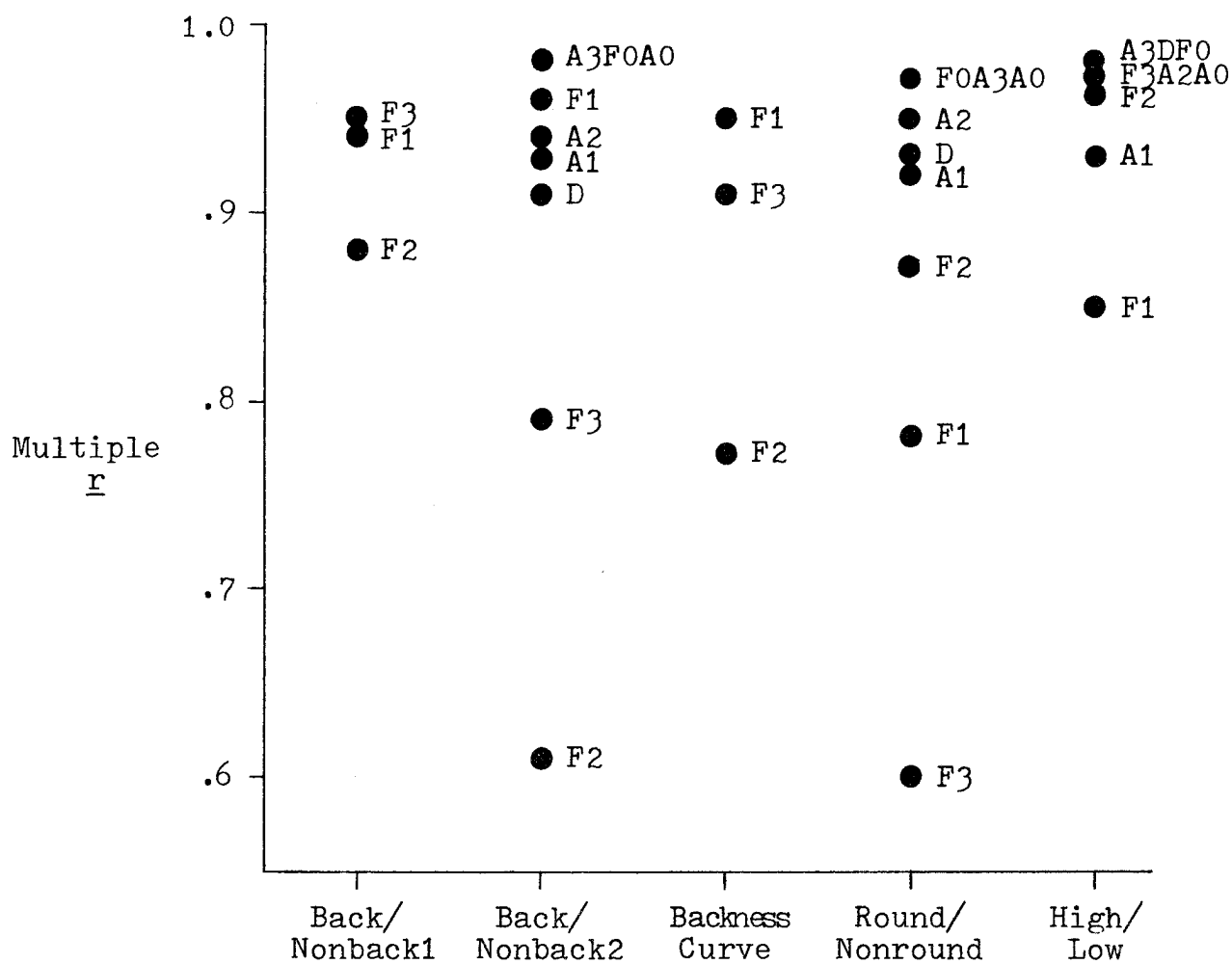


Figure 6.9: Summary of regression analyses, 4D PARAFAC solution for 7 Turkish listeners.

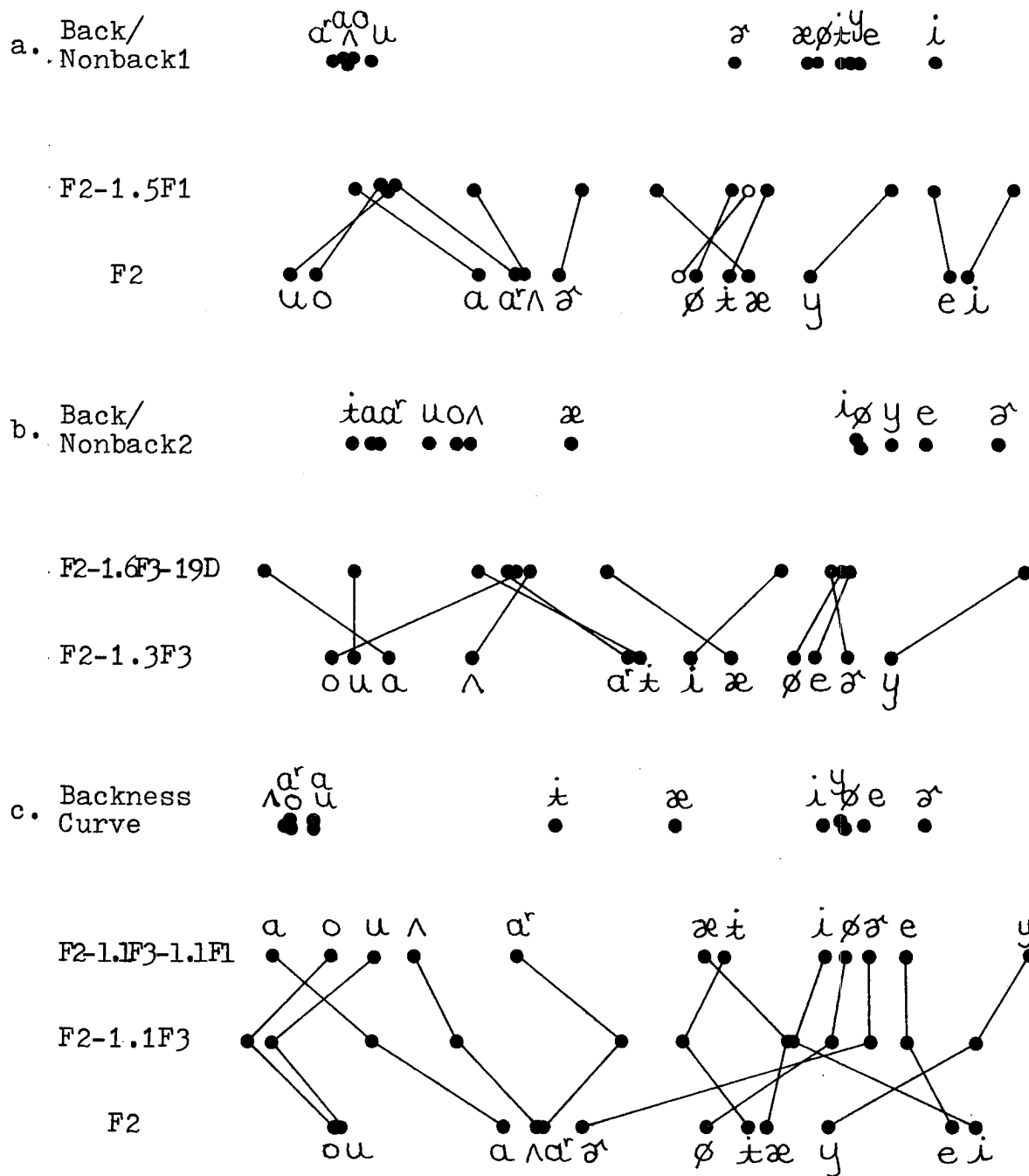


Figure 6.10: Selected comparisons of Turkish PARAFAC dimensions with their predictions by acoustic variables. o in part a shows position of σ based on hypothesized apparent F2 (see text).

that the duration of [ɰ̇] was probably anomalous, as discussed in Section 3.4. This casts doubt on any analyses in which the duration of [ɰ̇] plays an important part.

Again, there are alternate explanations, and one involves the interpretation of the Back/Nonback1 and Back/Nonback2 patterns as one curved scale. The predictions for the curved backness scale are in Figure 6.10c. F1, F2, and F3 together predict the vowel pattern with an r of .95 (Figure 6.9). The gaps that exist in the F1-F2-F3 prediction do match those in the perceptual dimension, separating [ɑ ɔ u ʌ ɔ̣], [æ ɰ̇], and [ɪ ø ɐ e y].

But there is really no reason to assume that each vowel is evaluated according to all three formants in order to decide whether or not it is front, central, or back. Such an assumption is a consequence of the use of linear regressions. There is a certain appeal to another approach, namely, a set of strategies that involve as few acoustic variables as possible. Consider that F2 is the primary source for the placement of vowels on a Backness scale. Then the vowel will appear in four groups (c.f. Figure 6.10a or c):

[u ɔ] [ɑ ʌ ɔ̣] [ɐ ɰ̇ ø æ y] [ɪ e]

[ɐ] is included with the group [ɰ̇ ɰ̇ æ y], according to the above suggestion that [ɐ] is perceived as having a higher F2 than is actually the case. The first two groups are assigned a value of [+Back], the last two, [-Back].

Now if the vowel sound being processed falls into the third group, it is evaluated according to F3 as well: this group is thereby split up into subgroups [ɐ][y ø][æ ɰ̇], listed in ascending order of F3. Apparently, vowels in this set that have a high third formant, namely [æ ɰ̇], are regarded as not front. This may seem strange, especially for [æ]. In the case of [æ], the first formant probably enters in as well. No Turkish front vowel has as high an F1 as the [æ] stimulus, which fact serves as another reason for listeners regarding the vowel as not front. If attention is focussed on the native vowels, though, the function of F3 is really to distinguish [ɰ̇] from [y] and [ø], both of which have lip rounding, and therefore lowered F3 values.

The most satisfying aspect of this proposal, as opposed to one using all linear regressions, is that it describes the specific effects F1 and F3 have on the classification of vowels which cannot be accounted for on the basis of F2 alone.

The next scale to consider, Round/Nonround, finds the vowels in two clusters with a large gap between them (Figure

6.10d). The unsuccessful attempt at matching this vowel pattern with linear regressions is shown as well. The best two-variable prediction is a weighted sum of F1 and F3, with r at only .60, although it at least allows an arbitrary line to be drawn separating round from nonround. Another two-variable prediction, found by trial and error, is $F1 + .25F2$. This, too, allows round and nonround to be distinguished (c.f. Figure 6.10d). But it is difficult to explain why a small gap between [e] and [ɤ] or between [ʌ] and [ɤ] should be the source for the separation into two perceptual groups. The one-variable prediction just begins to account for the binary nature of the scale.

The difficulty disappears as soon as the requirements of a single linear function is dropped. Figure 6.11 shows two equally good ways to group vowels according to the classes round and nonround, based on F1 and F2. The dashed line shows a curved function that easily separates the two groups. The cut is an entirely plausible one. The dotted line represents several possible 2-step decisions, such as the following: If a vowel has a high F1 (in this case, higher than about 400 Hz), it is unrounded. With F1 lower than 400 Hz, a vowel is round if F2 is less than 2000 Hz, nonround if it is greater. Naturally, the frequency values named are not absolute, but are presumably different for each speaker heard.

As with the German results, "round" is only an approximate label for a well-defined acoustic quality. Classifying [ɤ] as [+round], however, is not satisfying. In addition to being truly unrounded, it functions in the phonological system of Turkish as a nonrounded vowel, according to the vowel harmony relationships described in Chapter 4. Since it was perceived as a good Turkish [ɤ] by only two out of the seven listeners (the same two that show the largest weights on Back/Nonback2 in the PARAFAC solution, c.f. Appendix II), the placement of the [ɤ] stimulus on Round/Nonround may not be a good indication of the perceptual attributes of native [ɤ]. It is probably true, though, that the [ɤ] stimulus falls somewhere within the range of an acceptable native [ɤ], if not in the center of that range.

The remaining dimension, High/Low, is based mainly on F1 (Figure 6.9), and can be predicted somewhat better by considering the amplitude of A1 as well ($r = .93$). But on the basis of Figure 6.10e, the F1-A1 combination does not really provide a good prediction of the grouping of vowels into [æ ʌ ɑ ɔ ɤ], [ø e ɪ ɔ] and [i u y]. If [ɔ] and [ɤ] are not considered, being non-native, then F1 alone predicts the separation of low from non-low quite well. The perceptual distance between high and mid vowels is not reflected in F1,

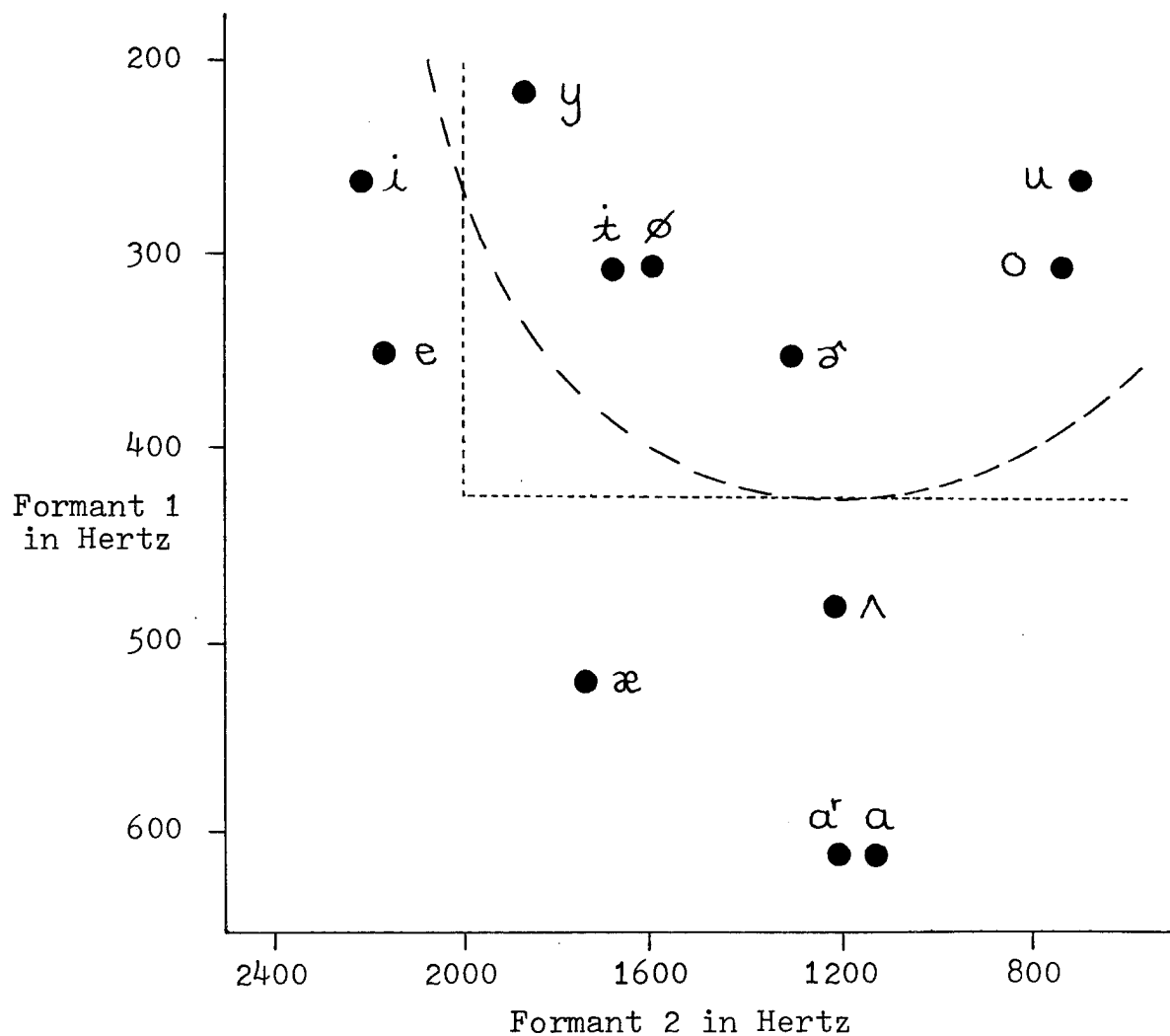
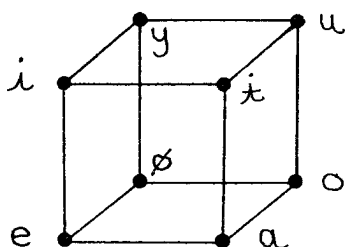


Figure 6.11: Two strategies for basing a rounding decision in Turkish on the F1-F2 plane.

though. This result was also found with German, where the effect was so great an extra dimension appeared.

It is curious that [ɤ] appears so close to [ɑ], when [ɑ] was explicitly stated by all the listeners to be an example of [α] with an unusual voice quality, and [ɤ] was heard as a version of [ø]. It would be gratifying, in some future experiment, to discover that retroflexion emerges as a separate dimension, given enough listeners. [ɑ] and [ɤ] would naturally appear at one end of this scale. Unfortunately, this is pure conjecture, and no good accounting for the position of [ɤ] on the High/Low scale suggests itself.

It appears that the 6-D PARAFAC solution of Chapter 5 corresponds well with the present one: the four dimensions of that solution most heavily weighted by the Turkish listeners emerge here as the only four interpretable dimensions. And yet, the solution is somewhat of a disappointment, if the expected result is a cube with each corner occupied:



[ɪ] was not perceived as high at all, though it was perceived by at least two listeners as back. [ɑ] is perceptually lower than [e], [ø], or [o], although it functions phonologically with those three vowels as low. It appears that the idealized linguistic structure for Turkish mentioned above appears at a more abstract level of cognition than the one accessed in this experiment, if it is psychologically real at all.

6.3 Thai

For the six Thai listeners, it is very difficult to decide which solution to treat as the correct one. The uniqueness-of-solution property is not as sure a criterion as it was in the case of German and Turkish. The interpretation of the scales that do emerge is difficult as well. One cause of this plight is certainly the low overall reliability of the Thai subjects, which can be confirmed by referring to the statistical tests in Tables 4.2 and 4.3.

There are two solutions at two dimensions. The most conservative course is therefore to regard the Thai responses as being one-dimensional. Interpretation is easy; the vowels form three clusters of High, Mid and Low. It is also of interest to note that the single most salient dimension for

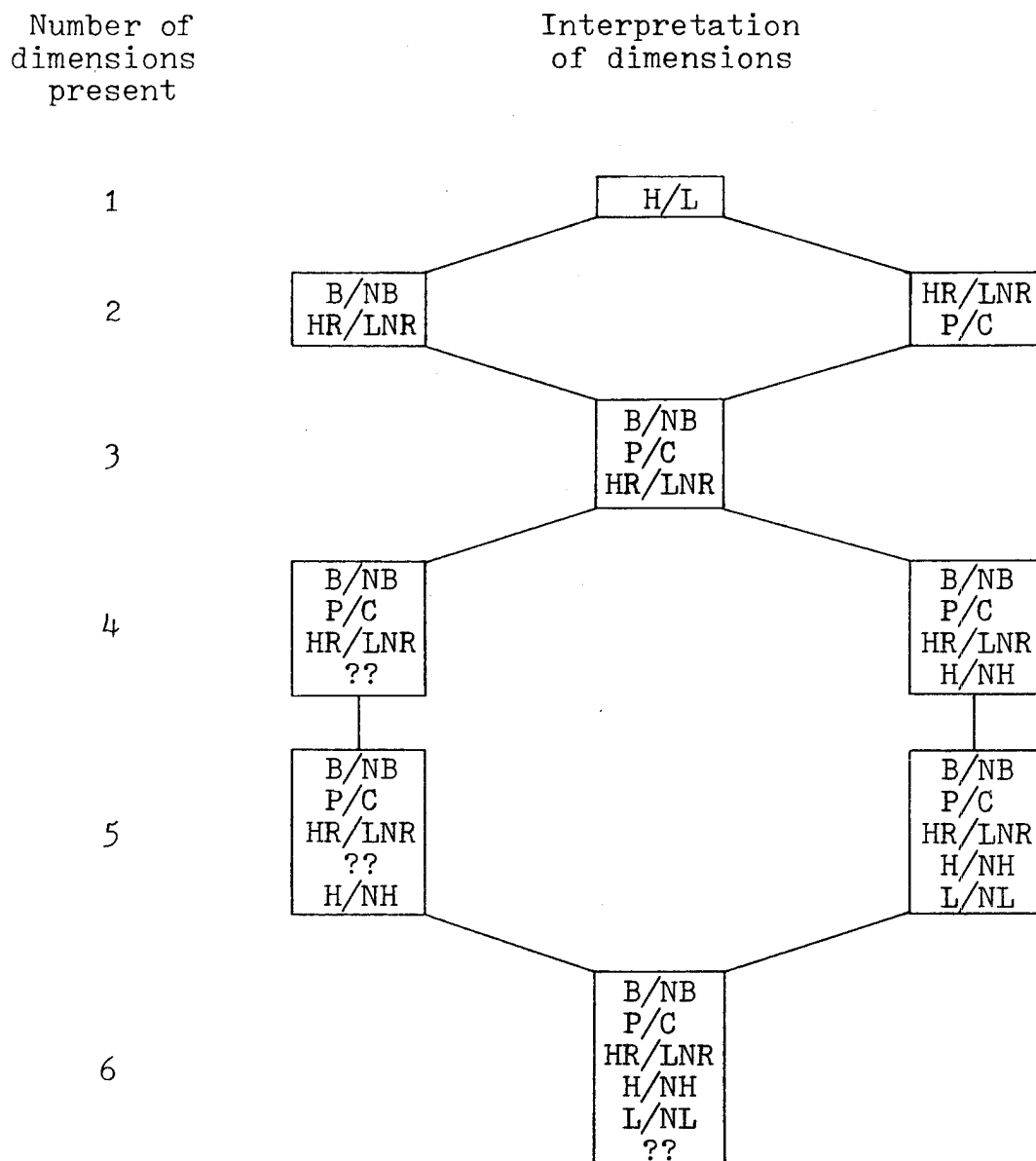
Thai listeners is not backness, as it was with German and Turkish. Certainly this point deserves some emphasis. And yet, one strongly suspects that there is more than one dimension to the Thai space, given the results for the other languages.

Figure 6.12 summarizes the properties of solutions in one through six dimensions. It appears that the two 2-D solutions each capture a different second dimension; in three dimensions, both of those scales appear, in addition to a dimension showing properties of both rounding and height. It is very likely that there are indeed three dimensions present, two of which are equally salient to the listeners as a group. In four dimensions, PARAFAC derives two solutions, one of which has an uninterpretable dimension ("??"). The vowels are separated into two main groups, $[\alpha \text{ } \text{a} \text{ } \text{æ} \text{ } \text{ɛ} \text{ } \text{ɪ} \text{ } \text{y}]$, and $[\text{e} \text{ } \text{o} \text{ } \text{ø} \text{ } \text{ɔ} \text{ } \text{u} \text{ } \text{ɯ}]$. The implications of the 5- and 6-D spaces will be discussed briefly after the analysis of the 3-D solution.

The two plots found in Figure 6.13 show the vowel space of the 3-D solution. Considering native vowels only, the vertical dimension on Figure 6.13a cuts the plane into the clearest groups, $[\text{ɪ} \text{ } \text{u} \text{ } \text{ɔ}]$ and $[\text{e} \text{ } \text{æ} \text{ } \text{ʌ} \text{ } \text{ɑ}]$. Members of the pairs $[\text{æ} \text{ } \text{e}]$, $[\text{u} \text{ } \text{o}]$, and $[\text{ʌ} \text{ } \text{ɑ}]$ are also distinguished in this direction, suggesting a further division of the two main levels into four. The most accurate label is the descriptive but awkward Highround/Lownonround, as this dimension appears to collapse two properties, rounding and height, into one scale. The horizontal dimension on both graphs is most related to the traditional term Back/Nonback, although $[\text{u}]$ is displaced towards the nonback side. Perhaps Palatal/Pharyngeal captures the vowel pattern better; this way, the extreme position of $[\text{ɪ}]$ has some meaning, as does the separating of velar $[\text{u}]$ from pharyngeal $[\text{o}]$.

And yet, when Figure 6.13a is considered as a whole, the vowel pattern lacks but a slight clockwise rotation to obtain more easily interpretable dimensions, namely height and backness. However, rotation is not permissible, and some other way must be found to deal with this pattern.

The Peripheral/Central dimension is found on Figure 6.13b. Vowels that line the perimeter of the F1-F2 plane, that is, $[\text{ɪ} \text{ } \text{e} \text{ } \text{æ} \text{ } \text{ɑ} \text{ } \text{o} \text{ } \text{u}]$ appear near the bottom of this dimension, while vowels central to the plane, $[\text{ʌ} \text{ } \text{ɤ} \text{ } \text{ɯ} \text{ } \text{o} \text{ } \text{ɤ}]$, are at the top. Its main function for the native vowels is to separate $[\text{ɑ}]$ and $[\text{ʌ}]$, although it redundantly helps distinguish $[\text{ɪ}]$ and $[\text{u}]$ from other peripheral vowels. A native Thai $[\text{ɤ}]$ would certainly be at the central end, as suggested by the $[\text{ɤ}]$ stimulus.



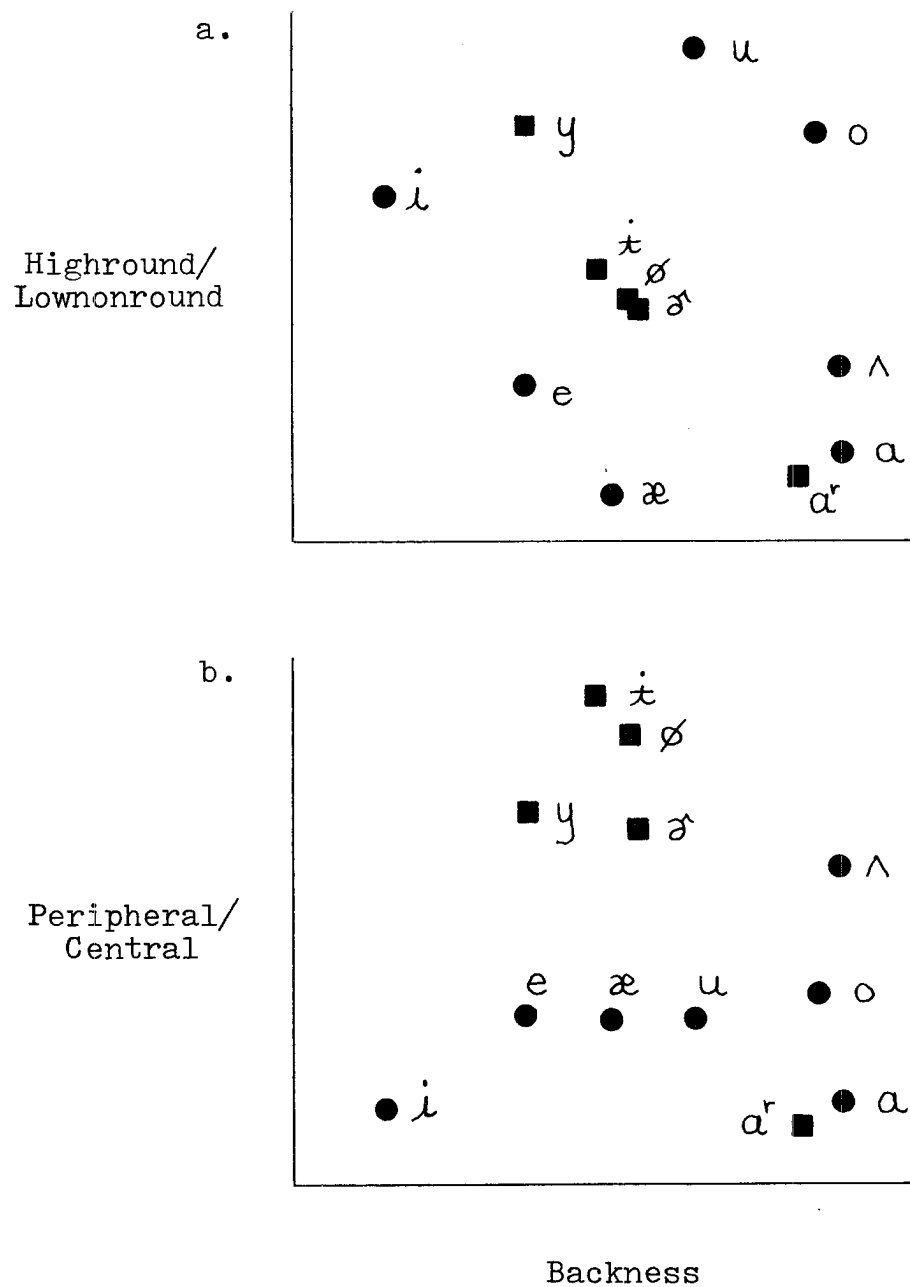


Figure 6.13: PARAFAC solution for Thai in 3 dimensions.
● = native vowels; ■ = non-native.

This scale is similar to the Peripheral/Central scale of the 35-subject space, as can be seen in Figure 6.14. The first noteworthy difference is that Thai listeners taken alone have a definite gap separating peripheral and central vowels. The presence of the gap, together with the fact that this scale appears as one of only two or three dimensions, supports the claim that Peripheral/Central is a real dimension for Thai subjects. A second difference is that [ɤ] does not appear at the extreme end of the Thai scale. This supports the conjecture made in Chapter 5 that the Peripheral/Central pattern that emerged there is a composite of a "real" Peripheral/Central dimension and one of retroflexion. This will be confirmed in the discussion of the results for English (Section 6.4).

Figure 6.15 summarizes the regression analyses. Highround/Lownonround and Back/Nonback are reasonably well predicted by acoustic variables, even though they pattern imperfectly in relation to linguistic features. Back/Nonback correlates well ($r = .94$) with a scale computed using $F1-.44F2$. The details of this prediction are found in Figure 6.16a. [u], grouped more closely with nonback vowels on the perceptual scale, is an anomaly that cannot be explained easily. The large perceptual separation of [u] and [o] has no obvious acoustic source. Highround/Lownonround (Figure 6.16b) is predicted extremely closely by $F1+.12F2$ ($r = .96$, c.f. Figure 6.15). Notice that all the high vowels, [i u y], are shifted further to the left than is predicted by $F1$ and $F2$. The basic property of this dimension, then, is a weighted sum of $F1$ and $F2$, supplemented by the provision that vowels with extremely low $F1$ are perceived as lower than predicted by $F1+.12F2$ alone.

Since Back/Nonback and Highround/Lownonround are predicted mainly by $F2$ and $F1$, respectively, with additional information supplied by $F1$ and $F2$, respectively, one would expect a plot of the two perceptual scales to resemble a rotated plot of $F1$ against $F2$. Figure 6.17 shows to what extent this is the case. The solid lines represent formant frequencies. The formant space must be distorted as well as rotated to match the position of the vowels on the plane. If $F1$ and $F2$ are actually the "true" perceptual dimensions, then the way distances are estimated by the listeners does not fit a Euclidean model, causing PARAFAC's inability to capture those dimensions.

It was impossible to account for all vowels on this graph. [ɔ], [ɤ] and [ʌ] could not be accommodated into any reasonable curvature of $F1$ and $F2$; the "x" entries for those vowels show the predicted positions according to $F1$ and $F2$. It is tempting to account for [ɤ] and [ɔ] by an appeal to $F3$.

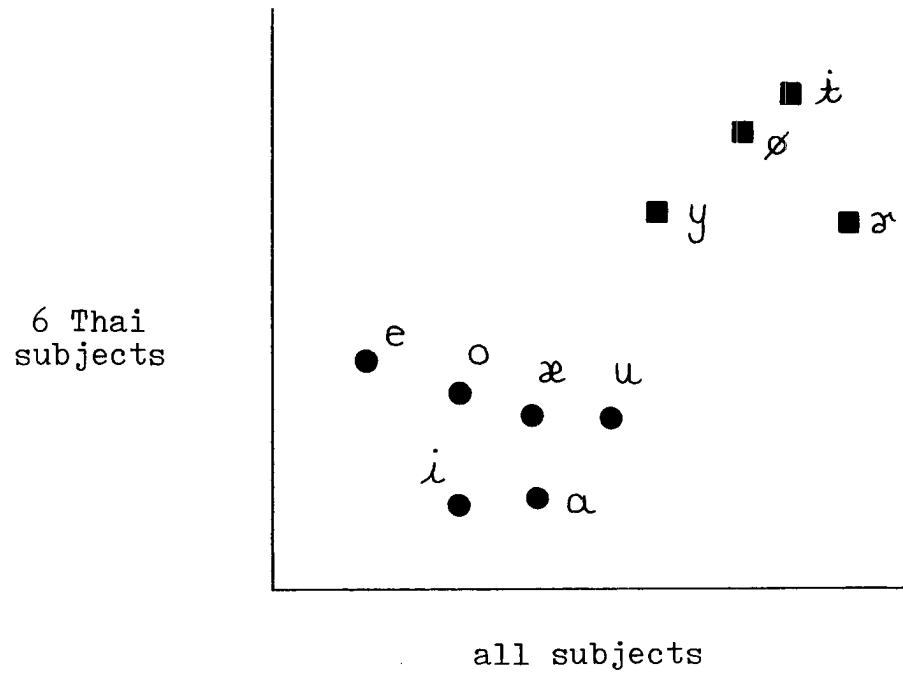


Figure 6.14: The Peripheral/Central dimension from 2 different solutions: the 3D space for Thai alone, and the 6D space for all 35 listeners. ● = native vowels; ■ = non-native.

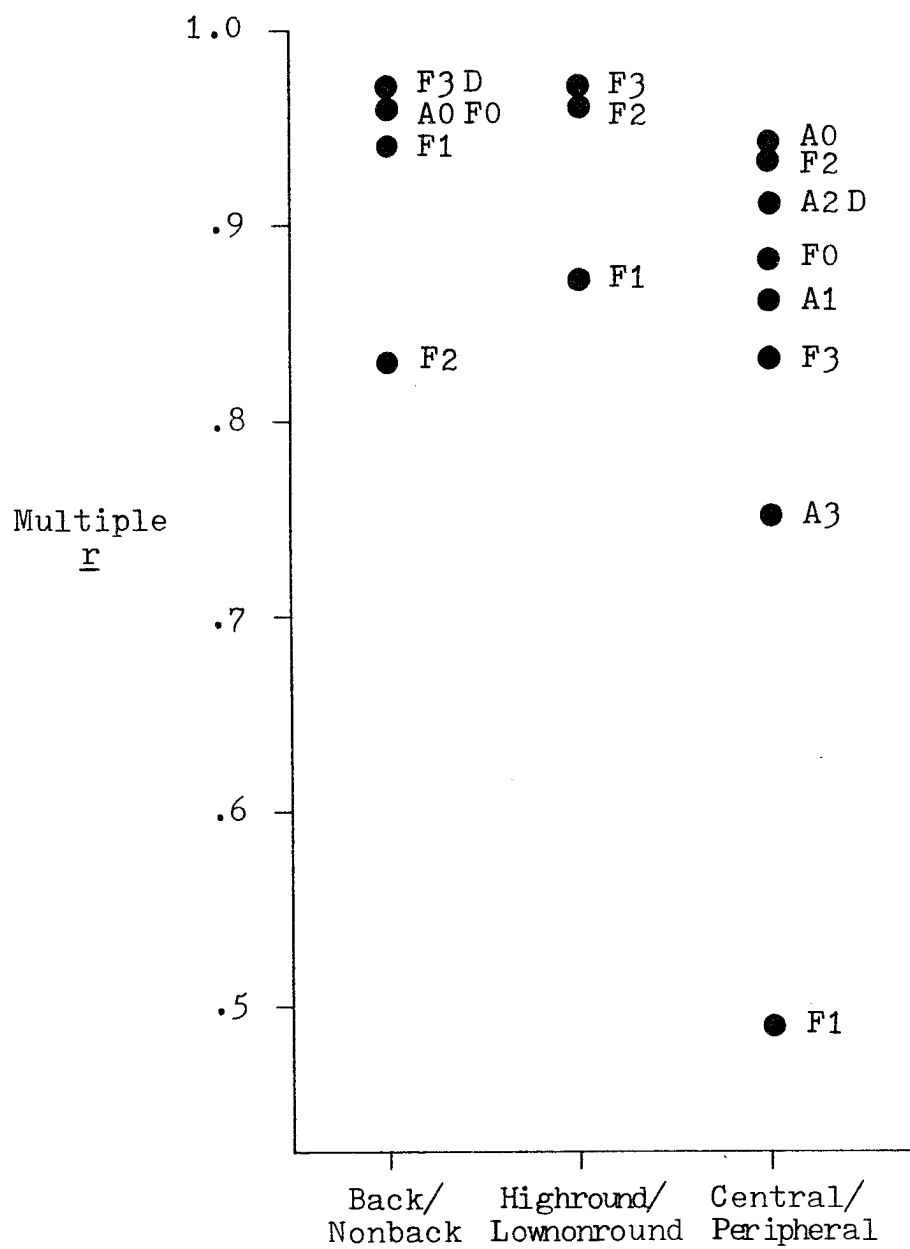


Figure 6.15: Summary of regression analyses, 3D PARAFAC solution for 6 Thai listeners.

The low F3 for [ɔ̌] may cause F2 to be perceived as higher than its true value, hence the closeness of [ø] and [ɔ̌]. The slight perceptual separation of [ɛ̌] and [ø] can be similarly explained in that [ɛ̌] has a higher third formant than [ø], in spite of their similarity in formants 1 and 2. Unfortunately, there is little satisfaction in proposing such an explanation for two non-native vowels, when the position of native [ʌ] is left unaccounted for.

Considering the relative noisiness of the Thai data, it is pointless to attempt too detailed an interpretation of the curvature. One aspect worth pointing out, however, is that the co-ordinates converge to shrink the perceptual distance in the F2 direction as F1 decreases. The upper central vowel region is thereby compressed. Is this because Thai only has one vowel in that region ([ɛ̌])? Or is this an error-laden reflection of an underestimation of large distances? Clearly, additional evidence is needed.

The Peripheral/Central dimension remains to be discussed. The quality of the predictions for this scale, based on linear regressions, is rather bad (Figure 6.15). Six variables are required to reach an r of .9, and all nine variables together show r of only .94. From an inspection of the pattern of predictions (Figure 6.16c), it appears that no regression can reproduce the clustering of vowels displayed by Peripheral/Central; although, to be charitable, one could point out that the 3-variable regression at least allows a cut to be made between [e] and [ʌ], dividing the vowels into the main groups seen in the perceptual dimension.

A more plausible account of the Peripheral/Central scale is a decision function illustrated in Figure 6.18. Rather than claim that the listeners compute Peripheral/Central values from F1, A3, F3 and perhaps other parameters as well -- and do it badly -- the following strategy, based solely on F1 and F2, is proposed as a better explanation:

- 1) Is the vowel back and rounded? (That is, is F2 sufficiently low?) If so, the vowel is Peripheral.
- 2) Does the vowel lie on the convex side of the acoustic boundary represented by the curved dotted line of Figure 6.18? If so, the vowel is Peripheral.
- 3) If the vowel is within the region defined by the dotted lines, it is Central, and its position along the scale is based on its distance from [ɛ̌].

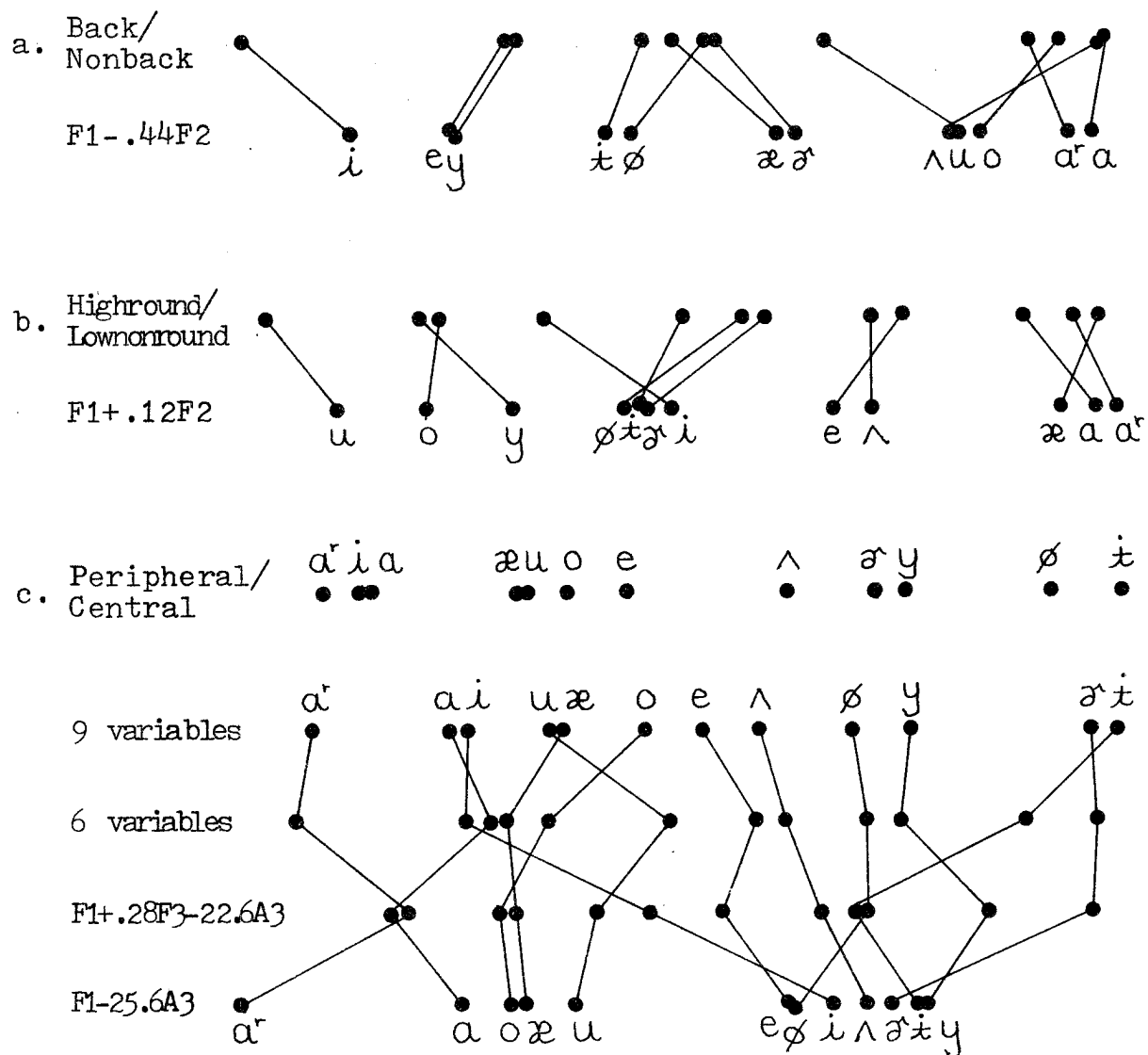


Figure 6.16: Selected comparisons of Thai PARAFAC dimensions with their predictions by acoustic variables.

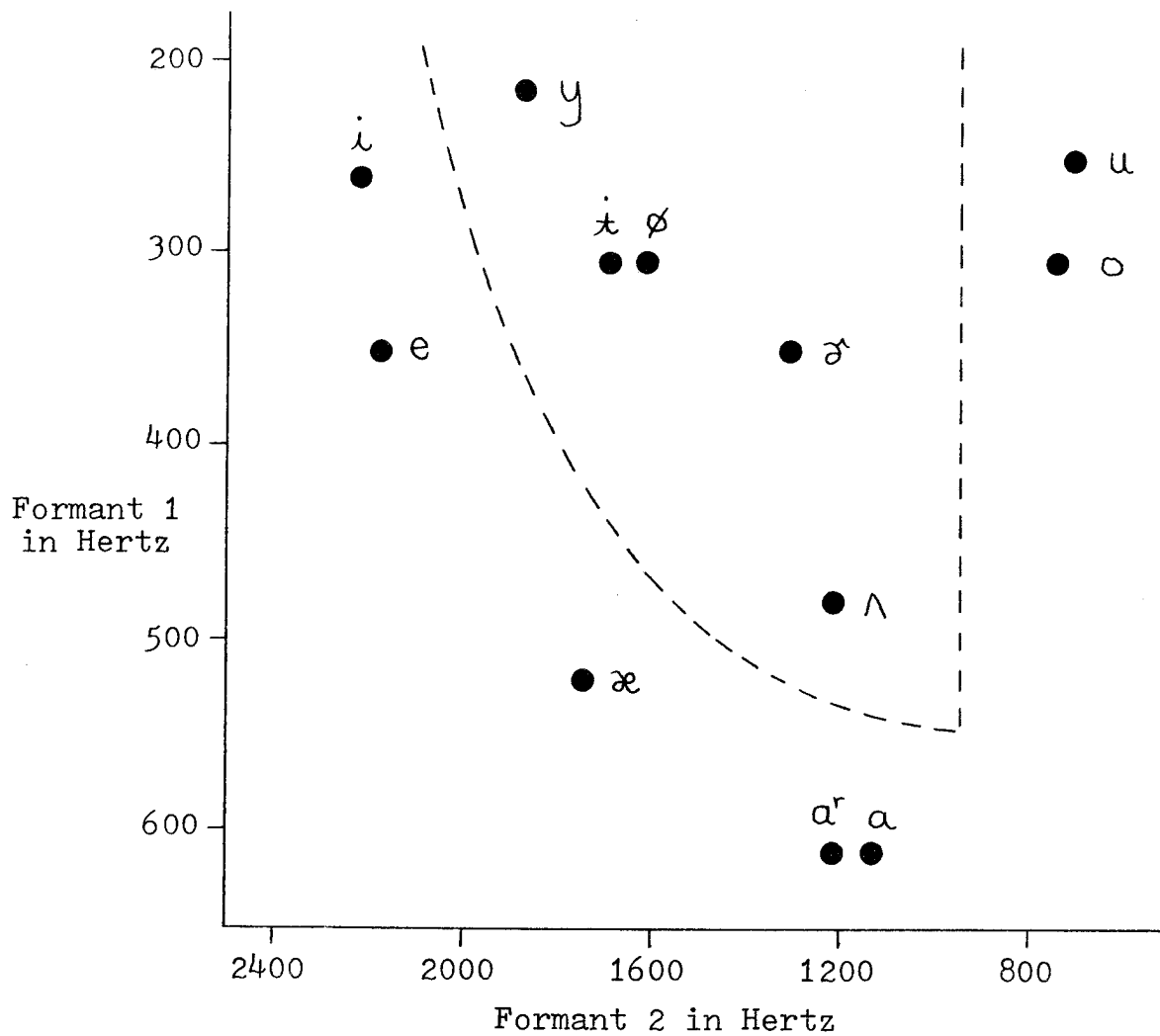


Figure 6.18: A strategy for Thai listeners' assessment of Peripheral/Central.

A few comments are in order. Step 1 predicts that the native Thai [ɔ], not present among the stimuli, would be classed as peripheral, with its low F2 (c.f. Figure 3.2b). This requires confirmation. With step 2, the curved dotted line may be awkward to refer to, but it is no less plausible a perceptual boundary than any straight line. Finally, some skepticism should temper the acceptance of step 3. It is certainly true that the distance from [ɛ] to the other central vowels corresponds strikingly to the distribution of vowels at the Central end of the scale. But there are only two native vowels in this region, [ɛ] and [ʌ], and it is doubtful that such an elaborate computation should be a part of one's natural strategies for dealing with so few sounds.

The difficult nature of the Thai results in three dimensions prompted the exploration of larger PARAFAC spaces. Often, when too few dimensions are extracted, the "real" component scales are distorted and combined with one another. It was hoped that in this case, the 5- and 6-D spaces would contain more intuitively satisfactory scales of height, backness, and perhaps rounding; a "better" version of Peripheral/Central might also have simpler acoustic correlates. Unfortunately, none of these hopes were realized. As summarized in Figure 6.12, all spaces from 4-D to 6-D inclusive contained essentially the same distorted and rotated F1-F2 plane found in three dimensions (that is, Back/Nonback and Highround/Lownonround) and the same Peripheral/Central scale. A successful interpretation of the 3-D space is therefore prerequisite to the eventual interpretation of spaces with more dimensions. (Both the 3-D and 6-D solutions appear in Appendix II, to allow further comparison.)

Although it is pointless to study the 5-D and 6-D spaces in detail, due to the apparent noise in the data, the emphasis on the height parameter in these and smaller spaces is too great to be overlooked. In six dimensions, four scales directly or indirectly refer to vowel height: Low/Nonlow, High/Nonhigh, Highround/Lownonround, and Back/Nonback. Only one dimension each refers to rounding, backness, or centralness. This prominence of vowel height is corroborated in the Thai vowel inventory: Thai has nine vowel monophthongs -- three front unrounded, three nonfront unrounded, and three back rounded (c.f. Figure 3.1b). Distinctions within each triple are made according to height alone. In the other four test languages, height distinctions are not independent of diphthongization, backness, or rounding, nor are they as systematic.

6.4 English

The data from the eight English listeners was difficult to analyze. According to PARAFAC, the solution in five dimensions was the prime candidate for discussion, uniqueness breaking down at six dimensions. The solution proved difficult to interpret. It was decided to reanalyze the data, deleting the subject whose weighting factors were most atypical of the eight.

The listener with the most anomalous weighting factors (E7) was in fact a professional singer who was proficient in the pronunciation of French and German. Her data were eliminated, in hopes of obtaining a more homogenous group of subjects. All solutions were unique until six dimensions, and some severe problems in interpreting the data still remained. Once more, one listener was eliminated whose weights were most unlike those of the other subjects. This subject knew no foreign languages and had no phonetic training. This time, there was again a unique solution in five dimensions, but the interpretability was much greater.

The results presented below are from the six-person solution in five dimensions. No apologies are made for totally ignoring the two excluded subjects. If their data were merely noisy (which is doubtful -- their reliability and validity scores were no worse than others', c.f. Chapter 4), nothing has been lost in the interpretation. If their responses are actually based on different perceptual properties, the present 5-D solution nevertheless represents the perceptual space of six native English listeners. It will be of great interest in future studies to determine how representative this sample is.

Figure 6.19 summarizes the evolution of the 5-D solution. Round/Nonround appeared in every solution. The next most salient dimension was High/low, followed by Back/Nonback. A second 3-D solution added Retroflex to Rounding and Height, rather than Backness. The 4-D space was unique and consisted of all dimensions found in the two 3-D solutions. With five dimensions, a scale labelled Mid/Nonmid was added. Two 6-D solutions emerged, which contained some uninterpretable dimensions.

The interpretation of the 5-D space is simplified by regarding High/Low and Mid/Nonmid as components of a single curved height dimension, as in Figure 6.20. The curve drawn on the figure, fitted by sight, does not allow a simple division into two or three clusters of high, (mid), and low vowels. Backness and rounding affect the positions along the curve considerably. For example, front vowels [æ e i] are placed

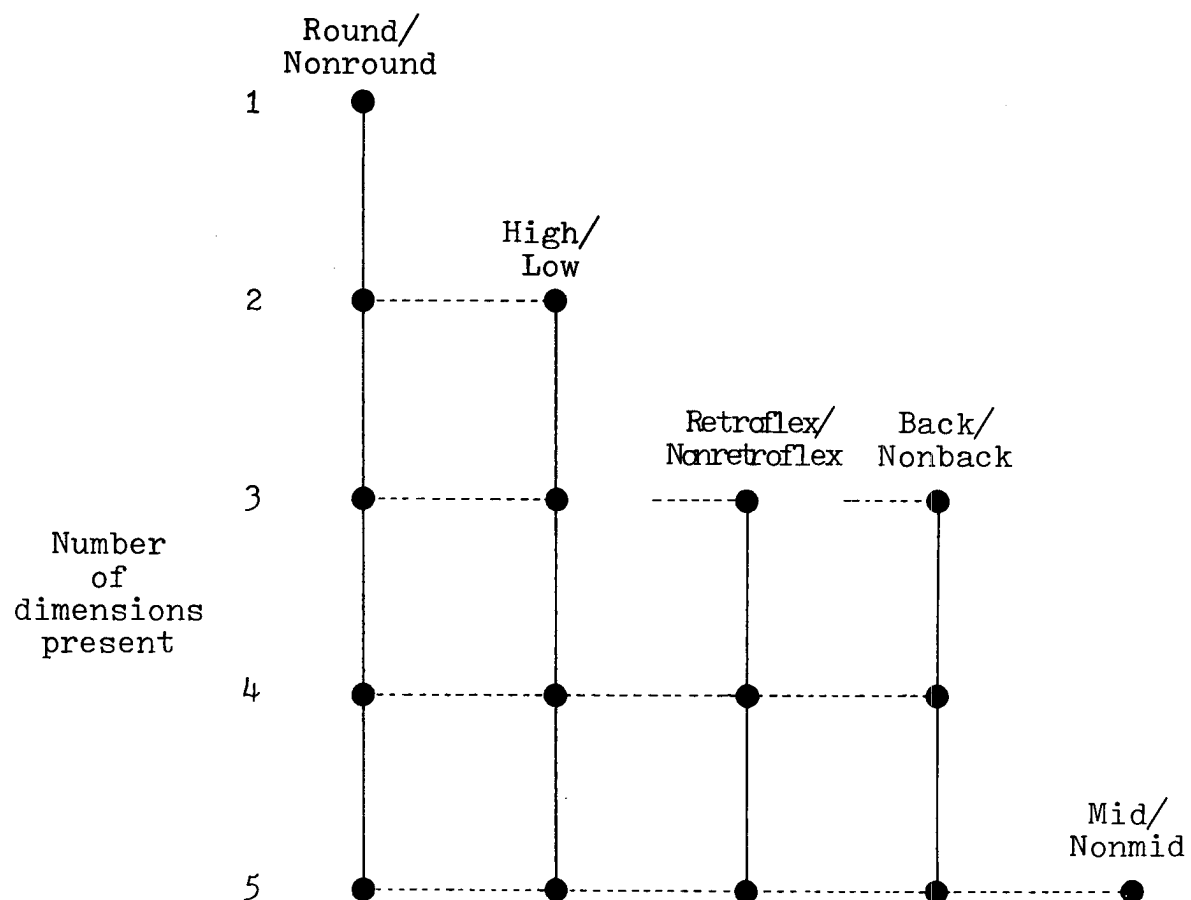


Figure 6.19: Summary of the PARAFAC solutions for English in 1-5 dimensions.

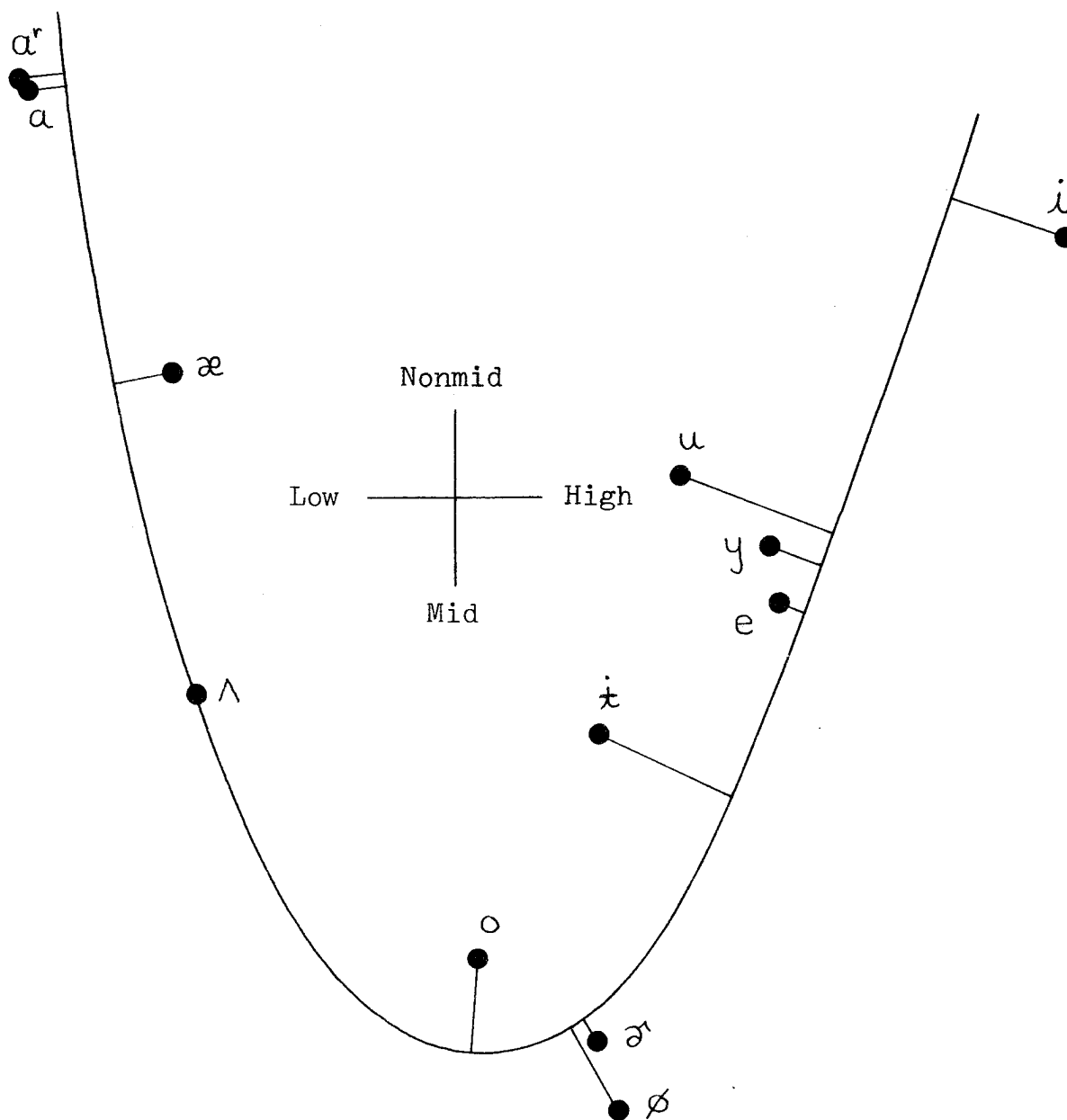


Figure 6.20: High/Low and Mid/Nonmid from the 5D solution for English, with curved line fitted by sight.

farther towards the right along the curve than their back counterparts [ɑ ɔ u]. Also, the difference in rounding seems to place [ɪ] and [ʊ] even farther apart than the curve itself can allow. In any case, the curved height scale is no less interpretable than the original High/Low and Mid/Nonmid dimensions.

The straightened height curve is displayed with the other three dimensions in Figure 6.21. Part a, height vs. rounding, reflects the two-dimensional plane most salient for the English listeners. (Planes formed from other combinations of dimensions are not as interpretable as the planes in the figure.) Vowels on this plane form three major clusters: [ɪ e], [ɛ ə ø u ɔ], and [æ ʌ ɑ ɔ̃]. The appropriate binary distinctive feature labelling for these groups would be, respectively, [-Low - Round], [-Low +Round], and [+Low -Round]. (The re-appearance of [ɛ] as [+round] should be no surprise.) Actually, the slight rotation of the plane, plus the fact that the clusters are about equidistant from one another, suggests that no set of axes -- no matter what the rotation -- will represent perceptually real dimensions. Rather, the three vowel groups may be perceived as qualitatively different from each other, requiring independent labels, such as Palatal, Round, and Low. Under these circumstances, PARAFAC would still be forced to derive some set of axes, but they would be largely arbitrary, devoid of any psychological import.

The data, however, do not completely vitiate the need for some explicit reference to vowel height. In attempting to provide all the native vowels with unique feature values, some reference to height is necessary within the main clusters, no matter how the clusters themselves are labelled. The pairs [ɪ e], [ʌ ɑ], and [u ɔ] require either a new feature such as $\pm$ Close, setting off [ɪ ʌ u] from [e ɔ ɑ], or a four-valued height feature.

The remaining two dimensions are shown in Figure 6.21b. Backness, the horizontal dimension, separates back from non-back vowels by means of a gap between [ɪ] and [ɑ]. But [ɛ] is grouped squarely with the back group, and [ɪ] is badly out of place, appearing as less front than [ɛ] or [e]. More curious yet is the fact that Back/Nonback separates several vowels which listeners informally called "the same": [y], while not a distinct English phoneme, was regarded as a native sound by the six listeners used in the present analysis. When asked informally to repeat the vowel, [y] was pronounced the same way as [u] (that is, [ʉ] or [ɯ]). Several listeners referred to it as "a kind of 'u'." Similarly, [ø] was heard by many listeners as [ɛ]. In these cases, the issue is not "foreign" vs. "native" sounds, but a wide range of acoustic shapes that are recognized as the same unit. Back/Nonback

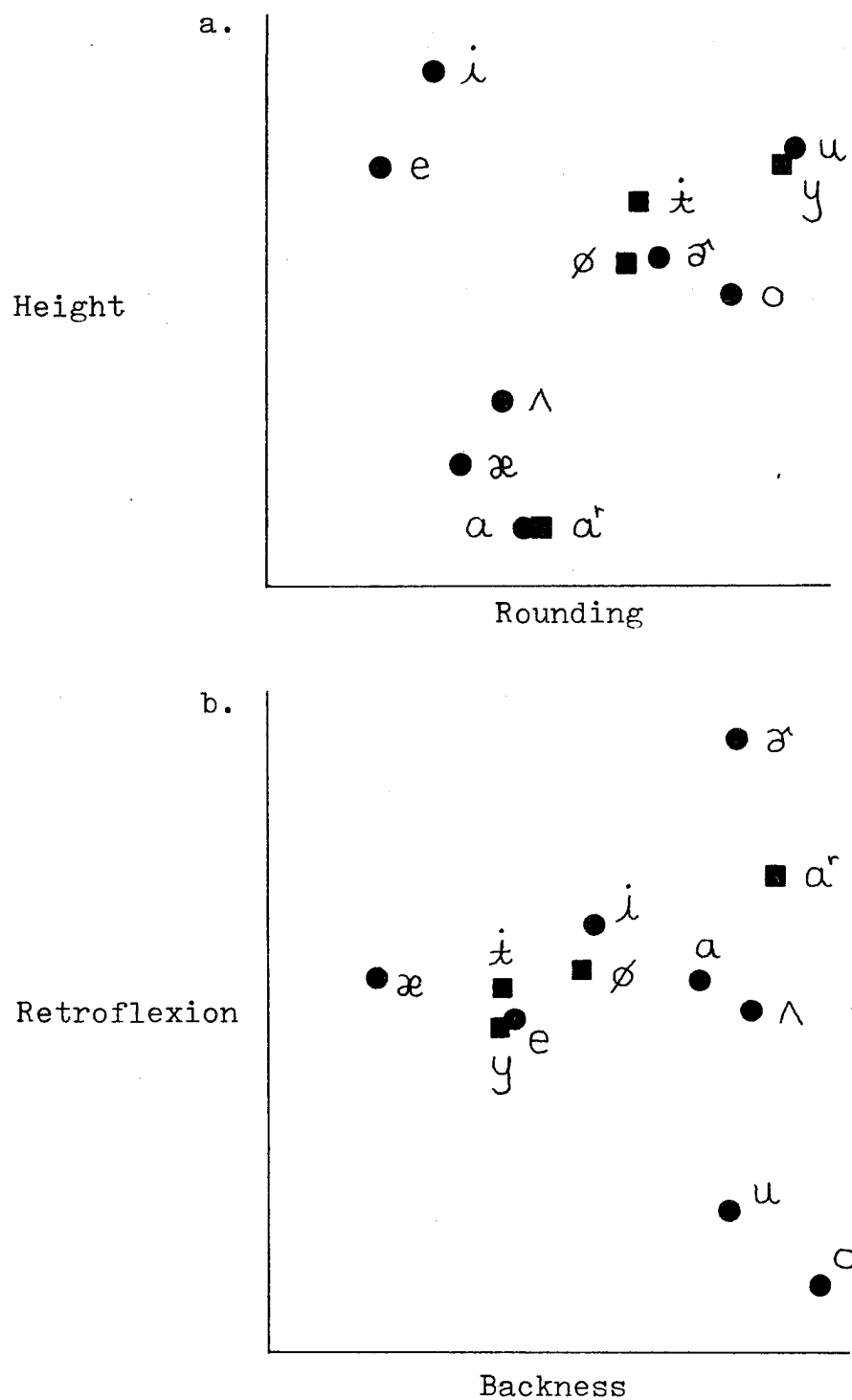


Figure 6.21: Back/Nonback, Round/Nonround, Retroflex/Nonretroflex, and straightened height curve from the 5D PARAFAC solution for English. ● = native vowels; ■ = non-native.

is therefore of no use in identifying these vowels.

The only apparent function of the Back/Nonback is to distinguish front [æ] from back [ʌ] and [ɑ]. This amounts to a hypothesis that a vowel is first evaluated according to the rounding - height plane. If it falls into the low group, it is then assessed for backness.

The vertical dimension of Figure 6.21b is called retroflex, mainly for its placement of [ɤ] at one extreme. Actually, the vowel that is crucial to the discovery of a Retroflex dimension is [ɑ̣]. It is gratifying to find [ɑ̣] as the next vowel in from [ɤ]. The right end of the scale does plausibly reflect a perceived degree of retroflexion; for no other language was such a pattern found. Acoustically, [ɑ̣] does not have a third formant as low as that of [ɤ], but it does have the lowest F3 of all the back vowels. One might argue that [ɑ̣] was merely perceived as "between" [ɤ] and [ɑ], incorporating attributes of both. But it is precisely on the dimension of retroflexion that [ɑ̣] appears midway between [ɤ] and [ɑ].

What of the [ɑ̣] end of the dimension? It is unlikely that this scale is used to identify back rounded vowels, although it certainly seems capable of doing so. As was noted in the discussion of the backness dimension, [y] and [u] are felt to be the "same" vowel. This scale contributes nothing to grouping [y] and [u] together, what with [y] falling in the large nondescript middle group. If [y] and [u] are the "same" vowel to English listeners, the listeners obviously disregard the separation of the two sounds on the dimension called retroflex.

The main function of this scale is to separate [ɤ] from [u] ([y]) and [ɑ̣], all of which make up the main cluster of rounded vowels (Figure 6.21a). This suggests a subordinate role for the feature retroflex, although it may just as well be an independent parameter which sets off [ɤ] from all other vowels.

The regression analyses, used to relate the perceptual patterns to acoustic properties, are summarized in Figure 6.22, which gives the stepwise regression coefficients for each dimension, and in Figure 6.23, which shows some of the predictions of vowel placement made by the acoustic variables.

Taking the five dimensions in order of emergence, Round/Nonround comes first. It appears to be difficult to account for purely on the basis of linear regressions (Figures 6.22, 6.23a). Notice, however, that F1 and F2 alone account for the pattern of Round/Nonround reasonably

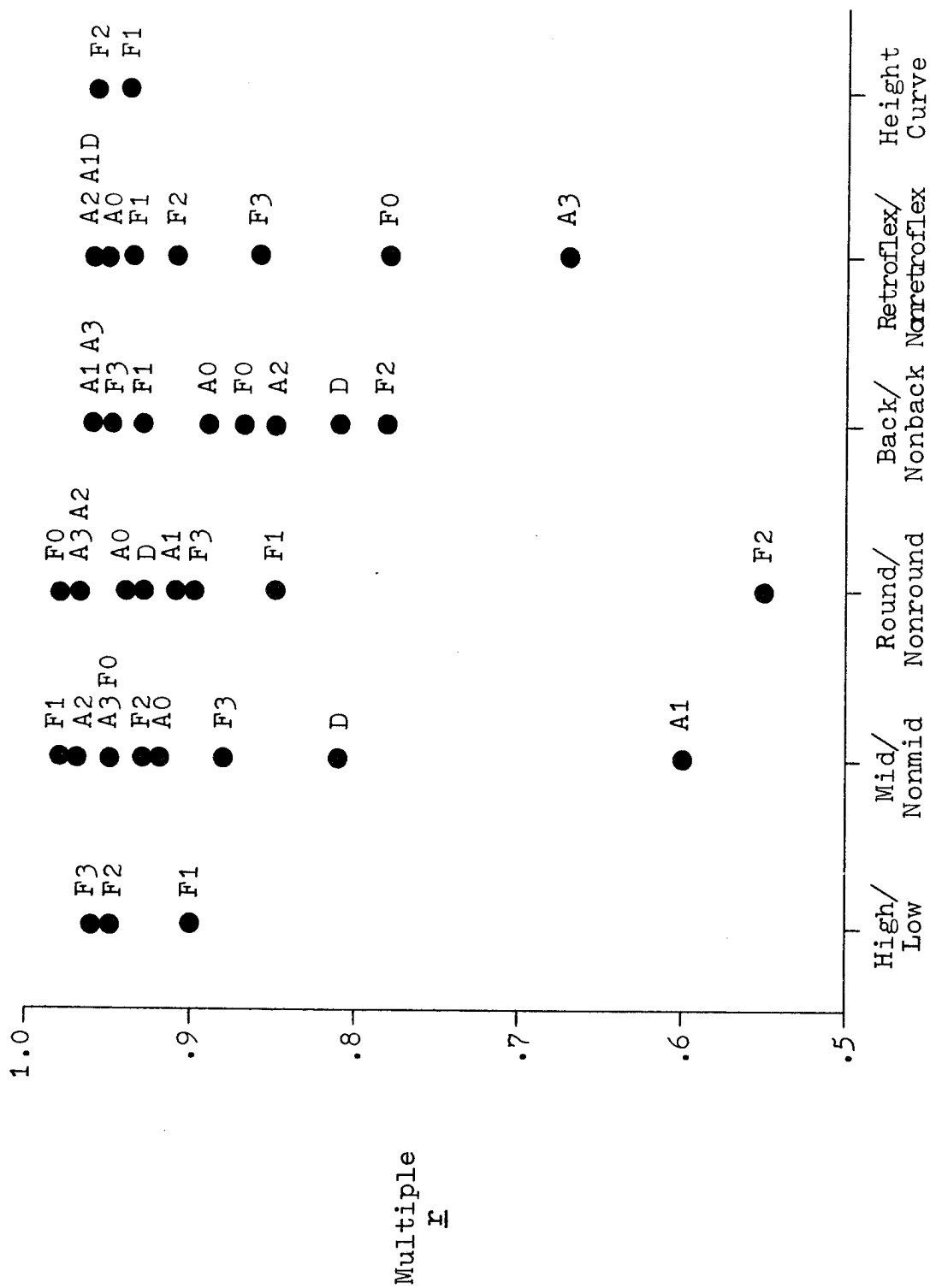


Figure 6.22: Summary of regression analyses, 5D PARAFAC solution for 6 English listeners.

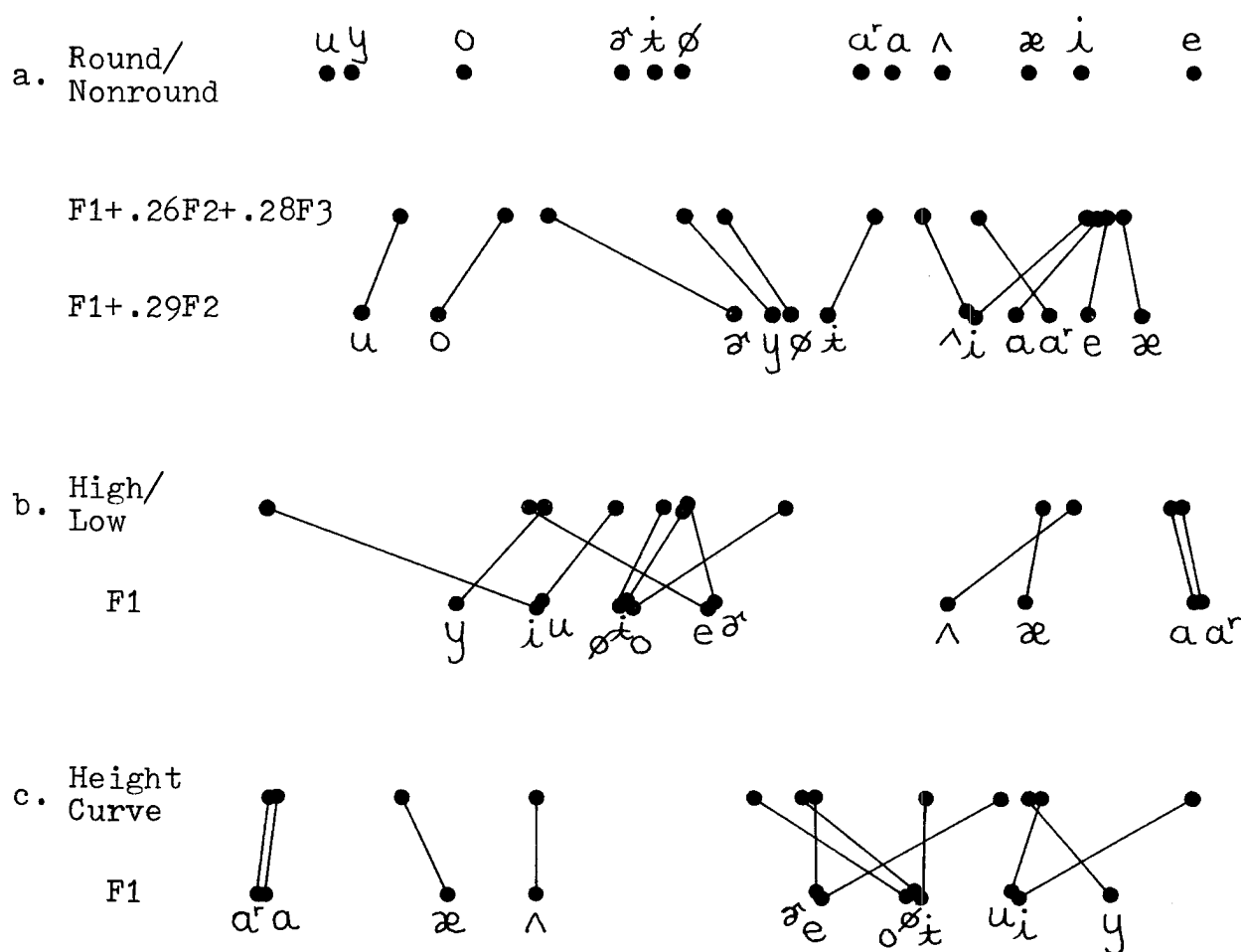


Figure 6.23: Selected comparisons of English PARAFAC dimensions (and height curve) with their predictions by acoustic variables.

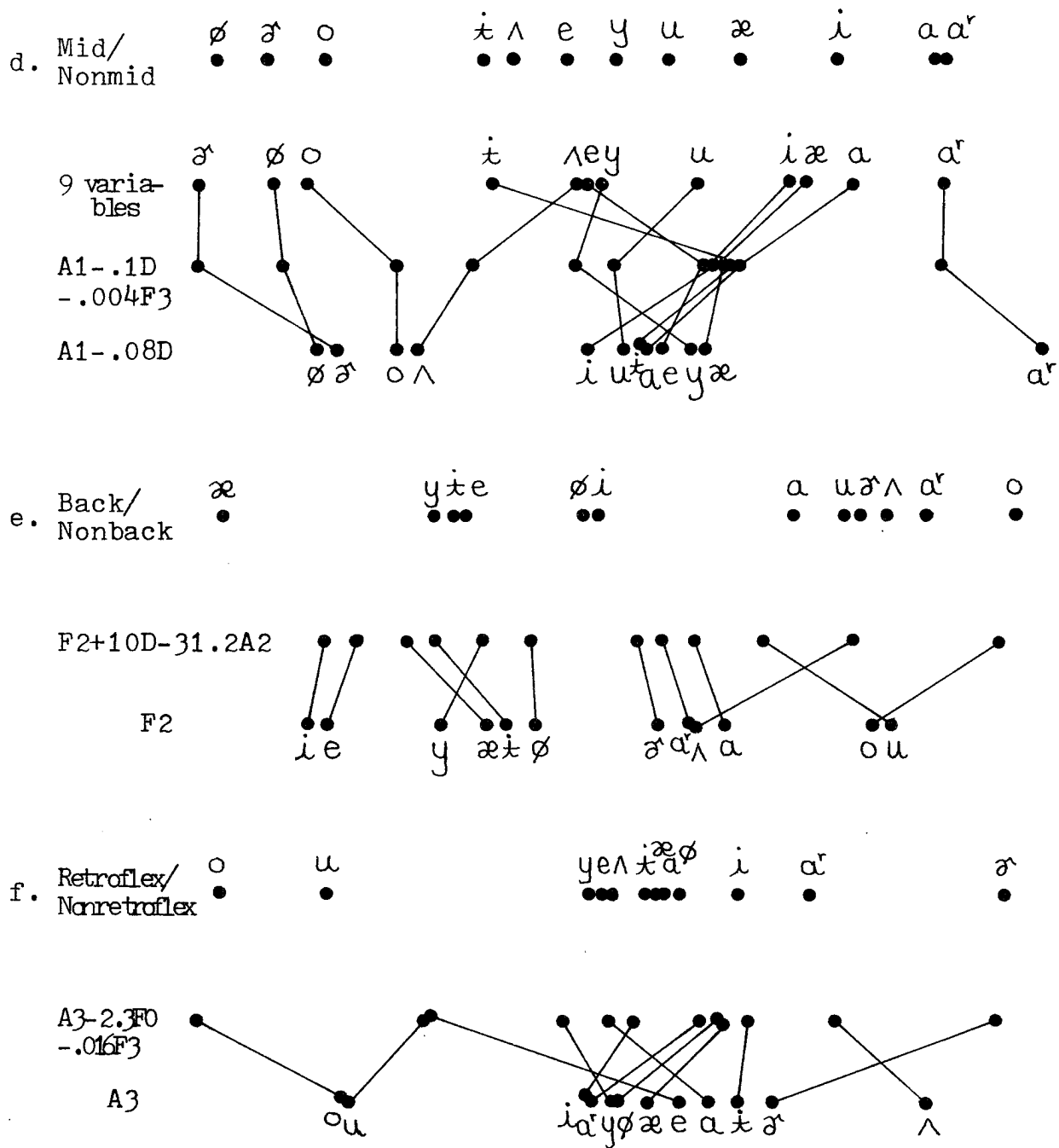


Figure 6.23: continued.

well, except for the position of [y]. The three-way separation of [u o] [ø ɨ ʌ] [ʌ ɨ ʌ ɔ̃ e æ], predicted from $F1 + .29F2$, is a good representation of the perceptual scale. The [y] can be satisfactorily explained with this proposal: A vowel falling in the middle group is re-examined according to F1. If F1 is very low, the vowel is shifted to the [u o] group. Since [ɔ̃] is the only native vowel remaining in the middle group, shifting the [y] aids in separating the distinctively rounded vowels [u(y) o] from redundantly rounded [ɔ̃].

The second dimension to emerge, High/Low, correlates with F1 values satisfactorily ($r = .90$, c.f. Figure 6.22), in that the perceptual separation into low vowels and high vowels is matched by the distribution on F1 (Figure 6.23b). When the plane formed by High/Low and Mid/Nonmid is regarded as a curved dimension of height, some improvement is made in both the correlation coefficient ($r = .94$ for F1 alone, c.f. Figure 6.22) and the vowel pattern (Figure 6.23c). Certainly, considering the Mid/Nonmid dimension by itself is very unsatisfying. Only eight or nine variables begin to account for the perceptual pattern, and the first few variables that account for the most variance make up an improbable, inexplicable set of predictors (c.f. Figure 6.22, 6.23d).

Back/Nonback, less salient than the first two dimensions, groups [ɑ u ɔ̃ ʌ ɔ̃ ɔ̃] together as back vowels, while placing the remaining, front vowels in a curious pattern at the opposite end of the scale (Figure 6.23e). The single variable F2 captures the major division into back and non-back. If non-native [ɨ] and [ø] are ignored, the gap in the pattern of F2 is slightly widened. Yet adding variables does not improve on the vowel pattern, and the leftmost [æ] is not accounted for at all. English listeners seem to apply a non-linear stretching to F2 differences, depending on the accompanying F1. The acoustically similar [æ] and [ɑ] are perceived as farther apart than the acoustically more distant pairs [ɨ u] and [e o]. Although the details differ, this phenomenon is similar to the compression of the [ɨ] vowel region reported for Thai.

Retroflexion, which emerged first in one of the 3-D solutions, is not well-predicted by the acoustic measures; A3, F0, F3 and F2, successively, must be included in order to reach $r = .91$ (Figure 6.22). But a much simpler relationship between perceptual scale and acoustic measures can be found than any regression analyses can express. According to Figure 6.23f, the vowels are arranged into three groups: [u o] at one end of the scale, [ɔ̃] at the other, and all the remaining vowels in between. [u] and [o] have very low relative A3 values, hence their isolation. [ɔ̃], on the

other hand, has an unusually prominent third formant, albeit in frequency, not amplitude.

Perhaps retroflex is really a scale of "high-frequency prominence," where the idea of prominence includes both frequency and amplitude. The non-prominent [ʊ] and [ɔ] lie at one end, vowels having ordinary second and third formant prominence are towards the middle, and the [ɤ] vowel, whose lowered third formant makes the F2 range unusually prominent, appears at the other end of the scale. However, since there is already a rounding dimension whose function it is to separate rounded [ʊɔ] (and [ɣ]) from the rest of the vowels, the only apparent function of the dimension of "high-frequency prominence" is the separation of [ɤ].

Before going on to the Swedish results, a brief comment is needed about the choice of the 5-D solution as the "correct" one. The input/output ratio of 4.4 is a warning that too much is being asked of the data. It is worth pointing out that the basic interpretation of the solutions does not change significantly from one to another (Figure 6.19). If the 4-D solution were prudently thought to be the one to present, what would be lost is the Mid/Nonmid scale, which is not interpretable in its own right. If an input/output ratio of 5.5 is regarded as too unstable, the investigator must choose which 3-D solution to discard. Choosing both effectively means choosing the 4-D solution. Stopping at two dimensions, in order to obtain a solid input/output ratio of 11, results in an interpretation of frustrating incompleteness. Considering all this, the focus on the 5-D solution seems to be well-justified.

6.5 Swedish

It may come as an anticlimax to have it reported that the most reasonable interpretation of the Swedish space encompasses the three dimensions of height, backness, and rounding. But arriving at this straightforward solution was an arduous task. For the reader willing to (temporarily) forgo the description of problems in obtaining the solution, Section 6.51 might be skipped. Section 6.52 discusses the relationship between the perceptual and acoustic spaces.

6.51 Problems in obtaining the Swedish solution

PARAFAC solutions were unique up to and including six dimensions, but only the 1-, 2-, and 3-D spaces were completely interpretable. The additional dimensions in the 4-, 5- and 6-D solutions were unfathomable. No attempts at searching planes in the spaces for curved dimensions, or at distorting an acoustic surface to conform to the solution, were successful.

And neither linear regressions nor special strategies based on acoustic properties were of any help. It seems very likely that the extra dimensions beyond three were unique because of noise in the data. The Swedish listeners showed relatively low reliability and validity scores (Tables 4.2, 4.3).

The 3-D solution carried its own difficulties, too. One dimension was easily interpretable as rounding; the remaining two scales could not be interpreted individually. But when plotted against one another, a plane was formed that obviously reflects the labels height and backness, if the plane is rotated through 45° (c.f. Figure 6.24). Of course, this violates the PARAFAC model. PARAFAC solutions may not be rotated without destroying the relationship between the dimensions of the vowel space and the personal weighting factors. Different, unknown person weights would apply to the new vowel dimensions, but the new solution in its entirety would then be simply wrong, because it would no longer provide the best fit of the model to the experimental data.

Three possibilities are open to the investigator at this point. One approach, fruitless in this case, is to search higher-dimension solutions for more satisfying results. A second is to accept as fact the idea that the two dimensions in question are true perceptual scales for Swedish listeners. If this is done, then no further interpretation is possible, since the patterns on the dimensions taken separately do not correspond well to any linguistic, articulatory or acoustic properties. The third and most interesting approach is to assume that the model is being violated and attempt to derive meaningful dimensions by either changing the model or by treating the listeners' data in some way intended to overcome the violation.

Following this last suggestion, it is noted that [ʉ] was present in the stimuli heard by Swedes and was not present for anyone else. Now [ʉ], as a fourth high tense vowel, next to [i y u], is difficult to describe using distinctive features. Chomsky and Halle (1968) invent the questionable feature "+Covered" to set off [ʉ] from [y], both of which are [+High, -Back, +Round]. Perhaps the linguists' difficulty in describing this sound has a parallel in the perceptual realm. Is the perception of [ʉ] based on some curious idiosyncratic strategy such that its presence vitiates the perceptual model that held up much more sturdily for the non-Swedish listeners?

To follow up this possibility, the original lists of triadic comparisons were re-computed into matrices, leaving

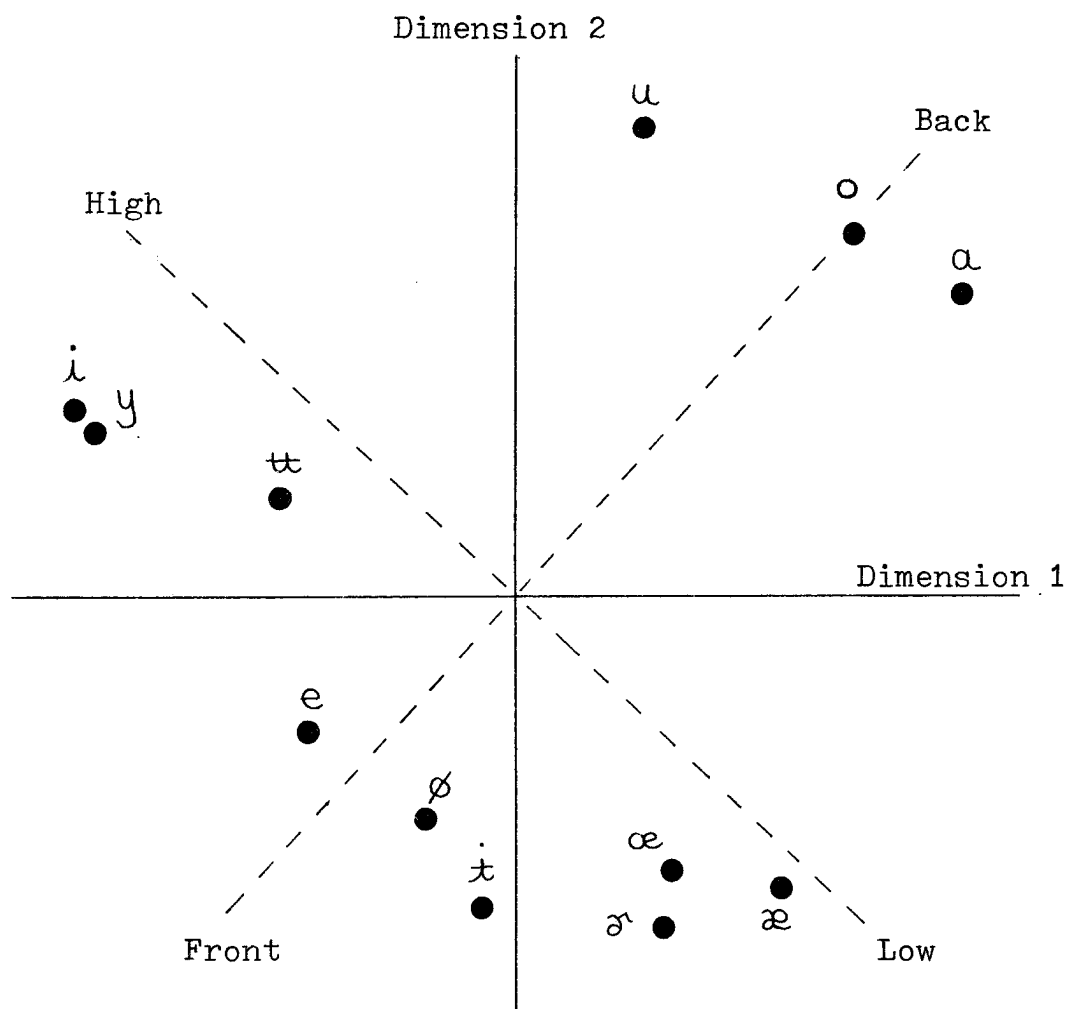


Figure 6.24: Height-Backness plane from the 3D PARAFAC solution for Swedish. Dotted lines show an interpretable but theoretically unacceptable rotation. Analysis includes all 7 listeners.

out all triads containing [ʌ]. COSCAL was applied to the vowel matrices to obtain additive constants for each listener in 1-5 dimensions. The constants were included in the computation of scalar products matrices. (See Section 4.4 for discussion of the additive constant problem and scalar products.)

When these revised data were analyzed by PARAFAC, it was found that there were two solutions at four dimensions, suggesting that the true dimensionality was three. The 4-, 5- and 6-D spaces were no more interpretable than the large 12-vowel spaces. But the 3-D space was essentially identical to the 3-D space it was intended to improve on. One dimension reflected rounding, and the other two effectively reproduced Figure 6.24 (with [ʌ] absent, of course, and with other trivial differences). The same unallowable rotation of axes stood between the results and a plausible interpretation. If the removal of [ʌ] reduced unsystematic tendencies in the data enough to eliminate uniqueness at four dimensions, it was still not enough to solve the main problem: Are there better dimensions for the 3-D space than PARAFAC has reported?

If one vowel is not the problem, perhaps some listener is. If one listener (or more) is behaving erratically, or in a consistent way that is not common to the others, his data will surely cause difficulties. But which listener(s)? None of the reliability or validity measures given in Chapter 4, nor the additive constants tabled there, nor any of the personal weights in the current solutions give strong evidence for discarding any particular listener. It will be shown that one listener was in fact behaving differently from the others, and that eliminating his data allowed PARAFAC to find a satisfying solution. But the discussion of this effort can only be made after the one remaining approach to the recalcitrant data is presented: the adoption of a more powerful MD-scaling model than PARAFAC.

As was reported in Terbeek and Harshman (1972), the Swedish results behaved very much like certain synthetic data analyzed by PARAFAC, namely, parallel sets of matrices (one matrix for each "person") representing various stretches and contractions of a common space, but constructed so that the axes were oblique. With both the Swedish and the synthetic data, there appeared (a) uniqueness beyond what was expected, (b) uninterpretable dimensions, and (c) planes that appeared to want rotation in order to improve their interpretation.

PARAFAC2, which relaxes the restriction that dimensions be orthogonal, was applied to the Swedish data (all seven persons, all twelve vowels) to determine whether an oblique orientation of axes can be found which gives a better solution.

The result in three dimensions consists of a dimension of Rounding, and one each of Height and Backness. The Rounding scale was found to be nearly orthogonal to the other two (92°). This was expected, given the original PARAFAC result, though it is gratifying to note that PARAFAC2 will allow axes to lie at right angles if that orientation gives the best fit. The rest of the space is plotted in Figure 6.25. Front/Back and High/Low do not meet at a right angle but at an angle of 71° .

In the article mentioned earlier (Terbeek and Harshman 1972), some attention was given to the meaning of the obliqueness itself. This is an important problem, especially if the oblique solution is taken as a reflection of each person's perceptual apparatus. The effect of the obliqueness on distances in the plane can be stated this way: The distance between two vowels that lie parallel to the bisector of the acute angle are expanded relative to the distance that would obtain if the angles were 90° . Examples from Figure 6.25 are [eɤ], [æɤ]. Distances more or less perpendicular to that bisector are compressed ([yɤ], [ɤœ]). The real question is, what in the sound system of Swedish would make the listener behave in this way?

There is good evidence that the obliqueness of the solution is an artefact caused by the data from one listener. The evidence appears in the form of one minus sign in the personal weighting factors (c.f. Appendix I) -- person 7, on the Back/Front dimension. With PARAFAC2, a negative person weight alters the orientation of two axes with respect to each other. The size of the angles involved remains the same, but the acute and obtuse angles exchange places. Therefore the negative weight for S7 on Back/Front means that while all other listeners are represented by Figure 6.25, S7 is represented by a version of that space in which the second and fourth quadrants (containing [yɤ] and [œœ]) have acute angles. For that person, back vowels in general are close to low front vowels. Low front vowels, in turn, are spread out considerably, compared to Figure 6.25.

When the angles between dimensions are sought by the program, one listener is therefore found to have a different set of angles, thanks to a curious mathematical exception to the PARAFAC2 restriction to the "same" set of angles for each person (namely, $|\cos \theta|$ is identical across subjects, but free to vary in sign).

The main point here is that one listener, S7, appears to have a perceptual space different enough from those of his compatriots to violate the assumptions of PARAFAC. Notice, however, that there is no evidence yet that each

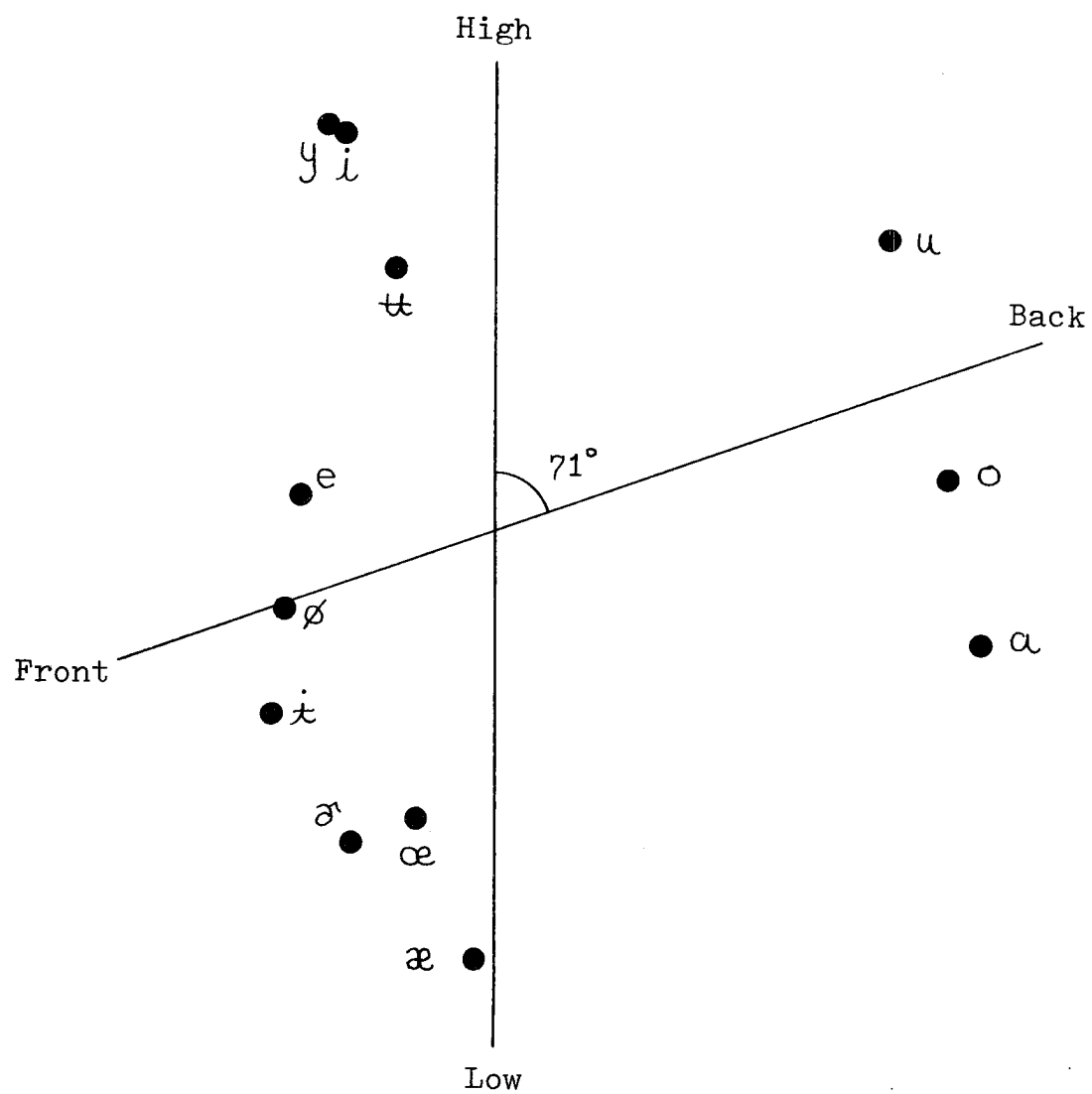


Figure 6.25: Height and Backness from the 3D PARAFAC2 solution for Swedish listeners.

listener uses an oblique space. What has been demonstrated is that if each listener's space is described using one common set of vowel dimensions together with personal weighting factors, the best fit is found by assigning one person a different orientation of axes. If the obliqueness is truly perceptually significant, an oblique solution ought to be necessary to adequately characterize the data when subject 7 is deleted.

This hypothesis was tested by submitting the data from subjects 1 through 6, all 12 vowels, to PARAFAC. The crucial two dimensions are plotted in Figure 6.26a. No obliqueness of axes is needed to enhance the easy interpretation of dimensions of the height-backness plane. The horizontal scale unambiguously represents a binary backness feature. The vertical dimension can only be interpreted as height, even though the scale of high-to-low is much smaller for back vowels than nonback ones. In choosing feature values to reflect vowel height, three degrees are necessary for the three equidistant back vowels. If the [e] and [a] stimuli correctly predict the perceptual relationship between native [ɛ] and [a], front vowels, too, need three feature values, giving high [i y u], mid [e ø], and low [æ]. [œ] is native but lax; it might be either mid or low.

As for the third dimension, rounding, two vowels disrupt the otherwise perfect binary division of the scale (c.f. Figure 6.26b). [y] and [o] appear halfway between the rounded and unrounded groups. Both vowels are apparently judged as being "between" two others. [y] lies between [i] and [ʉ]; [ɤ] and [ʉ] are distinguished only by rounding, hence the perception of [y] as "not as round as [ʉ]" but "rounder than [i]." Similarly, if it is postulated that [o] is perceived as "midway between" [a] and [u], then the position of [o] in the middle of the rounding scale as well as the height scale is understandable, in spite of being just as highly lip-rounded as [u]. These results support a three-valued rounding feature which is crucial only for distinguishing among [i y ʉ], and is redundant with height for the back group [a o u].

It seems very reasonable to conclude that the obliqueness of axes in the PARAFAC2 solution is not perceptually real. It appears to be the result of accomodating one listener whose vowel configuration was different from that of the six other listeners. This remark, however, is intended to imply neither that the explanation of the listeners' behavior is complete, nor that PARAFAC2 is somehow diminished in usefulness. More work is needed to discover what causes one person's perceptual space to differ from another's. And PARAFAC2 was the means by which it was in fact discovered

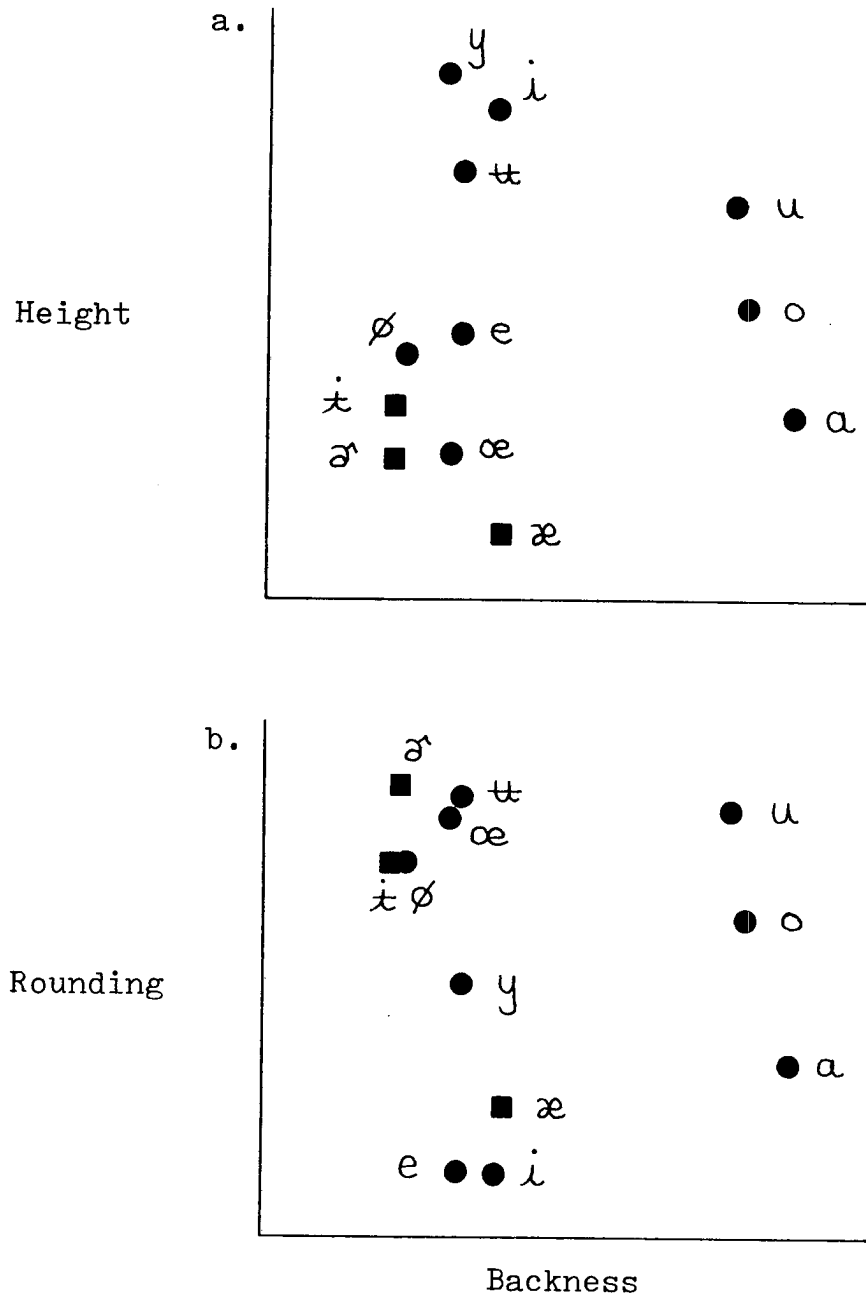


Figure 6.26: High/Low, Round/Nonround, and Back/Nonback from the 3D PARAFAC solution for Swedish using 6 listeners.
 ● = native vowels; ■ = non-native.

that listener #7 was responding in a way, relative to the other subjects, that violated the PARAFAC model.

Figure 6.27 summarizes the evolution of the perceptual space for each of the four analysis types: 1) PARAFAC with 12 vowels, 7 listeners; 2) same as (1) but with one vowel deleted; 3) same as (1) but with one person deleted; 4) PARAFAC2 with all 12 vowels and 7 listeners. Rounding is generally the most salient dimension, with Height and Backness about equal in importance. No 4-D solutions were interpretable, even though unique solutions existed beyond that point. This is attributed to a lack of consistency between and within listeners.

6.52 Acoustic sources for Swedish perceptual dimensions

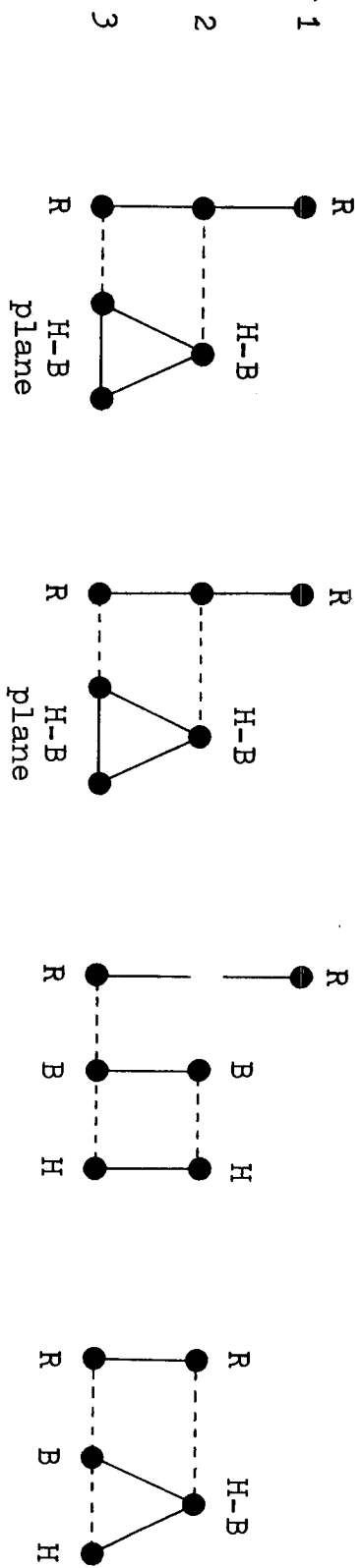
Figure 6.28 summarizes regression analyses performed on the 3-D PARAFAC2 space. Back/Nonback is the only one of the three dimensions that is well-described by linear regression. It is accounted for by a weighted sum of the second and third formants (Figure 6.29a). This is an unexpected result -- no other language used F3 as one of the main predictors of backness. There is another, equally plausible source for the backness dimension, however, as can be seen of Figure 6.30. Here the Swedish stimuli are plotted according to F1 and F2; the line which separates [ʊ ɔ ɑ] from the other vowels represents the simple expression $F2 - F1$ (the "dimension" of backness is perpendicular to the line on the figure). If it is postulated that all vowels on the left side of the backness boundary are heard as the "same" on this dimension, that is, that distance from the boundary is irrelevant, then F2 and F1 account for the backness pattern as well as F3 and F2.

For Rounding, F3, A0 and F2 combine to predict the pattern with a multiple r of .90. This is sufficient to account for the main gap in the pattern separating unrounded [i e æ ɑ] from the rest (Figure 6.29b). If only the formant frequencies are used, the predictions are not satisfactory. The two-formant regression places [ɤ] among the unrounded vowels, whereas the listeners heard it as rounded. The three-formant regression barely improves upon this incorrect location of [ɤ]. Once again, a non-linear combination of F1 and F2 easily distinguishes rounded from non-rounded vowels, illustrated by the curve moving downward from left to right on Figure 6.30. A straight line (approximately $F1 + .33F2$) could make the same cut, but without the elegance of having the cut fall midway between each of the pairs [ɪ y], [e ø], and [æ œ].

What this proposal cannot account for is the placement of [o] and [y] towards the middle of the rounding dimension. Regression analyses cannot account for it, either. If only the rounded vowels are considered, their position on the rounding scale is closely related to their respective distances

Number of
dimensions
present

Interpretation of dimensions



program	→	PARAFAC	PARAFAC	PARAFAC	PARAFAC2
number of listeners	→	7	7	6	7
number of vowels	→	12	11	12	12

Figure 6.27: Summary of the PARAFAC and PARAFAC2 solutions for Swedish, in 1 through 3 dimensions. R = Rounding; H = Height; B = Backness.

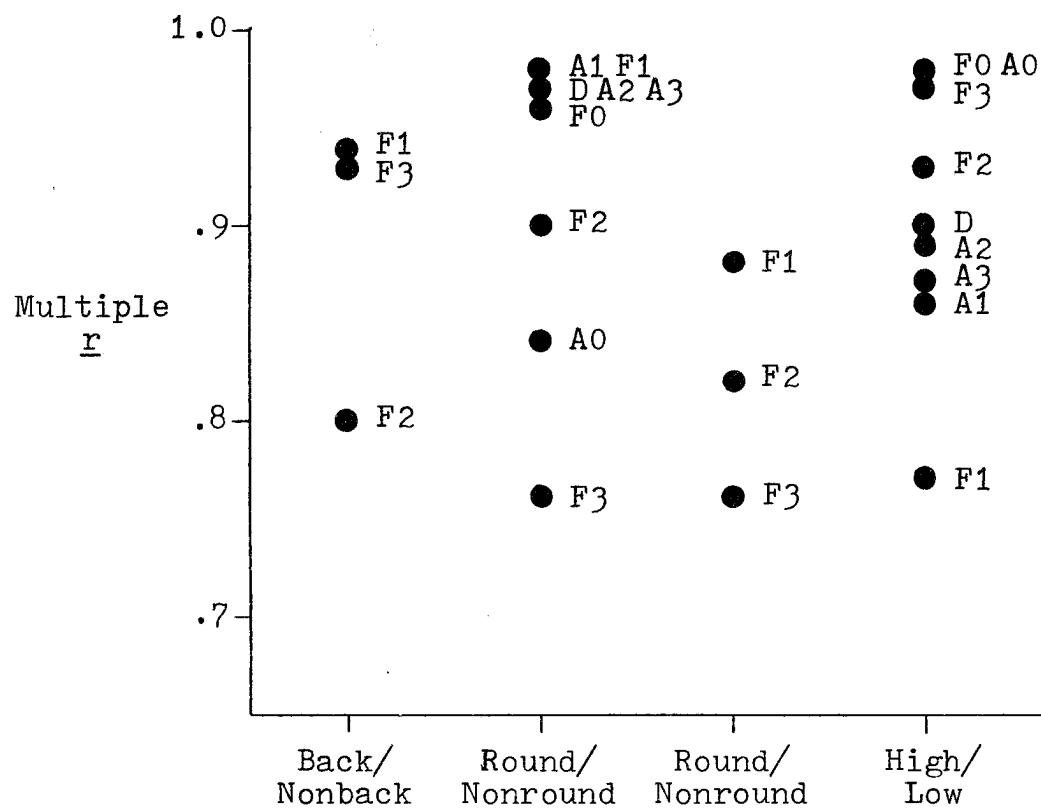


Figure 6.28: Summary of regression analyses, 3D
PARAFAC2 solution for 7 Swedish listeners.

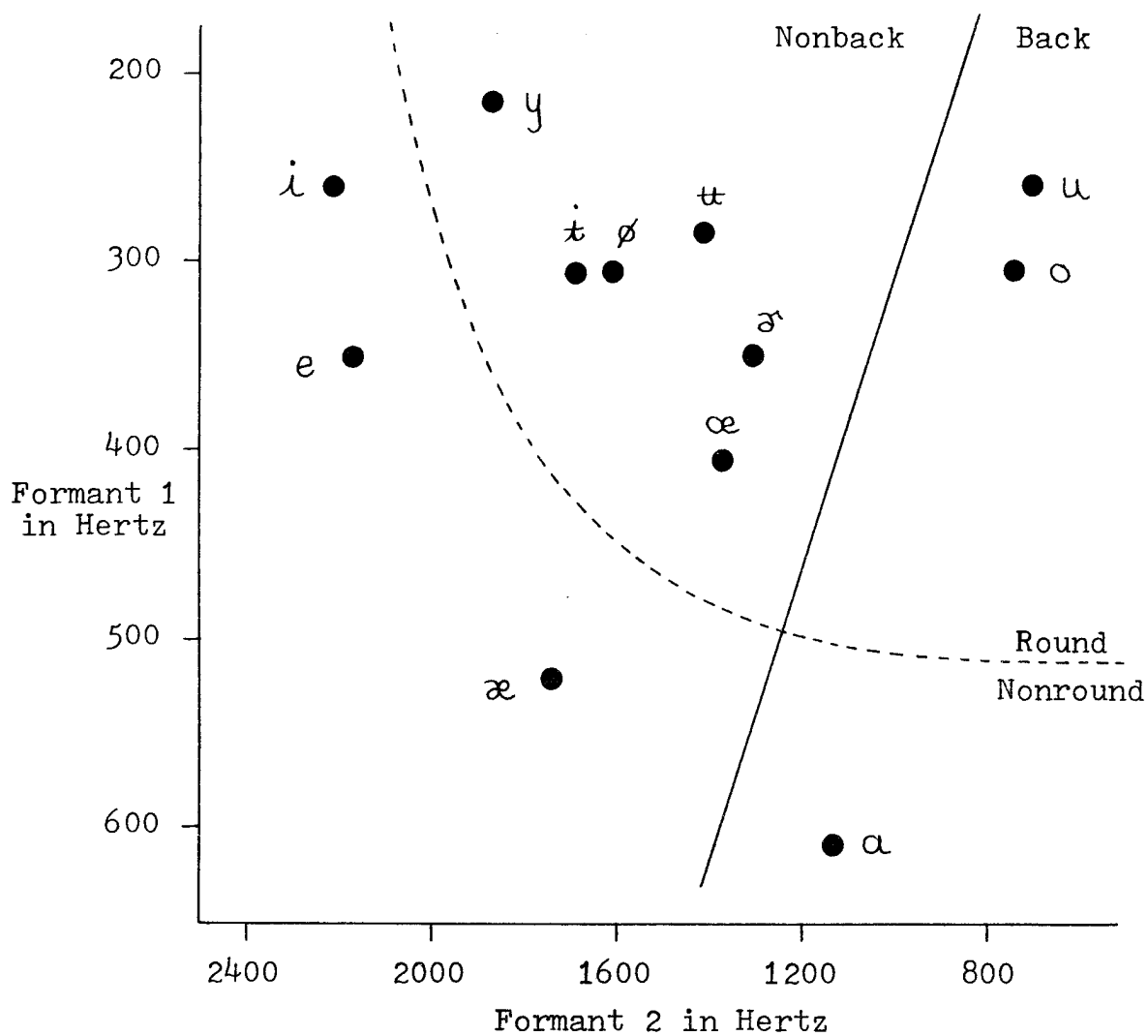


Figure 6.30: Possible strategies relating 2 dimensions of the PARAFAC2 solution for Swedish to the F1-F2 plane.

from a point on the F1-F2 plane near [ɤ] (say around 330 Hz, 1220 Hz). In this scheme, however, [u] is out of place. An ad hoc suggestion that works is that rounding values for high front vowels are based on F2 alone. For back vowels, rounding judgments inexplicably copy height judgments.

According to the regression analyses, the last dimension, High/Low, is only weakly related to F1 ($r = .77$), with four extra variables needed to reach .90 (Figure 6.28). But this poor correlation is deceptive; the order of vowels on High/Low (Figure 6.29c) is very similar to that for F1, with the important condition that back vowels be examined separately from front vowels. [u o ɔ] are obviously ordered according to their first formants. Perceptually, the distance from [u] to [o] is about the same as from [o] to [ɔ], although these near-equidistant steps are not at all reflected in the acoustic space. The F1 relationship for the front vowels is flawed by the placement of [ɛ] and [ɛ̃] (both non-native), but otherwise quite acceptable.

From the above discussion and the shape of Figure 6.26a, it is easy to conclude that the High/Low scale for front vowels is perceptually greater than the High/Low scale for back vowels. In other words, as one moves from back to front, F1 becomes more salient. Speculation as to the cause of this takes two directions. 1) Listeners may have come to regard the "distance" between adjacent vowels as roughly equal, although the actual acoustic distances between front vowels are smaller. 2) The effect of tongue and jaw movement on acoustic shape is smaller for back vowels than it is for front vowels. During language acquisition, one might become conditioned to match perceptual distances with some sort of "articulatory distance."

Interesting as these suppositions may be, it is difficult to answer them in a language-specific way, which is necessary. Why should the Swedish listeners behave supposedly according to facts about articulation, when the other four groups do not? Furthermore, Germans have more front vowels than back ones. Why are their results different from those of the Swedish listeners? There is not as yet enough information to answer this with any certainty.

6.6 Discussion

In comparing the results from the five languages taken separately, it is quite plain that neither Hypothesis 2 nor an extreme version of Hypothesis 3 is correct. The true state of affairs is a blend of the two. The findings that lead to this conclusion are summarized by the two remaining figures. Figure 6.31 gives a capsule account of the acoustic source

for each of the dimensions reported in this chapter. Figure 6.32 compares the personal weighting factors that emerged for those dimensions.

Hypothesis 3, stated as rigidly as possible, holds that no two languages can be expected to share any perceptual dimension. The data are hardly so anarchic as to support such a view (c.f. Figure 6.31). Only five phonological concepts are required to describe all the dimensions which appeared in this chapter: height, backness, rounding, retroflexion, and peripheralness. Furthermore, every solution presented involves information about height, backness, and rounding. There is little reason to expect such a convergence to highly similar dimensions by chance alone. A plausible conclusion is that there is a universal perceptual framework of possible dimensional types, based on a limited class of manipulations of the frequencies of the first three formants.

The exact patterns found certainly differ from one language to the next, as discussed below, but there can be no doubt as to the underlying consistencies out of which the differences grow. One prominent dimension (height) is based on F1, another (backness) on F2 or a weighted difference between F2 and F1, and another (rounding) on a weighted sum of F1 and F2. F3 can underlie a dimension independently, or contribute to distinctions made primarily according to F1 and F2. The sort of non-linear use of F2 and F1 found in Peripheral/Central seems to be relatively rare, but may figure significantly in the perceptions of some speakers.

Although there is a language-universal background of similarity, there is sufficient variation in the results to refute Hypothesis 2, the claim that all languages use the same dimensions, with the relative importance of the dimensions varying among languages. The most obvious evidence is that some dimensions are unique to specific languages. Retroflex (English), Back/Nonback2 (Turkish), and Peripheral/Central (Thai), are the three such dimensions that are most easily related to their respective vowel systems. Only English listeners need to distinguish retroflex [ʂ] from non-retroflex vowels. Only Turkish listeners need to perceive [ɰ] as phonologically [+back], which is reflected in the extra backness scale. And only Thais have two central non-round, nonlow vowels [ʌ ɤ]; these listeners have more interest in the non-peripheral region of the vowel plane than do any others. Highround/Low Nonround, the other dimension unique to Thai, is more difficult to account for, as is the rotated appearance of the plane formed by this scale and Back/Nonback.

The two remaining language-independent dimensions, according to Figure 6.31, are Mid/Nonmid (English) and Low/Nonlow (German). It is not difficult to accept these scales as components of curved dimensions of height, caused by listeners' underestimation of extreme height differences. But it remains for future studies to determine why the phenomenon ought to affect these two languages and not the others -- if indeed the results can be replicated.

A slightly more subtle kind of evidence against Hypothesis 2 is the fact that some dimensions "re-appear" in different solutions in broad outline only. For example, the acoustic source for the label Back/Nonback differs from language to language, although F2 is involved in every case. German and Swedish use different linear functions of F2 and F1, English and Thai involve F1 in a non-linear way, and Turkish uses F3 in judging certain vowels. The height dimensions, all based on F1, are not identical, either. German, Turkish and English incorporated some non-linear stretching or compression of distances, whereas Swedish uses information from F2; in Swedish, exact height values depend on backness. The Round/Nonround dimensions have the most consistent acoustic source, each being derived from a weighted sum of F1 and F2. And yet, two of those functions (for Turkish and Swedish) are only fair approximations to a more satisfying curve drawn through the F1/F2 plane. Thai, of course, does not seem to use rounding independently at all.

The personal weights given in Figure 6.32 afford a final comparison between the results of this chapter and the previous one. Two points are of primary interest. First, several solutions have patterns of personal weights that differ considerably from their counterpart patterns in the five-language space. For example, rounding is clearly the most salient dimension for Swedish listeners considered separately; the curious prominence of Peripheral/Central noticed in Chapter 5 is completely absent. This kind of finding supports the idea that the five-language analysis entails a distortion of a language's true perceptual space, producing spurious personal weights to accompany the misleading dimensions. In this case, the distortion was augmented by the data of listener #7, the inclusion of which made the PARAFAC model itself undependable. For English, the joint analysis showed high salience for Low/Nonlow, and considerably less for High/Nonhigh. But the separate English analysis shows high salience for High/Nonhigh, and comparatively little for Mid/Nonmid. One cause of this difference may be the absence of two listeners in the separate English solution. Nevertheless, in both analyses, vowel height emerges as a relatively important dimension.

With Thai, it is difficult to tell which scales count as "counterpart" dimensions in the two solutions, considering the large difference in dimensional structure. Except for Peripheral/Central, the personal weights are incomparable. What is maintained in both solutions is the slight prominence of height and rounding, even though these properties are compressed into one dimension in the separate Thai analysis.

The second main point of interest involves the difference between the three languages just mentioned and the remaining two, German and Turkish. The greatest saliences for German and Turkish are for Back/Nonback, whereas for English, Thai and Swedish, the dimensions containing rounding information have the highest values. This is consistent with the results of the five-language space (Chapter 5) and the earlier hierarchical clustering studies (Chapter 4). However, this result means little unless it can be shown that the difference between backness and rounding weights for individual listeners depends upon native language.

Matched-pairs t-tests were performed on the rounding and backness values to confirm or disconfirm such a relationship. German and Turkish listeners, regarded as one group, showed a significant predominance of backness over rounding ($t=2.24$, $p(13) < .05$, two-tailed). English, Swedish, and Thai were also treated as a single group. (Since Thai has no Round/Nonround dimension per se, Highround/Lownonround was used as the dimension representing rounding. The single English imaginary value was entered as zero.) Here, the results were significant in the opposite direction: rounding predominated over backness ($t=3.24$, $p(17) < .01$, two-tailed).

These statistics establish a clear difference in the relative importance of backness and rounding for German and Turkish on the one hand and English, Thai and Swedish on the other. There is nothing to be found in the structure of the languages that matches this difference except the presence vs. absence of a phonological rule which adjusts backness. The proposed explanation is that umlaut (for German) and backness harmony (for Turkish) have imposed a heightened awareness of the front-back dimension. In the absence of such a rule, rounding is the most salient scale.

It would be very satisfying to find that analyses of other languages support this conclusion. Unfortunately, the re-analyses of data from Pols et al (1969) and Hanson (1967), discussed in Chapter 2, are **equivocal**. Hanson's Swedish data yielded a straightforward backness dimension, but the rest of the 3-D solution consisted of a plane rotated 45° away from the expected dimensions of height and rounding. The personal weightings on backness are unusually high, but

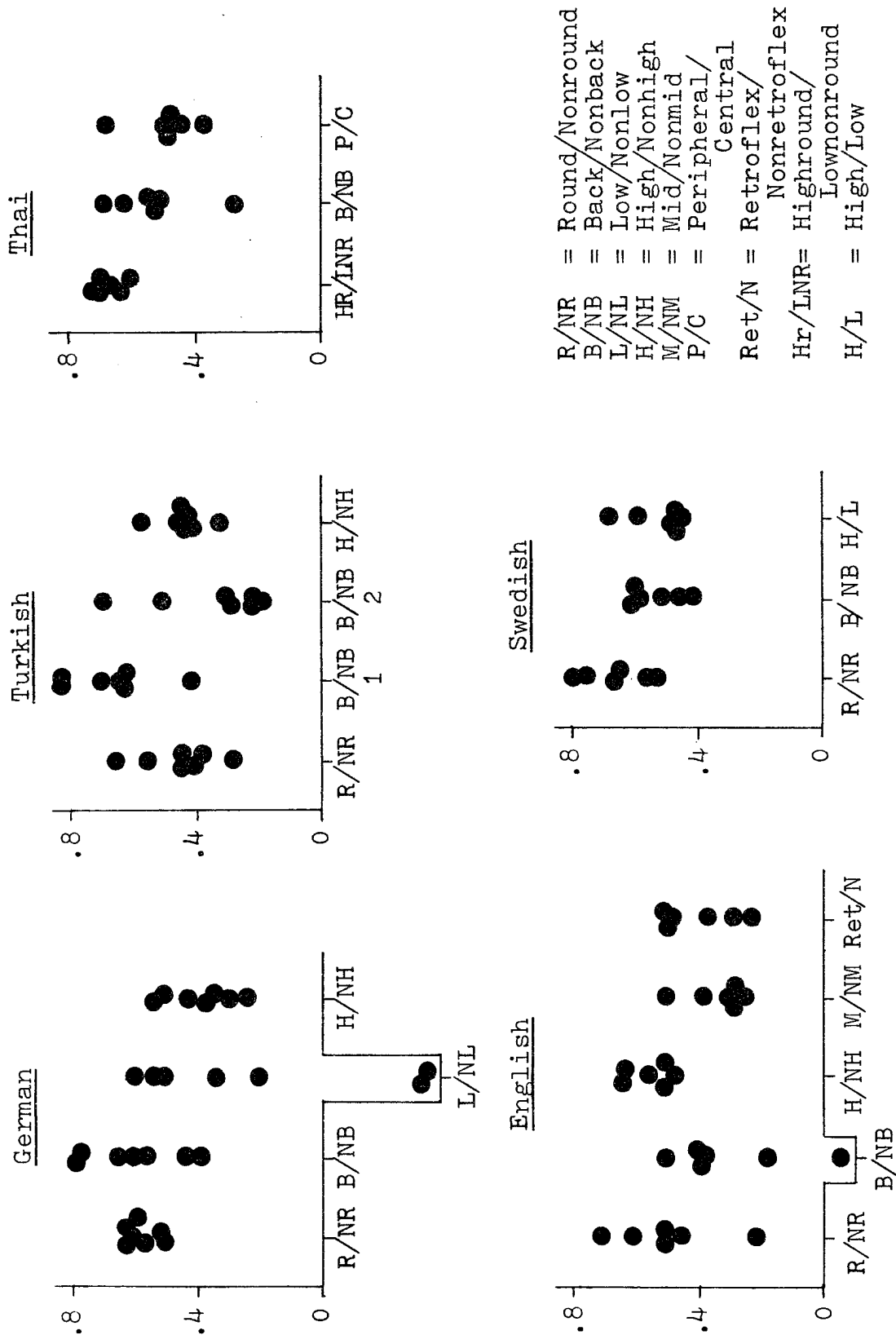


Figure 6.32: Personal weighting factors for language-specific solutions discussed in text. Swedish results are from the 6 person PARAFAC solution.

this may be because the other two dimensions are in some way spurious. It was too expensive to use PARAFAC2 to identify unusual subjects for deletion, a technique which proved successful with the present Swedish data. As for the Dutch data from Pols et al, the best solution was a four-dimensional space, re-interpreted as two curved dimensions. One of those contained both backness and rounding information; no comparison of backness and rounding as separate properties was therefore possible.

The relationship between phonological rules and dimensional salience needs to be explored further. Future studies should involve good-quality native vowels, a larger collection of test languages, and phonological rules in addition to the ones found to be of interest here.

Chapter 7

Conclusion

7.0 This final chapter relates the results of the MD-scaling analyses to several linguistic issues. In Section 7.1, observations regarding discreteness of feature values lead to support for the idea that features are not necessarily independent. Section 7.2 summarizes the evidence supporting the existence of multivalued features and opposing the Jakobsonian restriction to binarity. The phenomenon of maximal perceptual differentiation for phonemes is examined in Section 7.3. Section 7.4 summarizes the main findings of the entire experiment.

7.1 Discreteness and feature interdependence

It is generally agreed in linguistics that understanding spoken words is a matter of analog-to-digital conversion -- the continuous acoustic stream must be interpreted as a sequence of discrete segments, or phonemes. (It makes no difference to the point at hand whether the phonemes themselves are actual units of perceptual processing.) It is further agreed that there are properties, or features, which cut across the phonemes, grouping them into classes, defining their similarities and differences (c.f. Chapter 1). It follows that the features themselves operate with discrete categories: for a German vowel to be identified as either [y] or [ʏ], for example, it must be judged to be either rounded or not.

The present results are perfectly in line with this view. There are many instances of vowels clustering together on a perceptual dimension where such clustering could not be found in the acoustic information. The best examples of this phenomenon are seen in the Turkish rounding and backness dimensions (Figure 6.10a, b, d), and backness for Swedish (Figure 6.29a). The conclusion to be drawn from these examples is that listeners tend to disregard acoustic differences along a dimension if the vowels being compared fall into the same category. Subjects are demonstrating categorial perception for distinctive features.

This within-category clustering is not found for all dimensions discussed in Chapter 6. It is not present in the Thai space at all; it is hoped that further experiments with Thai will help to resolve this and other puzzling aspects of

the Thai data. The remaining instances where no categorial clustering is observed are the height dimensions for Turkish, English, and Swedish, and rounding for Swedish.

The lack of clustering is a direct result of considering the dimensions independently. This is equivalent to a demand that a particular range of values along a dimension must be assigned a certain feature value with no reference to any other features. Such a condition would cause difficulties in interpretation, if it were imposed strictly. With the English height dimension (Figure 6.23c), [u] and [e] are much too close together to justify a cut between them; yet when height is considered in connection with rounding (Figure 6.21a), it is clear that unrounded [ie] is parallel in height to rounded [uo]. Defining three levels of height for [ɪ] vs. [ue] vs. [o] is simply a mistake.

For Turkish and Swedish, positions along the height dimension are assigned feature values most efficiently if backness has been assigned first (Figure 6.3b, Figure 6.26a). In both languages, some clustering can be observed in the front vowels, once the back vowels are removed. Similarly, the Swedish rounding dimension (Figure 6.26b) shows clustering when front vowels are taken separately.

These observations imply that the real-time processing of vowel sounds by the listener may require sequential operations. To say the feature x depends on feature y suggests that y is determined before x. Choice reaction time is an appropriate method for exploring this hypothesis. Taking the English case as an example, it could be predicted that a listener requires more time to decide that [ɪ] and [e] are different than he needs to decide that [ɪ] and [u] are different. The reasoning is as follows: [ɪ] is presented, and the listener is given sufficient time to analyze it as +high and -round. Since the task is to respond as quickly as possible, assume that the response will be made as soon as the second stimulus is found to have any feature value that differs from the values for [ɪ]. Since rounding is hypothesized to be evaluated first, [ɪu] will have a faster reaction time than [ɪe]; in the latter case, two features must be examined before a difference is found. The proposed effect of backness on other dimensions in Turkish and Swedish is subject to the same kind of testing.

It would be gratifying to find that the present work, dealing with listeners' overt judgments of similarity and difference, is closely related to the study of mental operations performed in the act of speech perception. The choice reaction time approach is a suggestion with this purpose in mind.

7.2 Binary vs. multivalued features

How many distinct categories may appear on any one dimension? Much heat has been generated in the ongoing debate between those who insist that features must be binary and those who do not. The tradition of insistence upon binarity can be traced to Jakobson. In Preliminaries to Speech Analysis (1952), the following concise statement of position is found:

Any minimal distinction carried by the message confronts the listener with a two-choice situation.... The listener is obliged to choose either between two polar qualities of the same category, such as grave vs. acute, compact vs. diffuse, or between the presence and absence of a certain quality, such as voiced vs. unvoiced, nasalized vs. non-nasalized, sharpened vs. non-sharpened (plain). The choice between the two opposites may be termed distinctive feature [*italics sic.*] (p. 3).

Jakobson does not present any evidence for disallowing non-binary features, and the above statement must be taken as a postulate, not an empirical fact. Nevertheless, this position has become well-entrenched in modern phonology, as exemplified in Halle (1959) and Chomsky and Halle (1968).

The "opposing" viewpoint, represented by Ladefoged (1971) and Trubetzkoy (trans. Baltaxe 1969), agrees that binary distinctions play an important role in phonology. Ladefoged finds that twenty of the features he proposes are always binary (p. 91); Trubetzkoy grants that "gradual oppositions [discrete but multivalued. DT] are relatively rare and not as important as privative oppositions [presence vs. absence of a property. DT] (p. 75)." In terms of vowels, neither author sees any evidence for more than two degrees of backness ("timbre") or rounding at the phonemic level. But both Ladefoged and Trubetzkoy explicitly allow for the existence of some multivalued features. In particular, they agree that vowel height ("degree of aperture" for Trubetzkoy) often involves two, three, or four distinctive steps.

The perceptual spaces from Chapter 6 can be taken as evidence relevant to the multivalued vs. binary feature dispute only if it is agreed in advance that experimentally derived perceptual dimensions are equivalent to perceptual features. This condition is not a tautology. If a PARAFAC dimension were to show three unmistakable degrees of vowel height, a proponent of binarity might claim that the real perceptual features are higher-order linguistic functions derived from that dimension and entail the two binary deci-

sions of high/nonhigh and low/nonlow. Even so, however, a multivalued PARAFAC dimension argues for some level of perceptual organization which is not binary. Of course, if vowel height (to continue the example) for some language is truly a matter of binary features, PARAFAC analysis should show only dimensions with two vowel groups (such as high/nonhigh and low/nonlow).

With this condition firmly established, there is considerable support for the multivalued nature of vowel height. English (Figure 6.21a) requires four height levels for distinguishing [i e ʌ α]. For German, Turkish and Swedish (Figures 6.3a, 6.8b, 6.26a), there are three distinct levels of height for the back vowels [u o α], even though the height distinction for [o α] is redundant with rounding. [o] and [α] are judged to be too different in height to be collapsed to the same point on the height dimension. For the Thais (Figure 6.13a), no matter how the height/backness plane is interpreted, three levels of height are necessary for distinguishing [i e æ] and [ʌ o α], and a case could be made for a fourth level for [u].

In none of the analyses did the Jakobsonian binary features of ⁺high and ⁺low appear, including the German and English spaces, which each required two PARAFAC dimensions to express the listeners' judgments about height. The original, linear PARAFAC dimensions of Low/Nonlow for German and High/Nonhigh for English still require three and four height levels, respectively.

The backness dimensions, with the exception of Thai (which will not be discussed) and Turkish, can be interpreted best as binary. This is predicted by both sides of the binarity dispute. Swedish (Figure 6.29a) can be interpreted no other way; vowels cluster in two unambiguous groups. German would be just as good but for [æ], which lies midway between the front and back groups. As a non-native sound, located centrally on the acoustic scale underlying backness (Figure 6.5a), [æ] was not systematically attracted to either the front or the back group by the categorical perception phenomenon described in the preceding section. The English backness dimension is unusual. All vowels with a sufficiently low F2 cluster together to form a +back group, but the remaining vowels appear in three groups, [æ] constituting the frontmost one. As discussed in the previous chapter, backness has no clear function except to separate [æ] from [ʌ] and [α]. This dimension seems to be more salient for low vowels, causing [æ]-[α] to show a much greater perceptual distance than would be predicted by F2. The best conclusion is that backness is used in a binary fashion, to the extent that it is linguistically useful at all.

For Turkish, binarity is not in the ear of the hearer, but in the mind of the analyst: the analysis can be viewed as supporting either a binary or a ternary backness dimension. The two PARAFAC dimensions of Back/Nonback1 and Back/Nonback2 are each composed of two widely separated clusters, the difference being mainly in the position of [ɤ] (Figure 6.10a, b). To preserve binarity, one only needs to postulate a two-valued phonetic backness scale, which groups [ɤ] with front vowels, and a phonological scale, which Turkish speakers have developed to reflect the fact that [ɤ] behaves phonologically like back vowels. On the other hand, the backness curve based on these two dimensions can be defended as well (Figure 6.10c). No vowel can lie on the backness plane except along the curve. The acoustic source for the backness curve is just as reasonable as the source for Back/Nonback1; there is no reasonable acoustic source for Back/Nonback2. More work with Turkish is called for. But unless it can be shown that any given listener uses only one of the two backness dimensions, the binary vs. ternary issue for this particular problem will remain unresolved.

Rounding, Ladefoged claims, is always a binary feature. Three of the languages with a rounding scale conform to that claim; one, Swedish requires three levels. Turkish in this case is the textbook demonstration of binarity (Figure 6.10d). The English rounding dimension (Figure 6.23a) does not arrange the vowels as neatly as does the Turkish, but when considered as part of the height/rounding plane (Figure 6.21a), there are clearly only two degrees of rounding present. As for German, the round/nonround dimension taken alone (Figure 6.5b) consists of three groups: unrounded [i e æ ʌ a ɔ], rounded [y ø ɤ ɶ], and rounded [u o], conforming very closely to a linear function of the F1/F2 plane. When the rounding/backness plane is taken as a whole (Figure 6.3b), there are only two levels of rounding for front vowels, and two levels for back vowels.

The three-level structure of rounding for Swedish, with [y] and [o] taking the third, intermediate value, has been described explicitly in Section 6.5. This is not the first time it has been suggested that ɤ-round is not adequate for describing Swedish rounded vowels. Ladefoged (1971) credits Malmberg with the idea that [ɤ] is pronounced with lip compression, whereas [u o y ø] are pronounced with truly rounded, protruded lips (p. 77). Lindau (personal communication) points out that [y] and [ɤ] are often diphthongized (to the point of producing audible friction) in a way consistent with Malmberg's view. [y] becomes [yʲ] and [ɤ] becomes [ɤʲ]. But this phenomenon is completely different from the one at hand: [ɤ], for the seven Swedish listeners, groups with

[u] and [ø], while [y] and [o], which are said to have the same kind of lip activity as [u] and [ø], do not. It appears that the correct way to specify the production of rounded vowels and the correct way to specify their perception are different. Further research is necessary to determine if the perceptual dimension of rounding for Swedish constitutes a true counter-example to Ladefoged's restriction of rounding to two phonemic levels.

The remaining two dimensions that appeared in Chapter 6, retroflex and peripheralness (Section 6.4, 6.3), are binary in function, although the vowels do not form neat clusters. It would be surprising to discover a language that required more than two levels for these features.

7.3 Phonemic contrast and perceived distances

If any collection of three through twelve vowels could constitute a language's vowel system, there would be over 30,000 possible systems, given the unrealistic, conservative assumption that there were only fifteen vowels from which to choose. The only principle constraining the structure of observed vowel inventories would be that a set of phonemes must be consistent with the set that it evolved from, historically. But a series of small, unpredictable changes over many language families over many centuries ought to lead to a collection of vowel systems little different from a random sample out of the possible thirty thousand.

In reality, there are surprisingly few vowel systems, as attested mainly by Trubetzkoy (1969) and Hockett (1955). Many differences that do exist are dwarfed by persistent, recurring underlying patterns. The desire to formalize what is known about the recurring patterns (among other goals) led to Chomsky and Halle's theory of markedness (Sound Patterns of English, 1968). Markedness involves labelling phonemes and feature values as unmarked, or relatively common and therefore expectable, as opposed to marked, or relatively rare and therefore less expectable.

Unfortunately, markedness makes no contribution to an understanding of the principles that cause certain patterns of vowels to appear again and again, no matter how well it may (or may not) summarize those patterns. One example will suffice. The expression

$$[u \text{ round}] \rightarrow [\alpha \text{ round}] / \left[\begin{array}{c} \alpha \text{back} \\ -\text{low} \end{array} \right]$$

(part a of marking convention XI, p. 405), which states that rounding for nonlow vowels usually matches backness, does not

explain why a language L has [ɪ] and not [y], or why no language has [ø] unless it also has [e]. The marking convention simply reflects expectations.

Ladefoged (1971) speculates that facts such as the rounding-backness relationship just mentioned "are probably due to some kind of principle of maximal distinctiveness whereby the auditory differences between the vowels in a language tend to be kept at a maximum (p. 76)." This proposal was recently re-affirmed in more detail (Ladefoged, 1975):

Perceptually what matters is that sounds that affect the meaning of a word should be as distinct from one another as possible. This can be achieved by maximizing the perceptual distance between the sounds that occur in a contrasting set, such as the vowels in stressed monosyllables.

...If the vowels of a language are to be maximally distinct, the formant frequencies will be such that the vowels are as far apart as possible when plotted on a vowel chart. Consequently there is a natural tendency in languages for vowels to be spaced at approximately equal distances apart and for them to be on the outside of the possible vowel area (pp. 236-237).

Liljencrants and Lindblom (1972) explored this hypothesis by having a computer predict the location of vowels on the F1/F2 plane given (1) a fixed boundary for the vowel area, and (2) an inverse square repulsion law as a first approximation to maximizing intervowel distance. This was done for 3-12 vowels. The results are extremely difficult to evaluate, mainly because the process of comparing the predicted formant positions with published summaries of vowel phoneme systems is fraught with arbitrary decisions in matching one kind of description to the other. Nevertheless, the authors find that "the model produces about nine clear errors in a comparison involving 75 vowel qualities (p. 854)." They conclude that the idea of maximizing acoustic distance is a likely factor in determining vowel systems.

One kind of evidence against a distance-maximizing principle which operates even for large differences is the presence of curved perceptual dimensions. With such a dimension, the extreme values are perceived as being considerably closer to each other than predicted by the acoustic properties. (The English and German height dimen-

sions are good examples.) This seems incompatible with the idea that vowels have been forced to the ends of that dimension to meet the needs of the listener.

A criticism of Liljencrants and Lindblom's method which they themselves acknowledge is that the exact phonetic properties of the real vowel systems against which they measured their predictions were not taken into account at all. The phonemic symbols, except for [ə] and [ø], were all assigned formant values on the periphery of the formant plane. This greatly increased the chances of finding close matches, considering that all vowels generated by the computer program, except [ə], were pushed to the periphery by the distance maximizing procedure. (The model was incapable of generating an [ø]-like vowel.)

Their comparisons of real vs. predicted vowels cannot be taken at face value because real vowels do not necessarily behave in this way. Turkish speakers, for example, pronounce the /i/ and /u/ phonemes markedly lower and less tense than cardinal [i] or [u]. But this fact is concealed by the comparison of 8-vowel systems. Turkish /i/ and /u/ are claimed to match perfectly with the model's prediction of extremely high and maximally separated [i] and [u] (p. 853).

Evidently the principle of maximum acoustic contrast fails to account for some facts about vowel distances. It seems that there is a point of diminishing returns in maximizing acoustic distance. A plausible speculation is that some minimum distance between sounds must be attained, representing a tolerable level of confusability, beyond which there is no advantage to any further increase in distance, especially, perhaps, if that increase requires extra articulatory effort for one or both sounds. Unfortunately, this modified explanation taken by itself would predict the existence of vowel systems such as [ɛ ə ɔ] or [e ø æ ɑ ɔ], or [i e ɛ ʌ u]. These groups contain attested contrasts but do not occupy the whole vowel space the way real systems do.

A model of vowel distances that takes this observation into account ought to consider Stevens' (1971) hypothesis that vowels tend to maintain (or evolve towards) acoustic positions which are stable in spite of a large amount of articulatory variability. He finds that [i], [u], and [ɑ] are the main vowel target regions with this property (p. 221).

Stevens' acoustic stability principle, then, would account for the expected presence of [i], [u], and [ɑ]-like phonemes, and perhaps others as well. The principle of maximizing acoustic distance would act to keep relatively close vowels

above the level of confusability. Finally, a distance-minimizing tendency due to coarticulation and inertia of the vocal tract would keep vowels away from the periphery of the vowel region, subject to the opposing tendency of the first two principles.

The proposal given in the previous paragraph for explaining the distances between vowel phonemes is an improvement over the simple concept of maximizing acoustic distance which began this discussion. Nevertheless, that proposal still incorporates the assumption that changing the distance between two vowels means pronouncing the vowels differently -- changing the acoustic nature of one or both vowels. This view is too limited, in that it does not include the very real contribution made by the listener himself in establishing perceptual differences. The results presented in Chapter 6 unequivocally show that perceived distances are not based solely on acoustic properties, but depend upon the habits and expectations imposed by the listener's native language. The kind of listener contribution appropriate to maximizing perceptual distances can be stated explicitly as a testable hypothesis (HP), as follows:

Language L has two vowel phonemes, a and b, which are relatively similar acoustically. The native learner of L will acquire a sensitivity to that region of the acoustic space that allows him to distinguish between a and b. This learned sensitivity will be reflected in the listener's perceptual judgments. The distance from a to b will be judged to be greater by the native of L than by a native of a language without such a distinction.

Some evidence supporting HP has already been presented in Section 7.1. Clustering of vowels along a perceptual dimension clearly implies an increase in distance between two vowels that lay on opposite sides of a categorial distinction. A more direct way to demonstrate that HP is correct is to examine vowel distances computed from the individual language solutions given in Chapter 6. Eight acoustically close vowel pairs were selected for this purpose, taking Euclidean distances based on unscaled formant frequencies as a neutral measure of acoustic distance. Figure 7.1 gives the distances in both F1/F2 and F1/F2/F3 space for reference. Before the perceptual distances were computed, each dimension was multiplied by the mean personal weighting factor for that dimension. (Imaginary weights were ignored.) The distances are therefore based on both dimensional structure and relative

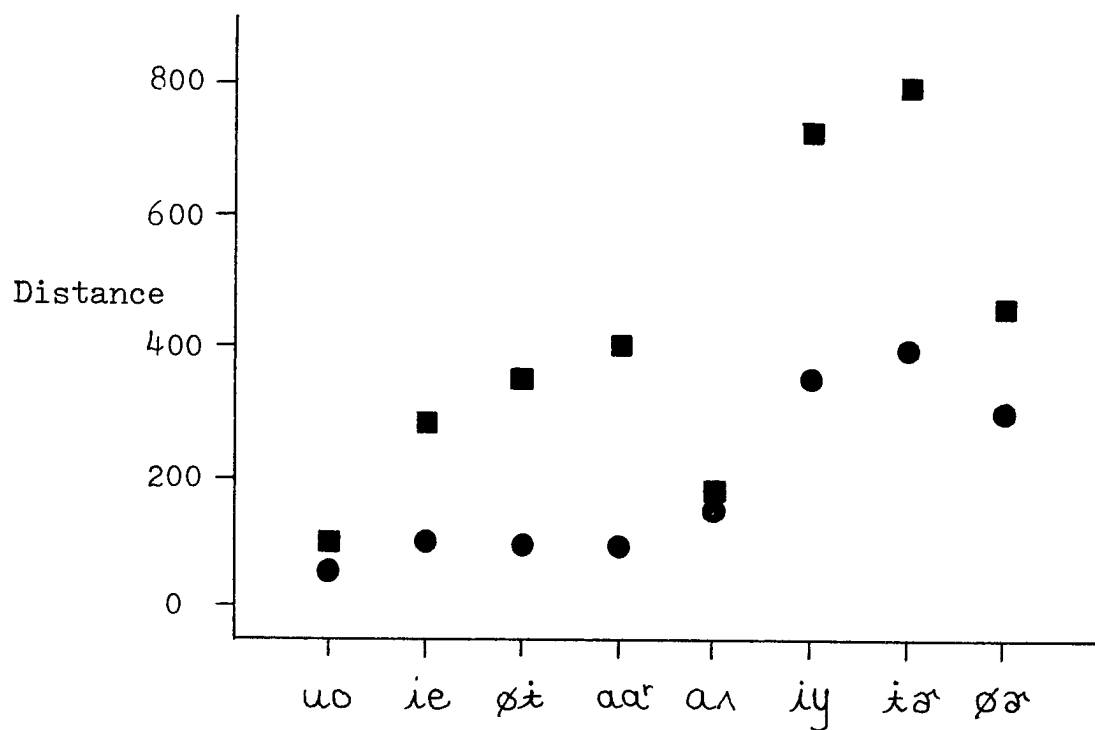


Figure 7.1: Euclidean distances in F1/F2 space (●) and F1/F2/F3 space (■) for 8 selected vowel pairs.

dimensional salience. For convenience, Euclidean distances were computed even in the case of curved dimensions. None of the vowel pairs involved an appreciable distance across the plane of a curved dimension, so this simplification has a negligible effect on the results. Since the personal weights had been normalized previously by making the sums of squared weights equal, the distances are comparable across languages, in spite of any difference in number of dimensions.

Distances for each pair are plotted vertically on Figure 7.2. Means for English, German, Turkish, Swedish and Thai are indicated by E, G, K, S and T, respectively. If the two vowels in a pair are both phonemes in a language, the position is marked with a circle; if only one vowel of a pair or neither is present, the point is marked with a square.

The best support for hypothesis HP comes from the two columns representing pairs which are contrastive in some languages but not in others. The distance between [ɤ] and [ø] (column 3) was greatest by far for Turkish listeners; only Turkish has [ɤ] and [ø] as separate phonemes. Similarly, in column 5, the English and Thai listeners, who have both [ɑ] and [ʌ] phonemes, find the [ʌɑ] distance to be large, relative to the distances computed for German and Turkish listeners, who have only [ɑ]. Swedish listeners did not hear [ʌ].

Furthermore, even though the acoustic differences constituting each pair are not the same, almost all points on Figure 7.2 representing phonemic contrasts have relatively large distances (1.7 or better). The only exception is the Turkish distance between [u] and [o]. This is almost certainly due to the inexact matching of the [uo] stimulus vowels to Turkish phonemes. Turkish /o/ is lower and less tightly rounded than the stimulus [o]; listeners must have heard the two vowels as being much more similar than their own [u] and [o]. This phenomenon may well have lowered other distances, such as [ie] for Turkish and [uo] for Swedish and Thai. But even these values are high, considering the distances shown for non-phonemic pairs with about the same amount of acoustic difference. [øɤ] and [ɑɑ'] have F1/F2 distances similar to the [uo] and [ie] distances (c.f. Figure 7.1), yet the corresponding perceptual distances, when there is no phonemic contrast, are all very low (c.f. Figure 7.2, columns 3 and 4). If the acoustic contribution of F3 is taken into account, the support for hypothesis HP is stronger yet, since the [ɑɑ'] distance in 3-formant space is greater than that for [ie], but the perceptual distances for [ɑɑ'] are small.

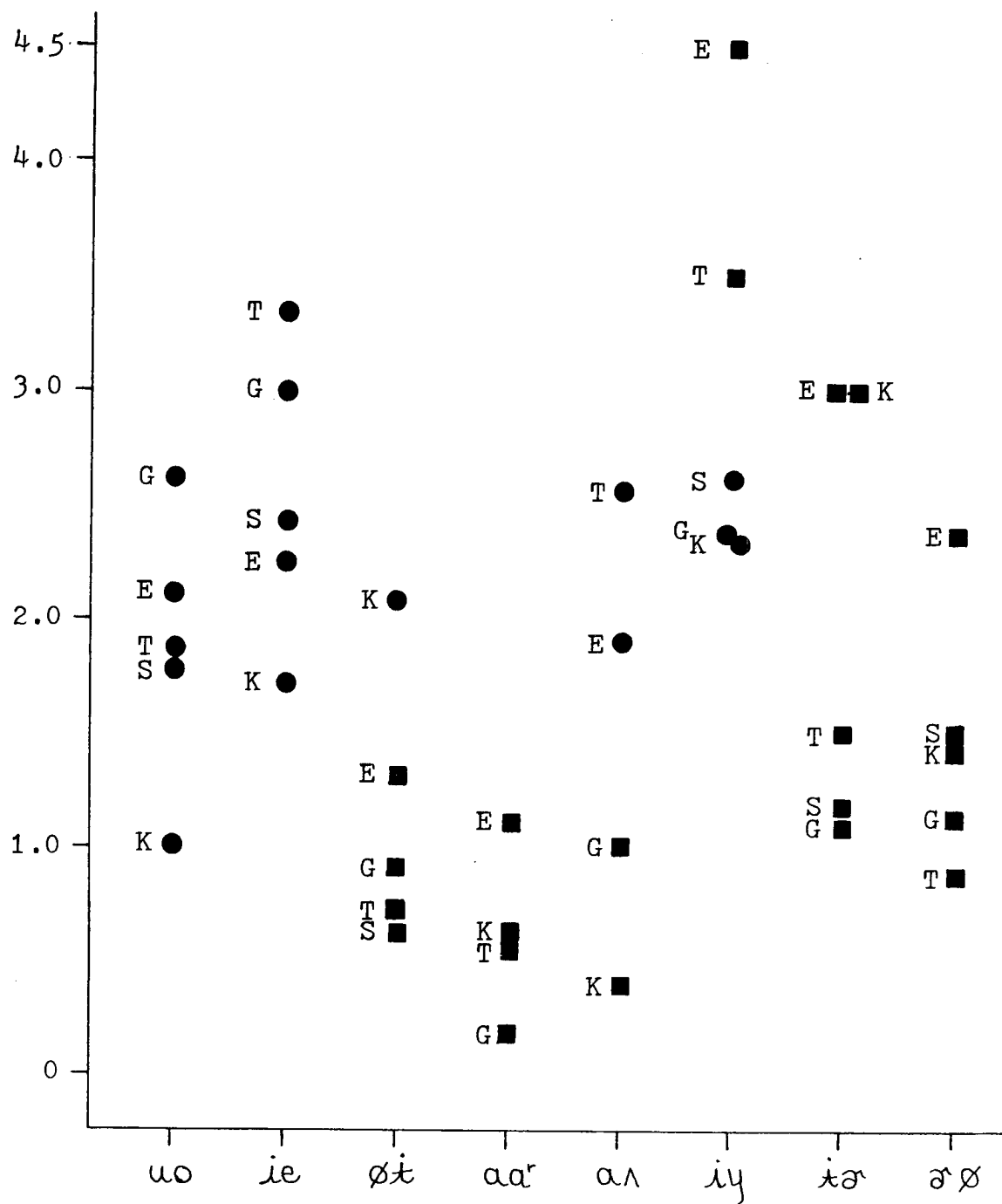


Figure 7.2: Mean perceptual distances for selected close vowel pairs. E=English, G=German, K=Turkish, S=Swedish, T=Thai. ● = both vowels present in the language; ■ = one or both vowels are absent.

Perhaps the most interesting aspect of Figure 7.2 is the presence of five squares with large distances in columns 6, 7, and 8. These are not counterexamples to HP, but they do mean that the presence of a phonemic contrast is not the only phenomenon that can increase a perceptual distance beyond what would be predicted from the acoustic nature of the sounds. To account for these five cases, individual features must be considered, not whole phonemes. In each instance, the vowels which form the pair are assigned to different categories along a feature used in the language.

The English and Thai distances for [ɪ ʏ] are large because those languages use rounding and peripheralness, respectively; the acoustic structure of [ɪ] and [ʏ] cause them to map onto opposite feature values. Admittedly, the extremely large English distance for [ɪ ʏ] is hard to explain, even though rounding is highly salient. Perhaps the listeners, in identifying [ʏ] as a version of [u], judged the distance from a native [ɪ] to a native [u]. Unfortunately, if this suggestion is correct, it must be explained why Turkish listeners did not similarly judge [ɪe] as though it were [ɪɛ]. The resolution of this sort of question awaits further work.

As for [ɛ ɤ], both English and Turkish listeners make use of F3 in identifying their respective native [ɛ] and [ɤ]. With Turkish, F3 is used to distinguish vowels in the [ɛ ɤ] region with respect to backness, and it will distinguish [ɛ ɤ] just as easily. With English, F3 is an independent parameter on which low values are used to identify retroflexion; [ɛ] maps with other non-retroflex vowels. The same reasoning holds for the pair [ɤ ɤ], which is given a large distance by English listeners. The fact that Turkish listeners do not find [ɤ ɤ] to have a relatively large distance rests on the way Turkish listeners use F3: they need only to distinguish high F3 (for [ɛ]) from relatively low F3 (for [ɤ]). The extremely low F3 of [ɤ] is classified as "not high," that is, identically with [ɤ].

It would be of interest to develop a quantitative model for predicting perceptual distances that takes into account the observed stretching of distances across phoneme boundaries and feature values. This would require answering questions such as these: Are some regions of the vowel space more susceptible to perceptual stretching than others? Does the answer depend on the phonetic or phonemic inventory at hand? Such research, it is hoped, would ultimately be integrated with work on articulation, such as the studies by Stevens and by Liljencrants and Lindblom mentioned earlier. The present results provide a framework for this kind of research.

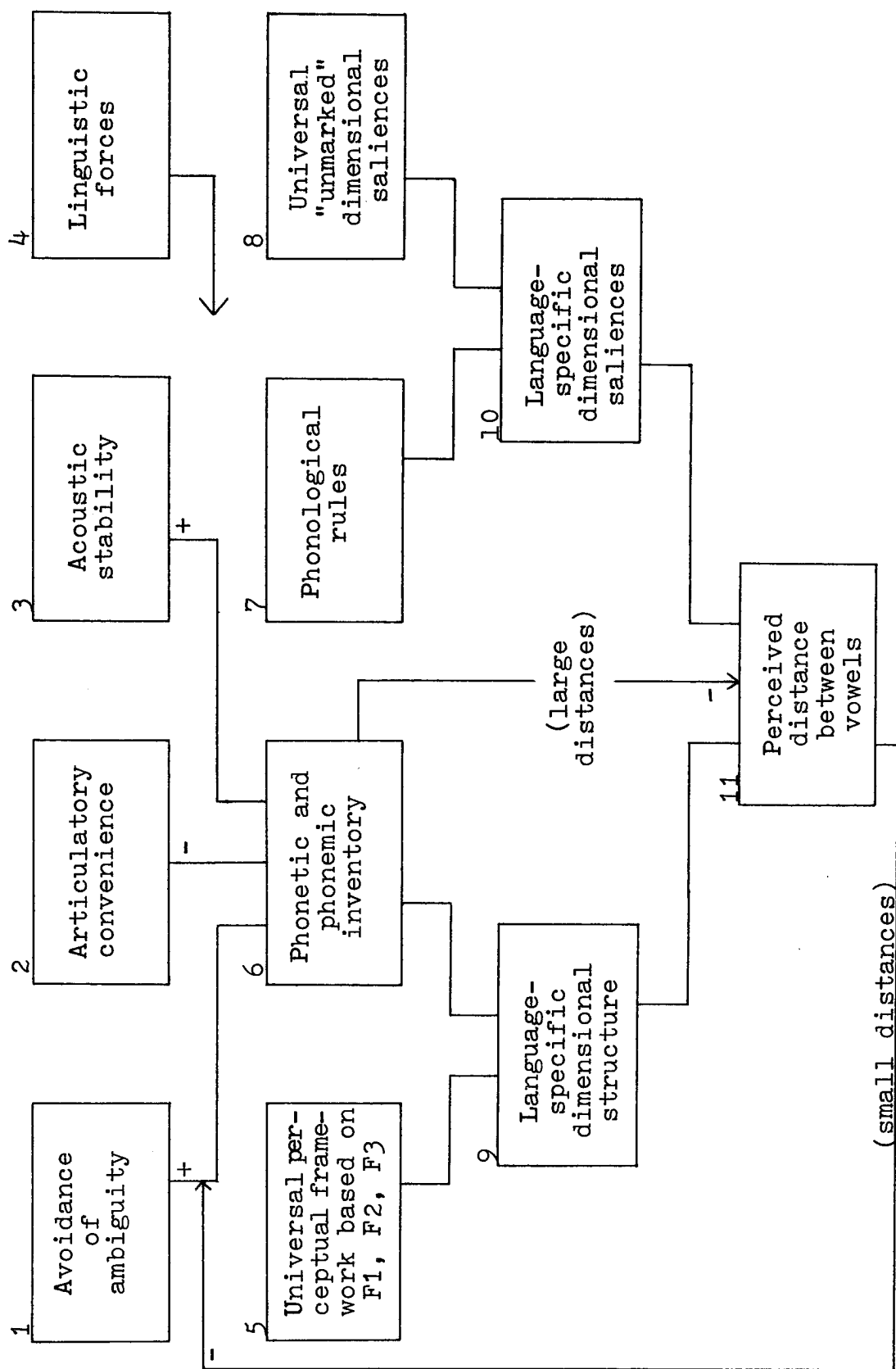


Figure 7.3: Proposed model of the phenomena influencing a listener's judgements of perceived distance between vowels.

7.4 Summary

Based on the results of the present research, the conclusions given below appear justified.

1. Perceptual features for vowels are easily described, for the most part, in traditional linguistic terms.

2. There are language-universal expectations for perceptual dimensions. The three main features expected to appear in a language are rounding, height, and backness. The acoustic cues for these features are generally based on $F1 + aF2$, $F2 - bF1$, and $F1$, respectively. a and b are often such that the range of the first and second formants are scaled to the same size.

3. Phonetic and/or phonemic inventory affect dimensional structure. There are language-specific dimensions; two are retroflexion ($F3$) and peripheralness ($F1$ - $F2$ plane).

4. The most salient dimension tends to be rounding, unless some special circumstance holds (c.f. Section 6.1, 6.2).

5. Language-specific phonological rules can affect dimensional salience. Languages with backness-adjusting rules have significantly greater salience for backness.

6. Perceptual features are not necessarily independent of one another, and sequential processing is implicated.

7. Perceptual features are not restricted to binary values. Vowel height in particular appears as a 3- or 4-valued perceptual dimension.

8. The principle of maximum acoustic separation is opposed by the perceptual expansion of small distances which span phoneme boundaries and the perceptual compression of large distances (curved dimensions).

Figure 7.3 is a graphic model of the forces acting to affect a listener's judgments of perceived vowel distances. Boxes 1-3 represent the pressures acting to change the phoneme inventory itself. The maximum distance principle (box 1) and the acoustic stability principle (box 3) tend to move phonemes apart, hence the plus signs. The phenomenon of articulatory convenience (box 2), tends to allow phonemes to become more similar, hence the minus sign. Box 4 merely recognizes that many linguistic phenomena such as language contact and grammatical change can have their effect on phonetics and phonology. No attempt has been made to portray the undoubtedly complex ways such forces may interact

with other parts of the model. Boxes 5 and 6 represent the universal and local influences on the language-specific dimensional structure (box 9). In parallel fashion, boxes 7 and 8 do the same for dimensional saliences (box 10). Saliency and structure act together to define the actual perceptual space (box 11). The line feeding back to the output of box 1 represents the tendency of the perceptual stretching of small distances to counter an increase in acoustic distance. The remaining line, connecting boxes 6 and 11 directly, attempts to reflect the existence of curved dimensions -- the compression of perceptual distance associated with large acoustic differences.

It should be clear that the conclusions listed above and the model shown in Figure 7.3 are in no way finished products. They represent rather a number of avenues of future research, an interim set of working hypotheses for the continuing study of speech perception from a linguistic viewpoint.

Appendix I

Part A

Raw dissimilarity matrices for each listener.
(E=English; G=German; K=Turkish; S=Swedish; T=Thai)

El:

E2 :

[illegible]

```

12
15  4
 9  1  0
10  6 10  6
 5  9 15 15 14
 4  7  8  6  8  7
14  5  2  5 14 16  9
16  9  4  9 17 11 13  7
 0  8 13 10 14  5  2 12 16
14  9 13  7  5 12 10 14 16 13
19  8 11 10 11 13 15 16 19 17  6

```

E3:

E4 :

```

14
17  5
13  1  2
13  4 10  5
10  5 11 10 16
  1  7 14  6  9 11
15  9  1  8  6 17 15
  7  6 12  7 10 17  8  6
  0  6 16  8 14 10  2 16 10
14  4  9  3 10  7 14 12 13 14
20 11 12 10 13 12 18 15 15 19  5

```

```

12
10  4
10  1  0
  8  4  5  3
10  7 13  8 14
  2  8 12  6 10  6
16  3  3  8 11 16 12
14 10 13  9  9 18 13  6
  0  7 11  7  8  7  1 14 14
12 15 15  6 15  7  9 19 18 12
15  9  9 10 14 15 12 18 20 15 12

```

E5:

E6:

```

16
14  3
14  6  6
12  9 12  4
 4 11 12 14 11
 9 12 13 10 9 11
13 10 10 9 15 15 9
18 12 11 5 14 15 6 1
 0 10 14 16 16 5 6 9 12
14 3 3 7 13 8 10 10 12 12
13 6 0 10 10 15 14 9 15 12 1

```

```

12
12 8
12 4 5
8 7 13 6
9 3 17 9 9
2 10 11 8 9 3
13 13 8 15 12 15 12
12 10 12 12 14 14 10 2
0 12 12 7 7 8 3 12 15
11 8 13 10 11 5 8 13 11 16
13 7 15 9 13 11 11 13 17 13 5

```

E7:

```

16
11 5
14 3 0
10 8 5 4
 3 14 18 11 11
 4 5 11 6 9 8
16 4 7 10 14 18 9
14 9 10 9 12 14 3 0
 1 12 17 15 12 3 0 13 8
 7 18 13 6 13 8 15 19 18 13
16 11 1 4 9 12 16 20 18 15 2

```

E8:

```

13
16 5
14 5 2
16 3 12 5
 5 4 13 10 4
 4 8 11 8 12 2
13 8 0 4 7 12 11
 7 14 10 18 15 8 5 10
 0 11 16 12 13 4 2 15 6
 9 15 14 17 16 8 12 19 11 10
16 7 15 11 11 9 16 15 19 17 0

```

G1:

```

αr
16 ˆ
16 5 y
17 2 3 ø
10 5 8 0 æ
10 9 10 7 9 æ
 2 9 19 15 10 6 ^
11 14 12 14 11 18 8 u
 7 11 11 9 7 16 10 0 o
 0 16 17 15 16 6 1 10 10 a
10 4 8 7 8 0 13 19 18 10 e .
17 6 0 7 11 9 16 20 19 17 3 λ

```

G2:

```

10
18 9
13 1 11
14 2 11 3
 6 5 17 11 11
 2 7 15 10 8 10
20 12 1 7 11 17 16
11 9 16 4 3 12 8 6
 0 12 18 11 13 4 1 16 14
14 0 12 2 6 3 10 17 13 12
16 10 0 13 15 15 15 9 17 18 7

```

G3:

```

15
18 6
13 0 7
16 2 3 1
 9 4 13 10 8
 2 9 16 12 12 8
19 9 5 11 9 13 13
13 7 12 9 8 12 13 3
 1 13 15 14 11 7 0 15 14
10 4 9 4 10 5 9 15 16 8
20 8 7 5 10 16 15 16 16 19 8

```

G4:

```

14
20 10
16 1 5
16 2 8 3
 7 7 13 8 9
 2 15 16 16 14 6
 4 15 5 13 10 17 8
 4 10 18 9 7 14 8 3
 1 16 15 14 10 9 0 6 2
13 4 14 0 7 5 12 17 13 11
16 10 0 9 16 18 17 18 16 14 4

```

G5:

```

13
16 8
14 0 8
13 2 12 2
  2 13 14 15 14
  1 10 17 13 9 4
15 9 4 13 12 16 12
  9 8 13 2 6 18 10 3
  0 11 18 14 15 3 2 13 11
  9 2 12 6 8 4 8 18 15 11
15 12 2 7 12 11 16 19 18 14 4

```

G6:

```

13
16 9
16 2 5
13 0 7 2
  5 11 8 8 8
  2 8 16 14 12 3
12 9 6 11 10 16 13
16 9 14 13 7 14 8 3
  1 10 19 16 10 9 0 15 13
10 4 8 3 10 3 12 16 15 12
18 9 1 9 12 12 17 18 17 18 4

```

G7:

```

8
17 11
14 4 6
11 5 8 0
  6 4 18 16 13
  3 4 16 10 8 5
18 10 9 11 10 18 13
12 10 11 6 10 14 10 8
  0 8 19 16 12 7 2 18 15
16 3 10 6 9 6 10 17 13 11
14 5 11 5 7 10 12 13 12 14 2

```

K1:

```

ar
14 i
16 12 y
15 2 2 ø
15 11 2 0 ø
10 9 12 8 7 æ
  2 11 17 14 13 12 ^
  7 3 8 15 14 16 7 u
  7 6 13 10 12 16 5 0 o
  0 12 16 15 15 11 1 10 5 a
14 9 7 4 6 5 15 18 16 18 e .
18 6 6 7 8 6 19 15 17 18 0 ʌ

```

K2:

```

9
18 6
12 0 3
  9 6 12 7
  9 4 13 7 4
  3 4 17 7 7 9
10 9 10 12 7 18 10
11 10 15 10 12 16 7 4
  0 5 19 12 9 7 1 14 13
13 7 10 9 8 1 15 18 18 14
20 9 4 8 12 12 18 16 19 17 5

```

K3:

```

15
17 4
16 2 4
15 4 5 1
12 5 9 4 6
  0 16 17 14 12 10
  4 14 10 12 14 18 6
  4 12 16 11 11 17 2 2
  2 17 19 16 14 13 1 5 4
17 5 6 5 7 7 18 13 15 15
20 3 3 9 9 9 18 15 15 19 0

```

K4:

```

16
17 4
16 3 3
13 1 2 0
6 10 11 10 8
2 18 18 15 13 7
6 7 10 12 8 18 12
6 13 13 10 8 18 6 0
0 16 18 15 11 4 1 6 4
11 11 6 6 13 2 15 15 19 14
16 8 3 8 12 11 19 20 18 17 1

```

K5:

```

10
12 3
11 1 6
11 2 3 0
10 11 18 13 14
2 10 16 13 12 10
8 12 8 13 12 14 10
4 8 14 11 12 14 4 4
0 13 13 9 9 5 1 6 6
8 17 15 12 16 0 10 14 13 7
16 15 9 13 16 16 13 16 16 12 8

```

K6:

```

13
12 14
14 16 2
19 14 3 8
4 14 13 11 16
4 6 16 8 15 7
7 7 13 10 12 11 9
5 6 7 11 13 11 7 1
0 8 15 12 16 2 1 8 6
15 11 5 8 9 8 12 14 15 16
18 9 7 5 11 14 12 14 16 14 0

```

K7:

```

14
16 3
16 0 2
16 1 5 2
14 8 8 7 9
1 14 16 15 16 12
7 16 11 13 11 17 5
6 11 16 11 9 15 5 0
0 15 18 16 12 14 2 8 7
15 7 5 8 7 1 14 14 14 16
18 6 4 7 10 4 18 15 17 18 2

```

S1:

```

α
2 ɛ
7 8 y
0 3 7 ø
6 4 10 9 æ
19 11 14 16 14 œ
12 4 10 7 1 16 œ
7 6 6 4 9 16 12 u
9 8 9 8 10 14 10 1 o
13 10 12 11 15 9 18 9 10 α
12 10 7 12 15 4 18 12 9 14 e
13 11 9 15 15 10 17 13 11 15 2 λ

```

S2:

```

6
9 13
5 0 7
7 8 15 3
8 10 11 7 7
9 5 11 3 4 12
9 9 15 12 10 11 8
11 11 12 14 9 12 11 2
8 14 16 12 16 12 11 11 9
12 1 13 12 11 11 10 17 11 11
14 12 6 17 13 14 15 12 12 9 2

```

S3:

```

5
1 11
3 1 9
11 3 14 5
16 11 16 10 11
6 3 10 0 1 6
4 12 8 13 14 17 12
10 16 13 11 13 13 11 1
14 14 14 15 18 7 15 10 3
10 8 13 6 18 1 12 14 12 12
7 11 5 10 20 12 16 11 15 15 1

```

S4:

```

11
4 15
7 1 13
6 5 12 4
9 6 13 10 8
10 0 15 2 4 6
5 9 11 11 10 12 9
14 13 13 12 9 15 13 5
13 7 19 9 12 3 12 9 11
13 12 8 10 12 9 10 14 14 13
10 9 6 12 16 15 16 10 14 14 6

```

S5:

```

2
9 12
0 3 9
5 8 12 6
13 8 10 10 9
6 3 14 3 7 7
4 10 7 8 9 15 7
8 14 13 10 11 9 12 8
13 14 9 14 15 7 11 16 11
14 11 5 12 13 0 14 15 12 12
18 17 4 15 18 11 15 19 11 8 5

```

S6:

```

4
4 12
1 2 9
6 3 10 5
15 8 13 10 11
8 0 12 6 2 9
9 11 14 12 14 13 7
9 9 12 9 14 10 8 0
20 18 17 16 18 4 15 7 7
15 8 9 8 14 1 12 19 10 7
15 18 1 13 17 8 14 14 13 14 7

```

S7:

```

2
3 6
1 2 7
10 5 11 8
11 9 11 9 11
6 3 11 1 2 8
9 10 15 11 8 17 6
10 13 16 8 7 3 7 8
16 13 14 14 16 4 13 17 6
8 7 8 5 15 16 13 13 14 17
11 13 7 8 17 15 18 19 16 17 5

```

T1:

```

ar
17 ɤ
12 5 ɣ
9 1 1 ø
10 3 7 2 ʂ
12 8 13 12 11 æ
1 8 14 11 8 12 ʌ
16 11 5 6 9 15 12 u
13 9 6 4 10 14 12 4 o
0 15 13 15 11 15 3 18 10 a
14 1 8 7 4 8 12 16 14 14 e
16 10 5 10 11 15 17 11 17 18 9 i

```

T2:

```

11
13 8
12 3 7
11 3 9 4
8 14 15 14 9
1 8 12 10 11 10
12 8 16 7 9 11 14
11 6 14 12 8 18 12 1
0 12 14 10 11 8 3 13 10
7 0 12 5 5 13 9 9 10 13
16 14 0 14 11 14 15 14 14 16 16

```

T3:

```

11
16 3
12 0 4
15 5 3 2
14 13 11 13 14
11 8 9 8 8 11
13 7 6 10 6 18 11
17 6 5 9 10 10 12 3
0 10 16 11 14 10 9 16 14
16 11 8 12 8 12 11 13 12 12
14 6 8 7 11 13 11 12 10 14 5

```

T4:

```

18
18 7
16 1 11
11 7 12 5
4 11 19 13 8
11 7 15 7 0 7
17 14 3 15 7 14 10
10 13 9 13 11 15 5 5
0 16 17 14 10 3 11 13 12
8 13 9 9 7 9 7 12 4 9
10 16 13 11 12 13 9 6 8 9 1

```

T5:

```

9
19 4
8 0 5
11 1 10 2
7 10 17 10 11
10 8 13 7 4 11
16 9 1 14 13 13 13
11 11 13 5 6 12 3 8
0 12 17 11 6 8 12 11 14
9 12 11 8 13 2 6 16 11 8
15 12 10 13 12 16 20 6 16 16 12

```

T6:

```

14
14 3
10 0 6
8 3 6 1
8 11 11 9 13
6 11 12 7 6 7
15 10 10 15 15 15 11
14 11 14 12 15 14 12 5
0 11 12 9 13 9 2 17 13
10 9 10 6 8 0 11 14 15 12
14 7 9 13 8 14 12 12 11 13 12

```

Part B

ENGLISH

GERMAN

THAI

248

TURKISH

91	46	22	18	64	67	59	11	45	100	e	.
108	46	22	18	64	67	59	11	45	100	e	.
100	24	22	18	64	67	59	11	45	100	e	.
98	39	32	18	64	67	59	11	45	100	e	.
65	61	84	60	64	67	59	11	45	100	e	.
14	79	117	86	88	67	59	11	45	100	e	.
49	68	70	87	78	112	59	11	45	100	e	.
43	66	94	74	77	107	36	11	45	100	e	.
2	86	118	95	86	56	8	57	45	100	e	.
93	67	54	52	66	24	99	106	110	100	e	.
126	56	36	57	78	72	117	111	118	115	16	j

SWEDISH

32	77	61	40	71	64	61	25	57	86	e	.
37	77	61	40	71	64	61	25	57	86	e	.
17	12	61	40	71	64	61	25	57	86	e	.
51	36	84	40	71	64	61	25	57	86	e	.
91	63	88	72	71	64	61	25	57	86	e	.
57	18	83	22	21	64	61	25	57	86	e	.
47	67	76	71	74	101	61	25	57	86	e	.
71	84	88	72	73	76	72	25	57	86	e	.
97	90	101	91	110	46	95	79	57	86	e	.
84	57	63	65	98	42	89	104	82	86	e	.
88	91	38	90	116	85	111	98	92	92	28	j

Appendix I.

Part C

Summed responses of all 35 listeners to 10 vowels.

i	u	ø	ɔ	æ	ʊ	o	a	e	ɐ
211	149	98	276	422	92	256	327	141	289
61	247	294	299	368	319	374	429	141	289
131	381	265	286	153	422	374	429	141	289
243	218	297	299	422	92	256	327	141	289
271	218	297	299	422	92	256	327	141	289
275	347	265	286	368	92	256	327	141	289
310	437	335	335	153	319	256	327	141	289
231	287	212	295	153	422	374	327	141	289
289	188	297	368	350	415	446	429	141	289

Appendix II

Co-ordinates for Selected Solutions

The solutions presented in the following tables include both vowel and subject co-ordinates. Vowel co-ordinates are normalized vertically. The values for any single dimension have a mean of zero, and are scaled such that the sum of their squares equals 100. In cases where two dimensions have been treated as components of a single curved scale, distances along those curves are given as well.

Personal weighting factors have been normalized horizontally, such that for any subject, the sum of their squares equals 1. Columns labelled $\sum \omega^2$ are the sums of squares of the unnormalized weights. To plot the perceptual space in n dimensions for subject S , whose weighting factors are $S_1, S_2 \dots S_n$, each vowel V , whose basic co-ordinates are $V_1, V_2 \dots V_n$, will be given new, personal co-ordinates $V_1 S_1, V_2 S_2 \dots V_n S_n$. Theoretically unacceptable weighting factors (negative values in the original computer output) are indicated by $\underline{i}$, the square root of -1. Small values can be treated as zero, effectively deleting the dimension involved. Large values indicate a more serious breakdown of the model.

The PARAFAC2 solution for Swedish has been normalized exactly as above. A negative personal weight changes the angle θ between that dimension and any other one to $180^\circ - \theta$. Negative personal weights can be effectively ignored if the dimension involved lies at nearly right angles to the others, as is the case with Round/Nonround. Personal vowel co-ordinates are computed as above. But in computing vowel distances or plotting, the angle between height and backness, especially, cannot be ignored.

The last table records the goodness of fit for nine series of analyses. "Fit" is

$$1 - \frac{\sum_{i=1}^m (D_i - P_i)^2}{\sum_{i=1}^m D_i^2} .$$

D is a value in the scalar products input matrices; P is the solution's prediction of that value. The summation is taken over all m input values and predictions. $m = sp(p-1)/2$, where s = number of subjects, and p = number of stimuli.

Table A II-1: PARAFAC analysis in three dimensions,
12 Swedish listeners (data from Hanson 1967)

Vowel space

	Height-Rounding plane		Back/Nonback
ʉ	3.58	5.95	-0.30
ø	4.50	-0.68	-0.03
y	1.67	0.87	-3.77
u	-1.55	4.10	4.36
o	0.49	1.07	4.33
a	2.40	-4.99	3.63
æ	-2.82	-3.84	-0.24
e	-1.52	-2.39	-2.78
i	-6.75	-0.09	-5.20

Subject space

	Height-Rounding plane		Back/Nonback	$\sum w^2$
* 1	.491	.493	.718	37.36
2	.403	.469	.786	10.79
7	.500	.576	.647	7.74
9	.233	.469	.852	6.09
10	.488	.514	.706	11.96
11	.491	.462	.739	18.81
12	.276	.463	.843	9.40
13	.549	.667	.504	32.67
14	.192	.472	.860	29.38
16	.217 i	.580	.843	12.30
17	.570	.404	.715	5.09
19	.263	.495	.828	3.72

* Hanson's original numbering. The 8 least consistent listeners were excluded here.

Table A II-2: PARAFAC analysis in four dimensions, 15 Dutch listeners (data from Pols et al, 1969)

Vowel space

	Background/ Front- nonround	High/ Low	e/a	Highlow/ Mid	Derived Height Curve	Derived Round- back Curve
œ	1.35	1.22	-1.24	4.21	4.60	5.70
o	-3.97	-0.76	-1.04	2.70	3.44	1.92
a	0.63	-3.61	-4.88	-1.10	1.00	4.50
u	-5.43	2.88	2.64	-2.40	7.98	0.00
y	-1.32	5.23	-2.87	-3.53	8.92	3.65
i	4.93	3.97	2.48	-3.77	8.84	7.95
ɔ	1.53	-4.67	2.68	-3.20	0.00	7.72
ø	2.50	1.02	1.15	3.42	4.68	7.05
ɔ	-2.85	-1.88	-2.48	2.29	2.88	2.87
ɛ	-0.47	-2.54	-2.17	-1.89	0.94	4.15
e	3.10	-0.86	5.73	3.27	3.86	9.35

Subject space

	Background/ Front- nonround	High/ Low	e/a	Highlow/ Mid	$\sum w^2$
* 3	.608	.640	.235	.405	9.17
7	.681	.671	.307	.086 i	9.80
8	.482	.584	.241	.607	9.71
9	.235	.633	.785	.268 i	8.33
10	.662	.567	.305	.383	10.55
11	.423	.701	.214	.533	8.20
12	.593	.574	.309	.472	8.82
13	.469	.632	.197	.584	8.87
14	.557	.545	.477	.407	4.48
15	.561	.585	.176	.559	11.50
16	.457	.602	.192	.626	5.37
17	.510	.562	.387	.524	6.46
18	.601	.592	.326	.426	5.26
19	.308	.618	.385	.613	3.83
20	.468	.584	.319	.581	7.00

* Pols et al original numbering. 5 subjects were not native speakers of Dutch and were excluded.

Table A II-3: Vowel space for PARAFAC analysis in six dimensions, 35 listeners

Vowels	Back/		Back/		Round/		Low/		Peripheral/		High/	
	Nonback1	Nonback2	Nonback2	Nonround	Nonround	Nonround	Nonlow	Nonlow	Central	Nonhigh	Nonhigh	Nonhigh
i	-3.05	4.29	4.29	0.23	0.23	0.23	1.95	1.95	-3.71	-2.70	-2.70	-2.70
y	-3.12	-3.11	-3.11	3.66	3.66	3.66	2.78	2.78	-0.88	5.20	5.20	5.20
ø	-1.81	-2.30	-2.30	1.05	1.05	1.05	2.67	2.67	-2.66	-3.05	-3.05	-3.05
ɘ	0.24	-4.68	-4.68	1.12	1.12	1.12	0.87	0.87	-4.96	-3.14	-3.14	-3.14
æ	-3.30	-1.03	-1.03	-2.25	-2.25	-2.25	-5.73	-5.73	1.71	-1.25	-1.25	-1.25
u	3.19	3.91	3.91	5.38	5.38	5.38	1.08	1.08	0.03	3.27	3.27	3.27
o	4.27	2.81	2.81	4.43	4.43	4.43	0.51	0.51	3.29	-2.28	-2.28	-2.28
a	4.11	2.49	2.49	-3.21	-3.21	-3.21	-4.42	-4.42	1.55	1.41	1.41	1.41
e	-3.23	-2.97	-2.97	-3.30	-3.30	-3.30	1.56	1.56	5.16	-2.11	-2.11	-2.11
ɪ	-3.33	-2.40	-2.40	-3.06	-3.06	-3.06	4.93	4.93	3.32	4.75	4.75	4.75

Curves derived from original dimensions (above)

Vowels	Distance along curve			Deviation from curve		
	Back	Height	Round	Back	Height	Round
i	1.65	2.13	1.99	0.0	-0.03	0.0
y	3.31	4.14	3.16	-0.18	0.0	0.15
ø	3.39	2.17	2.05	-0.19	-0.22	0.33
ɘ	4.03	1.88	2.22	0.0	0.0	-0.24
æ	2.95	0.21	0.84	0.12	-0.32	0.09
u	0.32	3.68	3.45	-0.06	0.47	-0.20
o	0.02	1.82	4.06	0.03	0.24	0.16
a	0.0	0.0	0.71	0.11	0.39	-0.12
e	3.30	2.17	0.0	-0.16	0.16	0.05
ɪ	3.18	3.98	0.38	-0.10	-0.54	0.04

Table A II-4: Subject space for PARAFAC analysis in six dimensions, 35 listeners

Subject	Back/ Non- back1	Back/ Non- back2	Round/ Non- round	Low/ Nonlow	Peri- pheral/ Central	High/ Non- high	Sum of Squared Un-nor- malized weights
<u>English</u>							
E1	.230	.248	.729	.415	.342	.246	14.93
E2	.192	.407	.578	.546	.302	.272	14.06
E3	.496	.225	.528	.474	.320	.311	13.64
E4	.385	.316	.592	.408	.350	.337	15.88
E5	.432	.349	.381	.540	.348	.366	15.04
E6	.468	.192	.544	.405	.364	.389	12.74
E7	.360	.453	.508	.479	.366	.211	25.60
E8	.405	.212	.525	.445	.551	.119	19.83
<u>German</u>							
G1	.564	.311	.488	.412	.255	.334	26.80
G2	.397	.068	.491	.507	.178	.555	21.44
G3	.426	.278	.523	.540	.229	.353	16.62
G4	.542	.284	.366	.436	.306	.482	31.58
G5	.409	.304	.525	.459	.262	.430	29.35
G6	.486	.327	.482	.453	.289	.368	21.66
G7	.321	.307	.611	.535	.265	.269	12.53
<u>Thai</u>							
T1	.505	.134	.502	.502	.200	.427	13.98
T2	.432	.359	.315	.554	.234	.472	12.62
T3	.261	.342	.559	.550	.443	.062	10.45
T4	.276	.074	.483	.510	.497	.423	16.56
T5	.348	.140	.450	.486	.417	.495	18.38
T6	.379	.379	.466	.438	.404	.374	11.63
<u>Turkish</u>							
K1	.499	.536	.403	.389	.237	.308	25.14
K2	.523	.347	.511	.428	.188	.355	18.11
K3	.678	.329	.264	.410	.302	.320	19.21
K4	.577	.321	.423	.451	.372	.209	27.62
K5	.387	.285	.420	.417	.538	.359	17.81
K6	.338	.618	.403	.536	.259	.111 i	16.07
K7	.697	.281	.327	.350	.335	.305	16.04
<u>Swedish</u>							
S1	.417	.252	.439	.407	.553	.314	12.46
S2	.462	.329	.392	.317	.443	.477	14.43
S3	.476	.190	.365	.385	.505	.448	24.27
S4	.410	.331	.379	.456	.412	.449	15.27
S5	.373	.271	.459	.287	.534	.458	14.85
S6	.466	.379	.333	.242	.505	.463	22.86
S7	.491	.240	.321	.531	.440	.350	15.25

Table A II-5: PARAFAC analysis in four dimensions,
7 German listeners

Vowel space

	Low/ Nonlow	Round/ Nonround	High/ Nonhigh	Back/ Nonback	Derived Height Curve
a ^r	-4.97	2.20	-0.14	-4.04	0.00
ɪ	0.11	-0.67	2.51	2.27	2.88
y	3.20	-1.83	-5.58	3.24	7.74
ø	1.39	-1.60	2.29	3.01	3.50
ɘ	0.81	-2.12	3.32	1.37	3.14
æ	-1.73	3.96	1.70	-0.38	1.88
ʌ	-2.93	1.54	0.65	-3.59	1.06
u	3.26	-4.98	-3.65	-2.13	6.76
o	0.17	-4.41	1.98	-2.13	2.90
a	-4.82	2.29	-0.36	-3.82	0.00
e	1.83	3.48	1.95	2.31	3.76
i	3.69	2.14	-4.67	3.88	7.28

Subject space

	Low/ Nonlow	Round/ Nonround	High/ Nonhigh	Back/ Nonback	$\sum \omega^2$
1	.341 i	.615	.350	.784	17.86
2	.547	.524	.529	.382	16.37
3	.507	.573	.294	.573	11.49
4	.333 i	.504	.511	.772	18.37
5	.199	.635	.434	.608	18.85
6	.341	.576	.361	.650	14.99
7	.595	.632	.238	.437	9.63

Table A II-6: PARAFAC analysis in four dimensions,
7 Turkish listeners

Vowel space

	High/ Nonhigh	Back/ Nonback1	Back/ Nonback2	Round/ Nonround	Derived Backness Curve
a ^r	-1.95	3.56	3.03	-2.30	3.08
ɪ	0.53	-2.49	3.25	3.04	1.81
y	4.20	-2.56	-3.13	1.91	0.36
ø	0.01	-2.20	-2.71	2.58	0.40
ɘ	-1.76	-1.18	-4.42	4.05	0.00
æ	-4.77	-2.07	0.73	-3.38	1.15
ʌ	-3.08	3.40	1.87	-1.55	3.14
u	4.03	3.06	2.43	2.30	3.03
o	1.90	3.38	2.14	2.64	3.12
a	-3.13	3.39	3.05	-1.67	3.03
e	0.39	-2.69	-3.53	-4.53	0.30
ɪ	3.63	-3.61	-2.70	-3.09	0.51

Subject space

	High/ Nonhigh	Back/ Nonback1	Back/ Nonback2	Round/ Nonround	$\sum \omega^2$
1	.424	.637	.500	.406	16.98
2	.574	.654	.308	.385	14.24
3	.437	.824	.220	.285	14.15
4	.437	.695	.217	.528	21.40
5	.450	.547	.285	.646	10.89
6	.404	.413	.688	.441	12.38
7	.329	.825	.199	.414	12.45

Table A II-7: PARAFAC analysis in five dimensions,
6 English listeners

Vowel space

	Back/ Nonback	Round/ Nonround	Retro- flex/ Nonret.	Mid/ Nonmid	High/ Nonhigh	Derived Height Curve
q ^r	3.07	1.12	2.72	4.28	4.28	0.00
t	-2.77	-1.05	0.28	-1.26	-0.94	6.90
y	-2.89	-4.32	-0.55	0.28	-2.36	7.95
ø	-1.08	-0.84	0.74	-4.61	-1.14	5.57
ɶ	2.34	-1.44	5.60	-4.02	-0.89	5.65
æ	-5.45	2.78	0.50	1.80	2.90	1.37
ʌ	2.60	1.93	-0.20	-0.96	3.17	2.77
u	2.11	-4.51	-4.53	0.87	-1.58	8.07
o	4.15	-3.10	-6.06	-3.33	0.22	5.05
a	1.47	1.42	0.36	4.21	4.18	0.05
e	-2.61	4.59	-0.44	-0.25	-2.50	7.72
i	-0.95	3.43	1.58	2.97	-5.33	9.67

Subject space

	Back/ Nonback	Round/ Nonround	Retro- flex/ Nonret.	Mid/ Nonmid	High/ Nonhigh	$\sum \omega^2$
1	.185	.714	.282	.367	.492	14.24
2	.057i	.518	.487	.303	.637	13.41
3	.500	.466	.228	.261	.642	15.19
4	.402	.612	.364	.281	.501	14.14
5	.388	.217	.481	.503	.564	15.17
6	.390	.515	.487	.285	.515	11.43

Table A II-8: PARAFAC analysis in three dimensions,
6 Thai listeners

Vowel space

	Peripheral/Central	Back/Nonback	Highround/Lownonround
ɑ ^r	-3.81	3.02	3.72
ɤ	4.96	-1.33	-0.61
ɥ	2.59	-2.71	-3.52
ø	4.18	-0.62	0.12
ə	2.34	-0.51	0.30
æ	-1.61	-0.97	3.98
ʌ	1.31	3.83	1.53
u	-1.58	0.73	-5.15
o	-1.13	3.27	-3.26
a	-3.34	3.76	3.22
e	-0.50	-2.79	1.78
i	-3.42	-5.69	-2.11

Subject space

	Peripheral/ Central	Back/ Nonback	Highround/ Lownonround	$\sum w^2$
1	.358	.684	.636	7.34
2	.487	.620	.615	5.06
3	.453	.509	.732	4.15
4	.688	.265	.675	7.75
5	.489	.541	.684	7.37
6	.469	.523	.711	5.37

Table A II-9: PARAFAC analysis in six dimensions,
6 Thai listeners

Vowel space

	High/ Nonhigh	Highround/ Lownonround	Back/ Nonback	Low/ Nonlow	Peripheral/ Central	??
a ^r	0.70	3.24	-4.51	-2.86	3.42	0.68
i	-0.17	0.30	1.89	1.27	-5.23	-2.44
y	-3.86	-1.95	1.32	3.94	-2.34	4.40
ø	1.49	1.53	0.41	2.86	-3.53	-1.71
ɔ ^r	1.18	0.98	0.63	0.96	-2.70	-1.30
æ	0.46	1.84	3.37	-6.40	0.55	2.21
ʌ	3.77	0.34	-3.29	-1.43	-1.40	1.89
u	-3.48	-6.34	-0.12	-0.18	0.16	-3.91
o	4.13	-4.94	-1.41	1.02	1.77	-2.70
ɑ	-1.09	2.87	-5.12	-3.30	2.46	1.48
e	2.40	2.45	3.73	1.03	2.83	-3.52
ɪ	-5.52	-0.33	3.09	3.09	4.03	4.90

Subject space

	H/NH	HR/LNR	B/NB	L/NL	P/C	??	$\sum \omega^2$
1	.327	.476	.581	.430	.297	.237	18.97
2	.267	.372	.422	.510	.284	.521	18.18
3	.161 i	.536	.490	.506	.417	.263	12.51
4	.349	.506	.263	.499	.589	.205 i	19.85
5	.558	.498	.363	.450	.450	.312 i	21.43
6	.278	.568	.469	.402	.399	.244	14.46

Table A II-10: PARAFAC2 analysis in three dimensions,
7 Swedish listeners

Vowel space

	High/Low	Back/Front	Round/Nonround
ʌ	3.01	-1.14	3.12
ɪ	-1.02	-2.52	1.70
ʏ	4.71	-1.89	-0.55
ø	-0.08	-2.38	1.66
ɘ	-2.71	-1.56	3.37
æ	-4.29	-0.21	-3.33
œ	-2.67	-0.91	2.54
u	1.58	4.27	2.82
o	-1.18	4.95	0.40
a	-2.94	5.37	-2.92
e	0.95	-2.30	-4.31
ɪ	4.64	-1.69	-4.50

Subject space

	High/Low	Back/Front	Round/Nonround	$\sum w^2$
1	.497	.452	.741	6.72
2	.487	.622	.613	5.16
3	.586	.593	-.552	14.72
4	.689	.527	.498	6.33
5	.478	.424	.769	7.54
6	.460	.611	-.644	12.06
7	.584	-.554	-.593	7.78

Angles between dimensions

Height-Backness: 71°
 Height-Rounding: 92°
 Backness-Rounding: 94°

Table A II-11: PARAFAC analysis in three dimensions,
6 Swedish listeners

Vowel space

	Back/Front	High/Low	Round/Nonround
ɤ	1.28	2.99	-3.05
ɛ	2.67	-1.85	-1.68
ɜ	1.64	5.00	0.76
ø	2.51	-0.70	-1.74
ɔ	2.55	-2.91	-3.26
æ	0.55	-4.28	3.32
œ	1.45	-2.76	-2.60
u	-4.28	2.29	-2.75
o	-4.62	0.15	-0.60
a	-5.63	-1.96	2.41
e	1.29	-0.28	4.55
ɪ	0.59	4.31	4.63

Subject space

	Back/Front	High/Low	Round/Nonround	$\sum w^2$
1	.451	.472	.757	6.52
2	.611	.483	.627	4.95
3	.586	.579	.567	13.78
4	.514	.680	.523	6.01
5	.402	.472	.785	7.26
6	.610	.462	.643	11.88

Table A II-12: Fit values for 9 series of solutions

<u>Language</u>	<u>Number of listeners</u>	<u>Number of dimensions</u>						
		1	2	3	4	5	6	7
All five	35	.359	.630	.707	.752	.784	.814	.839
German	7	.440	.780	.850	.884	.905		
Turkish	7	.662	.799	.836	.877	.901		
Thai	6	.384	.541	.651	.714	.768	.813	
English	6	.443	.707	.749	.796	.832	.863	
Swedish	6	.364	.578	.775	.821			
Swedish (PARAFAC2)	7		.568	.757	.820			
Swedish (Hanson)	12	.533	.689	.765	.820			
Dutch (Pols et al)	15	.233	.541	.648	.706	.744		

Appendix III

A Note on the I/O Ratio as an Index of Solution Stability

A serious consideration in MD-scaling is the selection of appropriate numbers of subjects and stimuli. With noisy (i.e., real) data, the fewer listeners and stimuli used, the more unreliable the solutions, especially those with many dimensions. Table A III affords an estimate of solution stability according to number of dimensions extracted, number of stimuli, and number of persons. The measure is the number of active cells submitted to the program, divided by the total number of co-ordinates the program is expected to derive (input/output=I/O). The number of co-ordinates for each dimension extracted is counted as the number of vowel stimuli plus the number of listeners.

The larger this ratio, the less sensitive the output co-ordinates are to noise in the distance values. If the ratio falls below 1.0, the solution is highly unstable, and should in all likelihood not have been attempted at all, since there are more output values than input values.

The I/O ratios for the 35 person solutions are found in column a. The ratios for individual language solutions are given in columns b, c, d. Boxes surround values representing solutions discussed in the text. The worst two of these values are 5.3 and 5.8. In both cases, the last dimension extracted is marginally interpretable, and the next smallest space is an adequate solution, as well. All boxed values are in the same range, though. By marked contrast, for any two-mode scaling of twelve vowels (column e), such as COSCAL on each person's matrix, no more than a 1-D solution can approach the dependability of the PARAFAC solutions presented.

Needless to say, the absolute value of I/O ratios is difficult to evaluate. The less reliable the data, the more menacing looms the I/O ratio. But for highly reliable data, an I/O approaching 1.0 will have little adverse effect on the results of a scaling solution. This is the standard expectation for pooled data in a two-mode analysis. Is the noise level of the present data sufficiently low to justify interpreting solutions with I/O ratios on the order of 5, 6, or 7? There are no guidelines in the state of the art of MD-scaling for weighing the reliability of data against the warning implied by the stability measure I/O. One can only approach the question by changing the point of view: If a solution is interpretable, it is likely that the I/O ratio was adequate. Since the solutions presented in the text are

reasonably interpretable, it can be declared post hoc that the I/O ratios were acceptable. Column f refers to data from Pols et al (1969) for comparison. The I/O ratios and accompanying interpretability are similar to those of the present experiment.

It is worth commenting that large I/O ratios do not necessarily mean easily interpretable results. Column g in the table represents the data collected by Hanson (1967). Although the I/O ratio for the "best" solution is 12.0, that 3-D space is still difficult to interpret. In planning MD-scaling experiments, it seems that an I/O ratio of about 6 is adequate, if a response task is used that is not intrinsically noisy. Hanson's free numerical estimation is not recommended.

Table A III: I/O Ratios for various solutions. Column e refers to COSCAL (or any two-mode procedure), the remainder to PARAFAC.

	a	b	c	d	e	f	g
	All 35 subjects	Thai, English	German, Turkish, Swedish	English	Single matrices	Dutch	Swedish
Number of stimuli	10	12	12	12	12	11	9
Number of subjects	35	6	7	8	1	15	20
Number of co-ordinates/dimension	45	18	19	20	12	26	29
Number of active cells	1575	396	462	528	66	825	1044
Number of dimensions							
1	35.0	22.0	24.2	26.4	5.5	31.8	36.0
2	17.5	11.0	12.1	13.2	2.8	15.9	18.0
3	11.7	<u>7.3</u>	<u>8.1</u>	8.8	1.8	10.5	<u>12.0</u>
4	8.8	5.5	<u>6.1</u>	6.6	1.4	<u>7.9</u>	9.0
5	7.0	4.4	4.9	<u>5.3</u>	1.1	6.3	7.2
6	<u>5.8</u>	3.7	4.1	4.4	.90	5.3	6.0
7	5.0	3.1	3.5	3.8	.86	4.5	5.1

c.f. Section # 5.1-4 6.3,4 6.1,2,5 6.4 4.4 2.3,5 2.3,5

BIBLIOGRAPHY

- Abercrombie, David (1967) Elements of General Phonetics, Chicago: Aldine.
- Abramson, Arthur (1962) "The vowels and tones of Standard Thai," Int. J. of Amer. Ling. 28.2, part 3.
- Attneave, F. (1950) "Dimensions of similarity," Amer. J. Psychol. 63:516-556.
- Bogert, B.P. (1973) "Practicing digital spectrum analysis," in David, E. Jr. and P. Denes, eds., Human Communication: A Unified View, New York: McGraw-Hill, pp. 225-282.
- Bricker, Peter (1971) "Individual differences and multidimensional scaling of speech perception data," J. Acoust. Soc. Amer. 49:96.
- Carroll, J.D. and J.J. Chang (1970) "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckhart-Young' decomposition," Psychometrika 35:283-319.
- Chomsky, Noam and Morris Halle (1968) The Sound Pattern of English, New York: Harper and Row.
- Cliff, N. (1966) "Orthogonal rotation to congruence," Psychometrika 31:33-42.
- Cooper, Lee (1970) "Metric multidimensional scaling and the concept of preference," Western Management Science Institute paper no. 163, Univ. of Calif. at Los Angeles.
- Essner, C. (1947) "Recherches sur la structure des voyelles orales," Neerland. Phon. Exper. 20:40-77.
- Fant, Gunnar (1956) "On the predictability of formant levels and spectrum envelopes from formant frequencies," in Halle, Morris, ed., For Roman Jakobson, The Hague: Mouton, pp. 109-120.
- _____ (1959) "Acoustic analysis and synthesis of speech with applications to Swedish," Ericsson Technics 1:106 pp.
- _____ (1968) "Analysis and synthesis of speech processes," in Malmberg, Bertil, ed., Manual of Phonetics, Amsterdam: North Holland, pp. 173-277.
- _____ (1970) Acoustic Theory of Speech Production (second edition), The Hague: Mouton.

- Fischer-Jorgenson, Eli (1967) "Perceptual dimensions of vowels," in To Honor Roman Jakobson, The Hague: Mouton, pp. 667-671.
- Flanagan, James (1957) "Difference limen for formant amplitude," J. Speech and Hear. Dis. 22:205-212.
- _____ (1972) Speech Analysis Synthesis and Perception, Berlin: Springer-Verlag.
- Flanagan, James, C. Coker, L. Rabiner, R. Schafer and N. Umeda (1970) "Synthetic voices for computers," IEEE Spectrum 7:22-45.
- Greenberg, Joseph and J. Jenkins (1964) "Studies in the psychological correlates of the sound system of American English," Word 20:157-177.
- Gulick, W.L. (1971) Hearing, London: Oxford.
- Halle, Morris (1959) The Sound Pattern of Russian, The Hague: Mouton.
- Hanson, Gøte (1967) "Dimensions in speech sound perception: an experimental study of vowel perception," Ericsson Technics 23: 175 pp.
- Harmon, H. (1960) Modern Factor Analysis, Chicago: Univ. of Chicago Press.
- Harshman, Richard (1970) "Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis," Working Papers in Phonetics 16: 84 pp.
- _____ (1972a) "Determination and proof of minimum uniqueness conditions for PARAFAC1," Working Papers in Phonetics 22:111-117.
- _____ (1972b) "PARAFAC2: Mathematical and technical notes," Working Papers in Phonetics 22:30-44.
- Hays, William L. (1963) Statistics, New York: Holt, Rinehart, Winston.
- Hockett, Charles F. (1955) "A manual of phonology," Int. J. of Amer. Ling. 21.4: 247 pp.
- Jakobson, Roman (1941) Kindersprache, Aphasie und allgemeine Lautgesetze, trans. A. Keiler (1968) as Child Language Aphasia and Phonological Universals, The Hague: Mouton.

- Jakobson, Roman, Gunnar Fant and Morris Halle (1952) Preliminaries to Speech Analysis, Cambridge: MIT.
- Jennrich, Robert (1972) "A generalization of the multi-dimensional scaling model of Carroll and Chang," Working Papers in Phonetics 22:45-47.
- Johnson, S. (1967) "Hierarchical clustering schemes," Psychometrika 32:241-254.
- Jones, Daniel (1967) An Outline of English Phonetics (first published 1918), Cambridge: Heffer and Sons.
- Joos, Martin (1948) "Acoustic phonetics," Language Monographs vol. 24.
- Jørgensen, H.P. (1969) "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen," Phonetica 19:217-245.
- Kaplan, Harold (1972) Anatomy and Physiology of Speech (second edition), New York: McGraw-Hill.
- Kruskal, J.B. (1964a) "Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis," Psychometrika 29:1-27.
- _____ (1964b) "Nonmetric multidimensional scaling: a numerical method," Psychometrika 29:115-129.
- Ladefoged, Peter (1962) Elements of Acoustic Phonetics, Univ. of Chicago.
- _____ (1967) Three Areas of Experimental Phonetics, London: Oxford.
- _____ (1971) Preliminaries to Linguistic Phonetics, Univ. of Chicago.
- _____ (1975) A Course in Phonetics, New York: Harcourt Brace Jovanovich.
- Ladefoged, Peter and D.E. Broadbent (1957) "Information conveyed by vowels," J. Acoust. Soc. Amer. 29:98-104.
- Ladefoged, Peter, Joseph deClerk, Mona Lindau and George Papçun (1972) "An auditory-motor theory of speech production," Working Papers in Phonetics 22:48-75.
- Lehiste, Ilse and George Peterson (1959) "Vowel amplitude and phonemic stress in American English," J. Acoust. Soc. Amer. 31:428-435.

- Liljencrants, Johan and Bjorn Lindblom (1972) "Numerical simulation of vowel quality systems: the role of perceptual contrast," Language 48:839-862.
- Lindblom, Bjorn and J. Sundberg (1971) "Acoustic consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Amer. 50:1166-1179.
- Miller, George and Patricia Nicely (1955) "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Amer. 27:338-352.
- Mohr, Burckhardt and William Wang (1968) "Perceptual distances and the specification of phonological features," Phonetica 18:31-45.
- Osgood, Charles, George Suci and Percy Tannenbaum (1957) The Measurement of Meaning, Urbana: Univ. of Illinois.
- Peterson, Gordon and Harold Barney (1952) "Control methods used in a study of the vowels," J. Acoust. Soc. Amer. 24:175-184.
- Peterson, Gordon and Ilse Lehiste (1960) "Duration of syllable nuclei in English," J. Acoust. Soc. Amer. 32:693-703.
- Pike, Kenneth (1943) Phonetics, Ann Arbor: Univ. of Michigan.
- Pols, L.C.W., L.J.Th. van der Kamp and R. Plomp (1969) "Perceptual and physical space of vowel sounds," J. Acoust. Soc. Amer. 46:458-467.
- Schoenemann, P. and R. M. Carroll (1970) "Fitting one matrix to another under a choice of a central dilation and a rigid motion," Psychometrika 35:245-255.
- Shepard, Roger N. (1973) "Psychological representation of speech sounds," in David, E.Jr. and P. Denes, eds., Human Communication: A Unified View, New York: McGraw-Hill, pp. 67-113.
- Shepard, Roger, A. Romney and S. Nerlove (1972) Multidimensional Scaling: Theory and Practice, New York: Seminar.
- Singh, Sadanand and George Woods (1971) "Perceptual structure of 12 American vowels," J. Acoust. Soc. Amer. 49:1861-1866.
- Stevens, Kenneth (1971) "Perception of phonetic segments: evidence from phonology, acoustics and psychoacoustics," in Horton, D. and J. Jenkins, eds., The Perception of Language, Columbus, Ohio: Merrill, pp. 216-235.

- Sweet, Henry (1890) A Primer of Phonetics (third edition), Oxford: Clarendon.
- Terbeek, Dale (1973) "Six dimensions of vowel quality," Proc. Ninth Regional Meeting, Chicago Linguistic Society.
- (1974) "Multidimensional scaling of the perception of diphthongs," paper read to the Classification Society, Ann Arbor.
- Terbeek, Dale and Richard Harshman (1972) "Is vowel perception non-Euclidean?" Working Papers in Phonetics 22:13-29.
- Thurstone, L. (1947) Multiple Factor Analysis, Univ. of Chicago.
- Torgerson, W. (1958) Theory and Methods of Scaling, New York: Wiley.
- Trager, George and Henry Smith, Jr. (1957) An Outline of English Structure, American Council of Learned Societies Studies in Linguistics: Occasional Papers vol.3.
- Trubetzkoy, N.S. (1969) Principles of Phonology (trans. C. Baltaxe), Berkeley: Univ. of California.
- van der Kamp, L.J.Th. and L.C.W. Pols (1971) "Perceptual analysis from confusions between vowels," Acta Psychologica 35:64-77.
- Wickelgren, Wayne (1965) "Distinctive features and errors in short-term memory for English vowels," J. Acoust. Soc. Amer. 38:583-588.
- Wiener, F. and Singh, S. (1974) "Multidimensional analysis of choice reaction time judgements on pairs of English fricatives," J. Exper. Psychol. 102:615-620.
- Young, F.W. and W.S. Torgerson (1967) "TORSCA, a Fortran IV program for Shepard-Kruskal multidimensional scaling analyses," Behavioral Science 12:498.