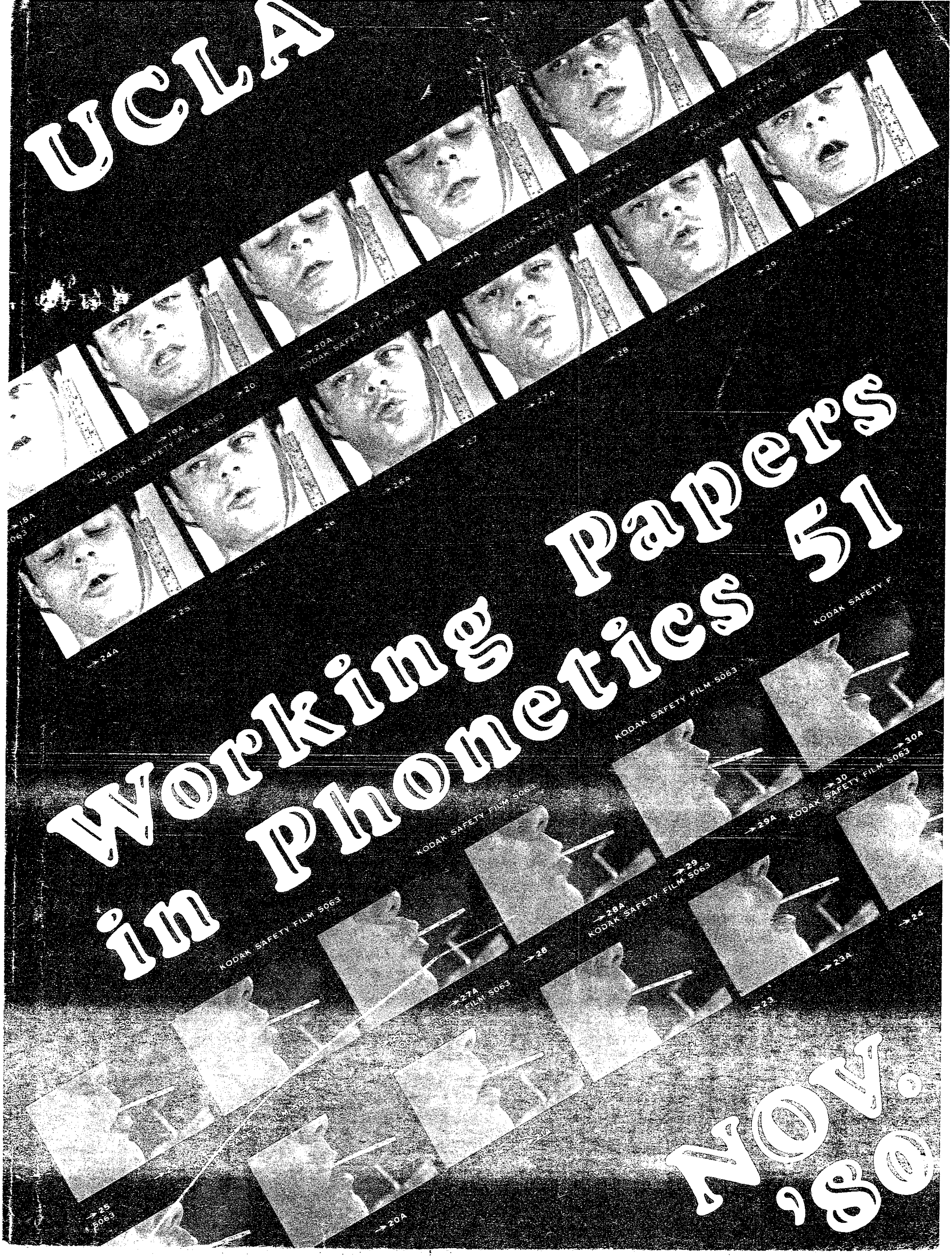

UCLA
Working Papers
in Phonetics 51
NOV. '80
KODAK SAFETY FILM 5063

UCLA Working Papers in Phonetics 51

November 1980

Table of Contents

The UCLA Phonetics Laboratory Group

Diane Ainsworth
Ron Anderson
Stephen R. Anderson
Sandra F. Disner
James F. Fordyce
Vicki Fromkin
Manuel Godinez
Steven Greenberg
Kathleen Houlihan
Richard Janda
Hector Javkin
Peter Ladefoged
Mona Lindau(-Webb)
Wendy Linker
Ian Maddieson
Priscilla McCoy
Jonas Nartey
George Papcun
Lloyd Rice
Diane Ridley
Mary Smith
Anne Wingate
Andreas Wittenstein
Eric Zee

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

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS Grant NS09780
NSF Grant BNS-78-07680
NSF Grant BNS-79-10051

and the UCLA Department of Linguistics.

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

The Phonetics Laboratory
Department of Linguistics
UCLA
405 Hilgard Avenue
Los Angeles, CA 90024

A Cross-Linguistic Study of Lip Position in Vowels

Wendy Linker

I. Introduction

The purpose of this investigation is to discover whether languages may differ in the ways in which they exploit variations in lip position in the production of vowels. There have already been numerous instrumental studies done on lip position in vowels in various languages (particularly French (Abry, Boë, Gentil, Descout and Graillot, 1979), Swedish (McAllister, et al, 1974; McAllister, 1978), and English (Fromkin, 1964)). In addition, traditional phonetic descriptions of vowels in different languages have either implicitly or explicitly stated that there are, in fact, systematic differences among perceptually similar vowels in different languages, as indicated, for example, by Sweet (Henderson, 1971, pg.64): "In English the lips are less spread than in many other languages, such as French and German". Thus, the broad question raised in this study is the following: Are there in fact measureable differences in lip positions in vowels in different languages? First, however, it may be useful to review a small portion of the research on lip position in vowel production, as a starting point from which the present study is undertaken.

Traditional phonetics generally embraced the view that although in principle any lip position could co-occur with any tongue position, there were in fact more "normal" or at least more probable combinations to be found in the vowels of different languages. For example, Henry Sweet writes, "Rounding can, of course be added to all the tongue positions. The degrees of rounding are infinite. As fixed points we distinguish three, corresponding to the three heights of the tongue, the general rule being that the higher the tongue position of the round vowel, the narrower the lip passage...but abnormal rounding also occurs. There is no difficulty, for instance, in combining mid position of the tongue with high rounding as in the second element of ou in no...it is also possible to under-round. The vowel in good is ´under-rounded´ in the dialects of the North-West of England: the high position of the back of the tongue is retained, while the lips are relaxed almost to low rounding" (ibid,pg.86). As another example of "abnormal" rounding in language Sweet cites both Danish and Swedish /o/, where the "lip narrowing in each case [is] a degree above the height of the tongue" (ibid, pg.64).

Sweet also distinguishes between "inner" and "outer" rounding: "In outer-rounding, with which front vowels are rounded - the lips are brought together vertically...Back and mixed vowels, on the other hand, are rounded by lateral compression of the corners of the mouth and, apparently, of the cheeks as well" (ibid, pg.63).

Daniel Jones, in his description of the cardinal vowels, also sets up an abstract system whereby tongue position and lip position are closely correlated, in a manner similar to Sweet. Jones distinguishes three kinds of lip rounding: "spread", "rounded" and "neutral", although "spread" and "neutral" are grouped together as "unrounded" (Jones, 1960, pg.39). As this system was intended to aid phoneticians in their descriptions of actual languages, it cannot be an accident that Jones set up these reference points with such close correlations between tongue height and lip position. Thus it is clear that both Sweet and later Jones had a strong notion of naturalness with regard to the use of lip rounding in vowels in the languages of the world; both were concerned with degrees of lip rounding, and, in the case of Sweet, with different types of rounding as well.

Fromkin's article entitled "Lip positions in American English vowels" (1964) represents one of several careful investigations into the movement of the lips in the production of vowels within a given language. This study will be described in some detail, as it is one of the most relevant to the present investigation. Simultaneous frontal and lateral photographs were taken of the production of 12 American English vowels produced by 10 subjects; in addition, lateral x-rays and stone casts of the subjects' lips were made. One main purpose of Fromkin's investigation was to measure a number of parameters of lip position and study the relationship among these parameters for English vowels. Seven measurements of the lips were taken: (1) width of the lip opening, (2) height of the lip opening, (3) area of the lip opening, (4) distance between the outermost points of the lips, (5) protrusion of the upper lip, (6) protrusion of the lower lip, and (7) the distance between the upper and lower front teeth. With respect to the present investigation, the relevant conclusions discussed by Fromkin are the following: (1) Where four subjects were considered, the absolute values of the measurements differed greatly; however, the <u>relationships</u> among the vowels when graphed along the lip parameters remained the same. (2) Only the English /u/ was distinguished by the parameters width and height of lip opening. (3) In general, lip positions do serve to distinguish <u>sets</u> of vowels in English (i.e., the front unrounded vowels from the back rounded vowels), but do not serve to distinguish vowels within particular groups (/u/ being the exception). (4) No correlation was found for English between protrusion and height of lip opening. (5) A linear relationship was found to exist between the product width x height and the area of the lip opening.

Thus, Fromkin's study showed that one can, in fact, quantify relationships among various parameters of lip position for the vowels of English; however, the question of how these parameters co-vary in other languages still remains.

Many studies have also been conducted investigating the activity of the lips within a single language (e.g., Fujimura, 1961, Fromkin, 1965, McAllister, Lubker & Carlson, 1974, McAllister, 1978). With regard to the present study, McAllister's work on Swedish rounded vowels is particularly relevant. Essentially, it was found that the traditional classification of the Swedish rounded vowels (i.e., inrounding vs. outrounding) was, in fact, reflected in the electromyographic activity of the orbicularis oris muscle. (This result will be discussed in more detail later, in relation to my own results). Thus it seems to be the case that one can measure differences in lip activity in vowels, using x-ray, EMG and photographic techniques, and that the results obtained from such studies are largely in agreement with traditional descriptions of lip gestures in vowel production.

The final type of investigation to be discussed involves the relationship between the acoustic nature of vowels and lip position. Recall that Sweet, for example, assumed a natural relationship between tongue position and lip position. Could it also be the case that there is some natural relation between the acoustic vowel space and the lip positions associated with points in this space? A study by Stevens and House (1955) represents a step in the direction of answering this question; they specified variations in the effect of the lips by the ratio A/l, where A represents an idealized area of the mouth opening (determined by adjusting parameters of a vocal tract analog) and l is equal to the length of an idealized vocal tract tube. They plotted this single number A/l against formant frequency, and thus found correspondences between changes in F1, F2 and F3 as A/l varied. However, because A/l represents only idealized lip shapes (which themselves are based on perceptual auditory judgements), it is obviously still necessary to discover the relation between actual measures of lip position and acoustic vowel quality in real languages, in order to approach the question raised above.

The next step in this direction was to investigate the relationship between lip position and acoustic properties of vowels within a single language. Ladefoged, et al (1978) measured lateral x-rays of five speakers of American English producing ten vowels, and found that the distance between the upper and lower lips could be predicted from a particular combination of formant frequencies. The correlation between the observed lip distance and the predicted lip distance was 0.78. However, as this study was not primarily concerned with lip articulation only this single lip parameter and its relationship to formant frequency was considered.

We have seen that much research has been devoted to the study of lip activity in vowels, and that some of this research has focused on the relationship between the acoustic nature of vowels and lip position. However, most of this work has been confined to the investigation of a single language at a time, and in most cases no more than five subjects were used (usually fewer). I would now like to go a step further, and compare measurements of both lip position and acoustic characteristics of vowels across languages, so that we may be in a position to verify, and if shown to be viable, to quantify the observations described in the traditional phonetics literature.

Before undertaking this cross-linguistic study, an investigation of lip position in American English vowels was carried out (Linker, 1978), in order to answer the following questions:

(1) Would it be possible to obtain a higher correlation between lip position and formant frequency than that found by Ladefoged, et al if more subjects were used and more measurements of lip position taken?

(2) Which measure or measures of lip position correlate most highly with formant frequency?

II. A Study of Lip Position and Formant Frequency in American English Vowels

Methods

Simultaneous frontal and lateral photographs were taken of eight male subjects pronouncing nine words: heed, hid, hayed, head, had, hod, hoed, hood, who′d. A device, which had two cameras mounted on it, was used to keep the subject′s head steady during the production of each word. The subject was instructed to sustain the vowel of each word until after the simultaneous photographs were taken. The entire session, conducted in a sound attenuated booth, was recorded. In the recording, an audible click from the camera shutters indicates the exact time at which the photographs were taken. In order to measure lip protrusion, it was necessary to decide on a fixed reference point to measure from. The upper front teeth were selected, since this structure moves less than any other structure in the front of the mouth. A narrow paper strip marked in 5 millimeter sections was attached to a very thin metal wire, and the wire was fastened between the subject′s upper front teeth. From this, relevant distances to the upper teeth could be calculated. Care was taken to ensure that the stick did not touch either the upper or lower lip of the subject. In addition, a millimeter scale was placed in the frontal view.

To obtain measures of lip position, negatives of the front and side views were enlarged, and tracings were made. Exact distances were calculated from the scales using a computer program written by Lloyd Rice. With this program it was possible to use a Grafpen connected to the PDP-12 computer to retrace the tracings of the negatives. The program used the points marked by the Grafpen as input to compute 24 lip measurements. The measurements in the side view were taken with reference to the five points shown in Figure 1: (1) Point B, a fixed point on the upper front teeth, (2) Point C, the corner of the mouth, (3) Point D, the most forward point of contact between the lips, (4) Point E, the outer point on the lower lip surface, and (5) Point F, the outer point on the upper lip surface. Every possible direct distance from each of these five points to one another was found. In addition, selected perpendicular distances were taken from points B, C, and D to: (a) the tangent line through points E and F and (b) perpendicular lines drawn through points B, D, and F. Finally, the area enclosed by the lips was calculated. Only 12 of these side view variables are considered in the analysis due to problems of scaling the values properly.

Figure 2 shows the front view measurements that were calculated. Measures 20, 21, and 22 represent the vertical distance between the lips taken at equal intervals, with measure 21 being at the center of lips. Measure 23 is the distance between the two corners of the mouth, and measure 24 is the area of the mouth opening in the front view.

FIGURE 1: A diagram showing the measurements taken from the lateral view of the lips.

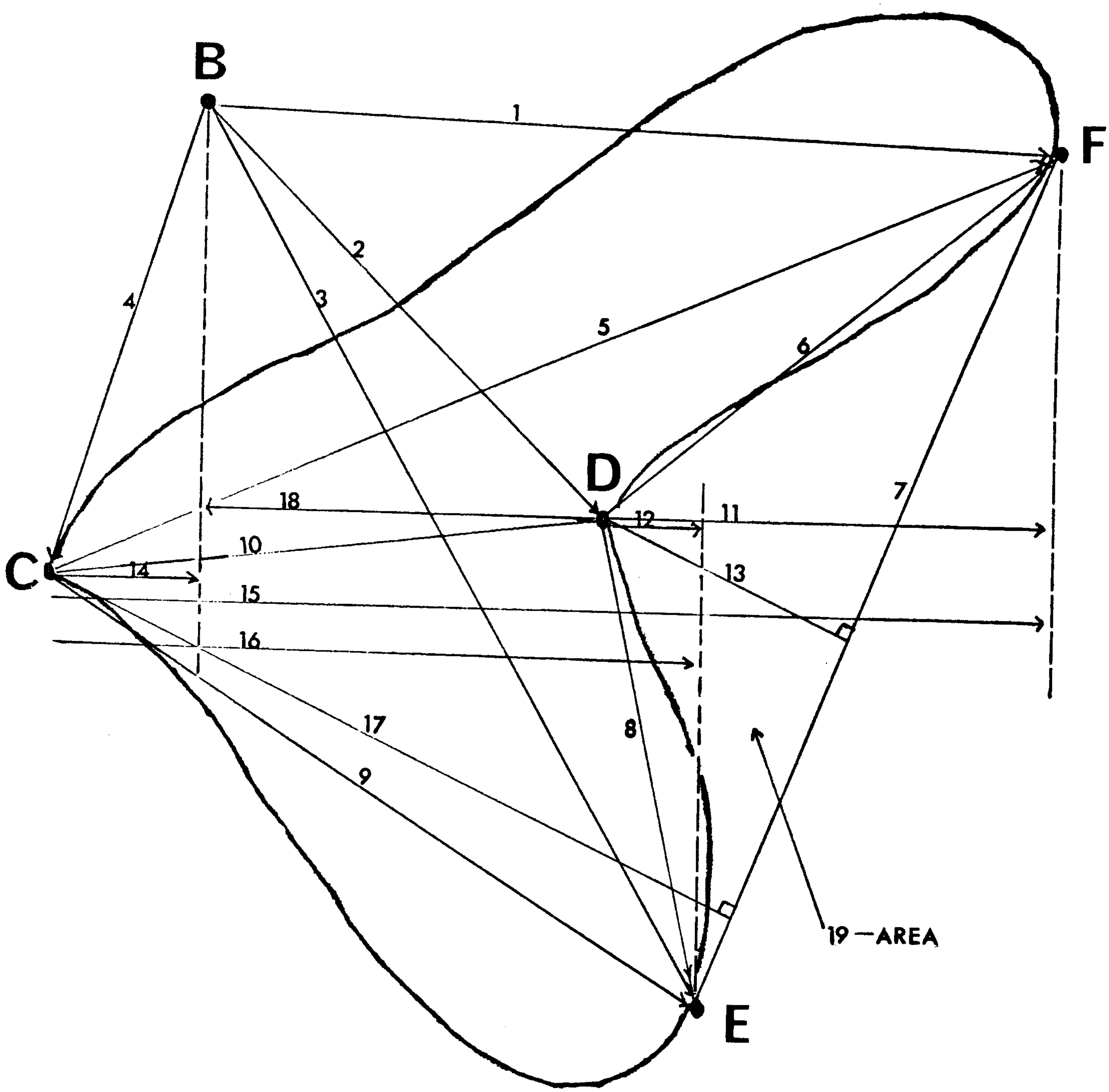
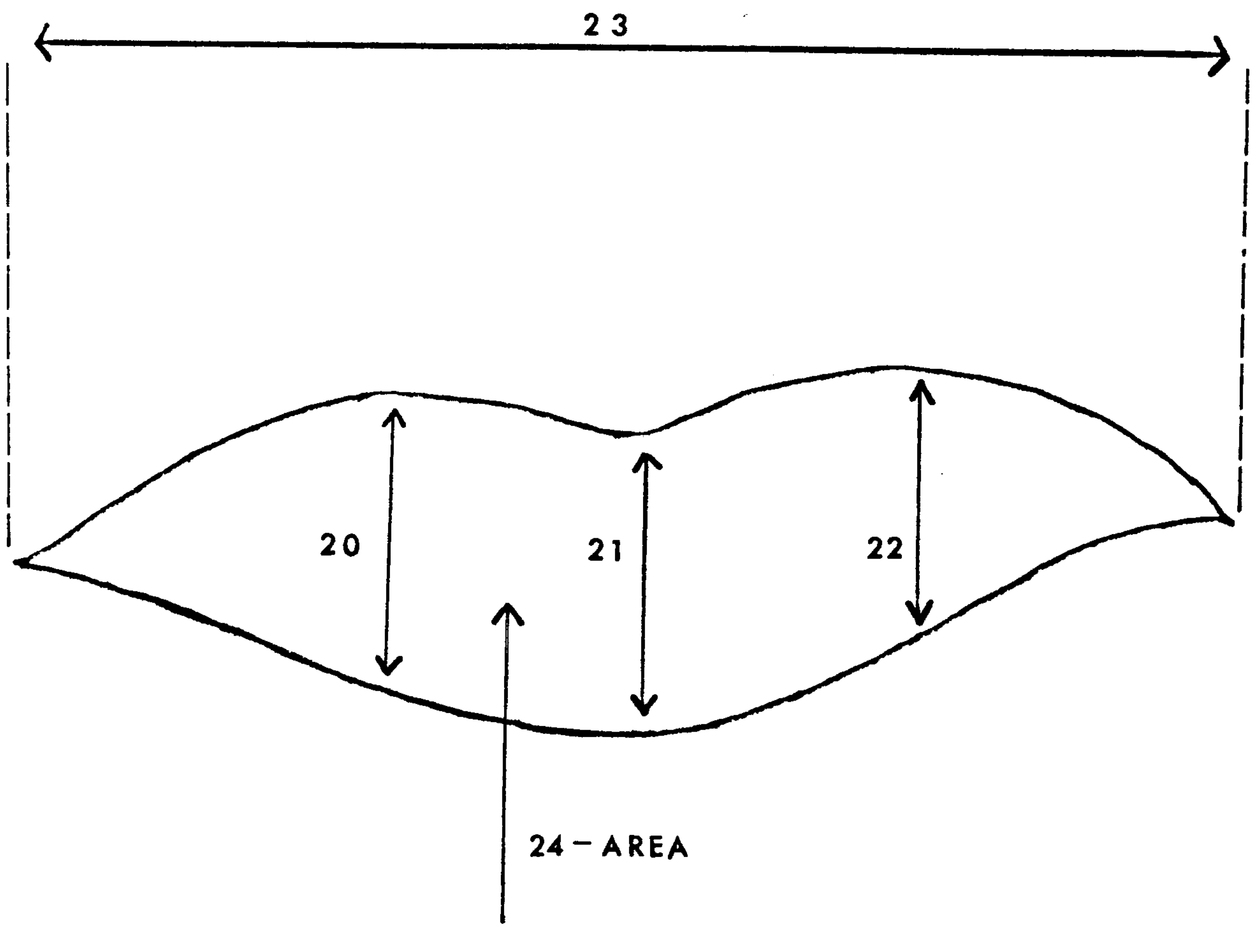


FIGURE 2: A diagram showing the measurements taken from the frontal view of the lips.



(The reader may find it useful to refer to figures 1 and 2 throughout this paper). Thus, 12 side view variables and 5 front view variables entered into the analysis of the English vowels, giving a total of 17 lip variables.

Wide band spectrograms were made of each utterance photographed and the formant frequencies F1, F2, and F3 were measured at a point immediately prior to the click of the camera shutters. (This click was easily visible on the spectrogram). In some cases, where interpretation was difficult, linear prediction analysis was used to supplement the spectrograms. From these three primary acoustic measurements, 22 additional acoustic variables were obtained by calculating products and ratios of combinations of formant frequencies. These were used in the regression analysis discussed below.

Analysis and Results

The resulting data consists of a set of three formant frequencies matched with a set of 17 measures of lip position for each of nine vowels for each of eight speakers. In order to see how the lip measures covaried with one another for different vowels and different speakers, the lip measurement data alone (i.e., not including the acoustic data) were subjected to a 3-mode factor analysis, PARAFAC (Harshman, 1970, Harshman, Ladefoged and Goldstein, 1977).

As shown below, PARAFAC gives as a result three sets of weightings, also referred to as ʹloadingsʹ: the first set (m) represents the weighting of each of the 17 lip measures on the extracted factors; the second set (v) represents the weightings of the nine individual vowels on the factors, and the third set (s) represents the weightings of the eight individual speakers with respect to the extracted factors.

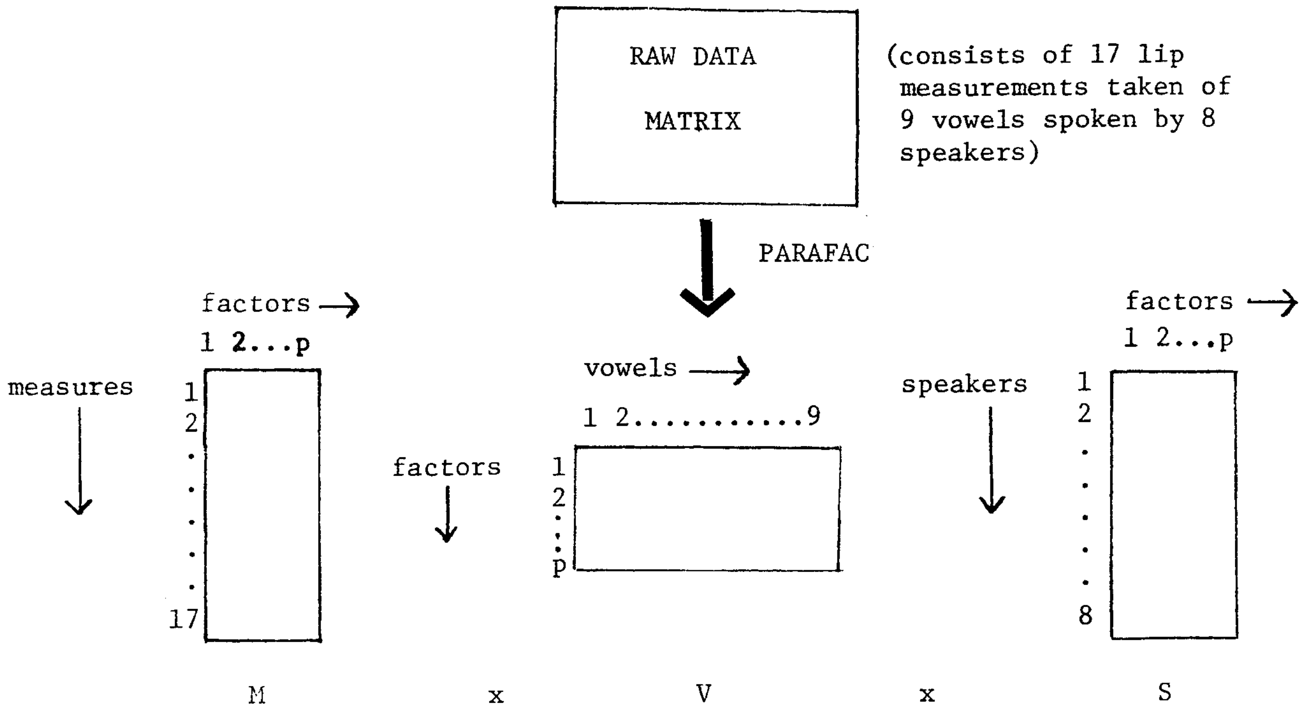
THE PARAFAC EQUATION

$$d_{ijk} = m_{i1}v_{j1}s_{k1} + m_{i2}v_{j2}s_{k2} + \cdots + m_{ip}v_{jp}s_{kp}$$

where d_{ijk} = a single data point for lip measure i for vowel j as spoken by speaker k;

m,v,s = the loadings (weightings) of lip measure, vowel, and speaker respectively on factors 1 through p.

Schematically, we are breaking down the original raw data matrix in the following way:



(matrix of 17 measure loadings on factors 1 through p)

(matrix of 9 vowel loadings on factors 1 through p)

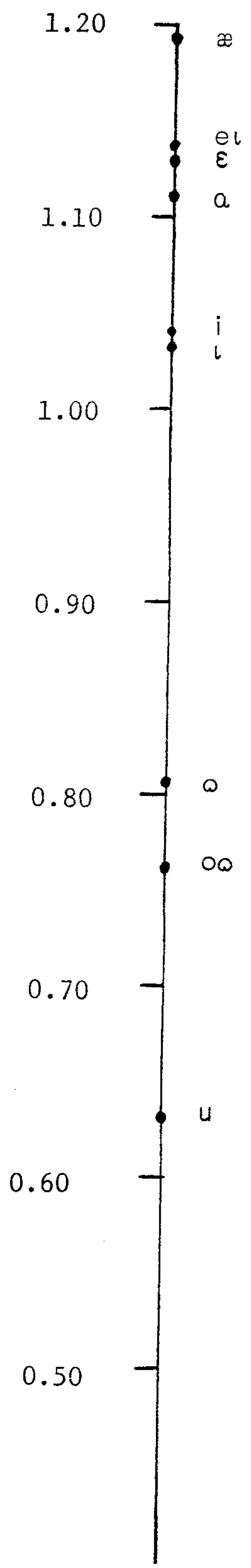
(matrix of 8 speaker loadings on factors 1 through p)

(That is, the raw data matrix D is equal to the product M x V x S. Thus, the factors can be thought of as <u>dimensions</u> which underlie the data set.)

Both 1-factor and 2-factor unique solutions were obtained from different random starting configurations, and the solutions for each number of factors converged with each random start. Only one factor was needed to account for 92% of the variance in the data (R^2 = 0.923), while R^2 for the 2-factor solution was 0.957. Because very little additional variance was accounted for by the second factor, it was decided that the 1-factor solution was the correct solution to work with. Although interpretation of the factor is not completely straightforward, it can be seen in figure 3 that the values for the vowel loadings tend to be higher for vowels with a large mouth opening and lower for rounded vowels. Thus, this factor is related to a traditional description of lip position in vowels. Note that the fact that only one factor was needed to adequately describe these data does not conflict with Fromkin's result that protrusion and vertical opening are uncorrelated for English. Further study would be needed to ascertain whether or not Fromkin's data could be described by the same factor.

Given that the 17 lip variables are so well determined by only one factor, and given that this factor lends itself to a reasonable interpretation, the factor may be taken to represent the underlying determiner of lip position for these data. Therefore, we would like to predict this one lip factor from formant frequencies, in order to arrive at a measure of the correlation between the acoustic data and

FIGURE 3: English vowel loadings from PARAFAC along the underlying factor L



the lip variables taken overall. It would be possible to predict each of the 17 lip measures from the formant frequencies, using 17 different equations. But these predictions would not use the fact that some of the 17 measurements are highly correlated, and should not be allowed to vary independently. It is therefore appropriate to predict them from the same equation.

To determine how to predict this underlying lip parameter from the formant frequencies, a stepwise multiple regression program was run, using the individual products of the 8 speaker loadings and the 9 vowel loadings given by the PARAFAC analysis. As can be seen under (1) in Table I, a fairly high degree of correlation (R = 0.831) was achieved with only two terms. Having predicted the values of the product vowel loading x speaker loading along the underlying lip factor, it was now possible to take the actual lip measurement loadings (i.e., the m loadings), given by PARAFAC, together with the vowel x speaker loadings predicted by the regression, to reconstruct a predicted set of lip measurements, that depend on formant frequency. In other words, the product M x V x S as determined by formant frequency was compared with the original raw data. The correlation obtaining between these predicted lip measurements and the actual lip data is 0.945.

We can now look at the difference between predicted and observed values for each individual lip measure across all vowels and subjects, to see which measures were best recovered. Here it is important to remember that only one regression equation is being used, namely the one that predicts the underlying lip parameter, and it is from this parameter that the other measures are predicted.

The correlation coefficients for the 17 lip variables are listed under (2) in Table I. The lip measurement that is best recovered is number 6, which is a measure of upper lip protrusion from the most forward point of contact between the lips (point D). The measure that is predicted most poorly is measure 1, the protrusion from the tooth (point B) to the upper lip (point F). Although it might be thought that both these measures are good descriptions of the amount of lip protrusion, apparently only the distance from the most forward point of lip contact can be predicted accurately for English vowels from the acoustic data.

This preliminary study of lip position and formant frequency in English vowels was encouraging in two respects: (1) it is possible, for English vowels, to use PARAFAC as a method of reducing a large number of lip measurements into a single, representative parameter and (2) it is possible, for English vowels, to predict lip positions from formant frequencies with a fairly high degree of accuracy.

TABLE I

(1) Regression equation to predict the underlying lip factor L from 72 tokens (8 speakers, 9 vowels). Multiple R = 0.831

$$L = 20.44 \frac{F2}{F3} + 0.005 \frac{F1 \times F3}{F2} - 1.610$$

(2) Correlation between predicted and observed measures = 0.945 (17 lip measures, 9 vowels, 8 speakers)

(3) Individual correlations for the 17 lip measures:

Side View

1.	Distance from tooth (B) to outermost point on upper lip	-0.191
2.	Distance from B to most forward point of contact between the lips (D)	0.683
3.	Distance from B to outermost point on lower lip (E)	0.332
4.	Distance from B to corner of mouth (C)	0.670
5.	Distance from C to F	0.641
* 6.	Distance from D to F	0.828
7.	Distance from E to F	0.683
8.	Distance from D to E	0.744
11.	Distance from D to perpendicular line through F	0.810
12.	Distance from D to perpendicular line through E	0.690
13.	Distance from D perpendicularly to line connecting E and F	0.738
18.	Distance from D to perpendicular line through B	0.684

Front View

20.	Vertical distance between the lips, 1	0.409
21.	Vertical distance between the lips, 2 (center)	0.426
22.	Vertical distance between the lips, 3	0.425
23.	Distance between the corners of the mouth (horizontal opening)	0.610
24.	Area of mouth opening	0.514

III. The Cross-Linguistic Study

Given the success of this approach with English vowels, it was decided to use PARAFAC as a tool for comparing lip positions in vowels of different languages. An additional advantage to this is that PARAFAC deals effectively with the great amount of variation among individual subjects that can obscure trends that we might observe across languages. My intention in this study has been to discover general patterns of lip articulation in the vowels of the languages I have investigated. In particular, I have been concerned with two fundamental questions: first, do similar vowels in different languages - that is, vowels transcribed in the same way by phoneticians - make use of different lip gestures? For example, if we look at the degree of lip protrusion used in French vowels and compare it to lip protrusion in Swedish vowels, will we find differences? The second question I want to briefly examine is the following: if in fact we do find differences in lip positions across similar vowel systems, to what extent do these differences reflect the fact that the vowels in these systems are not precisely the same? That is, if the actual vowels under investigation were acoustically identical, would we observe differences in lip position anyway?

Methods and Analysis

The languages selected were Swedish, Finnish, French and Cantonese. All have vowel systems which contain at least vowels similar to [i,e,a,o,u,y,ø]. The languages chosen are genetically diverse in order to maximize the chance of finding differences in lip position. Wordlists were prepared containing minimal or near-minimal contrasts of all stressed vowels in these languages, long vowels in open syllables being preferred where possible. (See Appendix I). In some cases nonsense words were used. Eight male speakers of each language were rehearsed in reading the relevant wordlist. Subjects were instructed to sustain the vowel of each word. After training, a subject was seated in a dentist's chair which held his head steady while simultaneous frontal and lateral photographs of his lips were taken during the production of each word. The entire session was recorded. Measurements of the first three formants of each vowel were made on wide band spectrograms prepared from the recordings. The moment at which the photographs were taken was determined from the noise of the camera shutters on the spectrograms, and measurements of formant frequency were made just before that instant.

Measures of lip position were obtained using a somewhat modified version of the same computer program as was used with the English data. However, the measures that were omitted before were included this time because it was possible to scale them properly. It was, nevertheless, not possible to include measure 10 in the analysis. Measure 10 is the distance between points C and D (the corner of the mouth and the most forward point of lip contact; please refer to figure 1) and is equal to zero for many unrounded vowels. The version of PARAFAC used here does not accept a value of zero as data.

The data thus consist of a set of formant frequencies matched with a set of 23 measures of lip position for each of n vowels for each of 8 speakers of each of four languages. The lip measurement data from each language was then analyzed separately by PARAFAC. Between three and five solutions of 1, 2, and 3 factors were obtained for each language, as well as two solutions of 4 factors for each language. The 4-factor solutions are not considered further because a unique solution was not obtainable, nor did the solutions converge. Table II lists the squared correlation values obtained for the 1, 2, and 3 factor solutions, and figure 4 plots the dimensionality of the solutions against the squared correlations.

Choosing the Solutions

The criteria generally used for choosing a PARAFAC solution are the following: (1) uniqueness (the same solution is obtained when the analysis is carried out from different starting configurations), (2) convergence (starting from one starting configuration, after a certain number of iterations no further change occurs in the solution), (3) the value of the squared correlation (amount of variance accounted for) and (4) interpretability of factors. The 4-factor solutions were discarded immediately because these were either non-unique or did not converge, there was little increase in the squared correlation as compared to the 3-factor solutions, and the fourth factor was uninterpretable.

It can be seen from Figure 4 and from Table II that (1) for all languages there is a sharp increase in the squared correlation in progressing from a one dimensional analysis to a two dimensional one, and a smaller but still substantial increase in the squared correlation (especially for French and Swedish) in going from two dimensions to three dimensions, and (2) convergence and uniqueness are obtained at three dimensions in all languages except Cantonese. These facts leave us with a choice between the 2- and 3- factor solutions.

Here it is important to clarify the role PARAFAC is playing in this study. The goal of the factor analysis is twofold, namely: (1) to provide an adequate description of the lip measurement data; that is, there should be a high correlation between the original data and the lip measurements predicted by PARAFAC, and (2) to analyze each language in such a way that the analyses will be compatible enough for comparison across languages. Thus, in choosing between the 2 factor and 3 factor PARAFAC solutions, we are concerned that (1) the solutions adequately represent the data and that (2) the solutions are comparable.

Let us postpone for a moment the decision about dimensionality and examine the method chosen to compare lip position across languages.

After PARAFAC has been carried out, there are four separate factor spaces,one for each language, each of which correlates very highly with its corresponding raw data matrix. These solutions cannot be directly compared, however, because they may represent different coordinate systems. In order to solve this problem of comparability, a generalized type of canonical correlation (Carroll, 1965) was used, in the form of a computer program called CANON, implemented by Louis

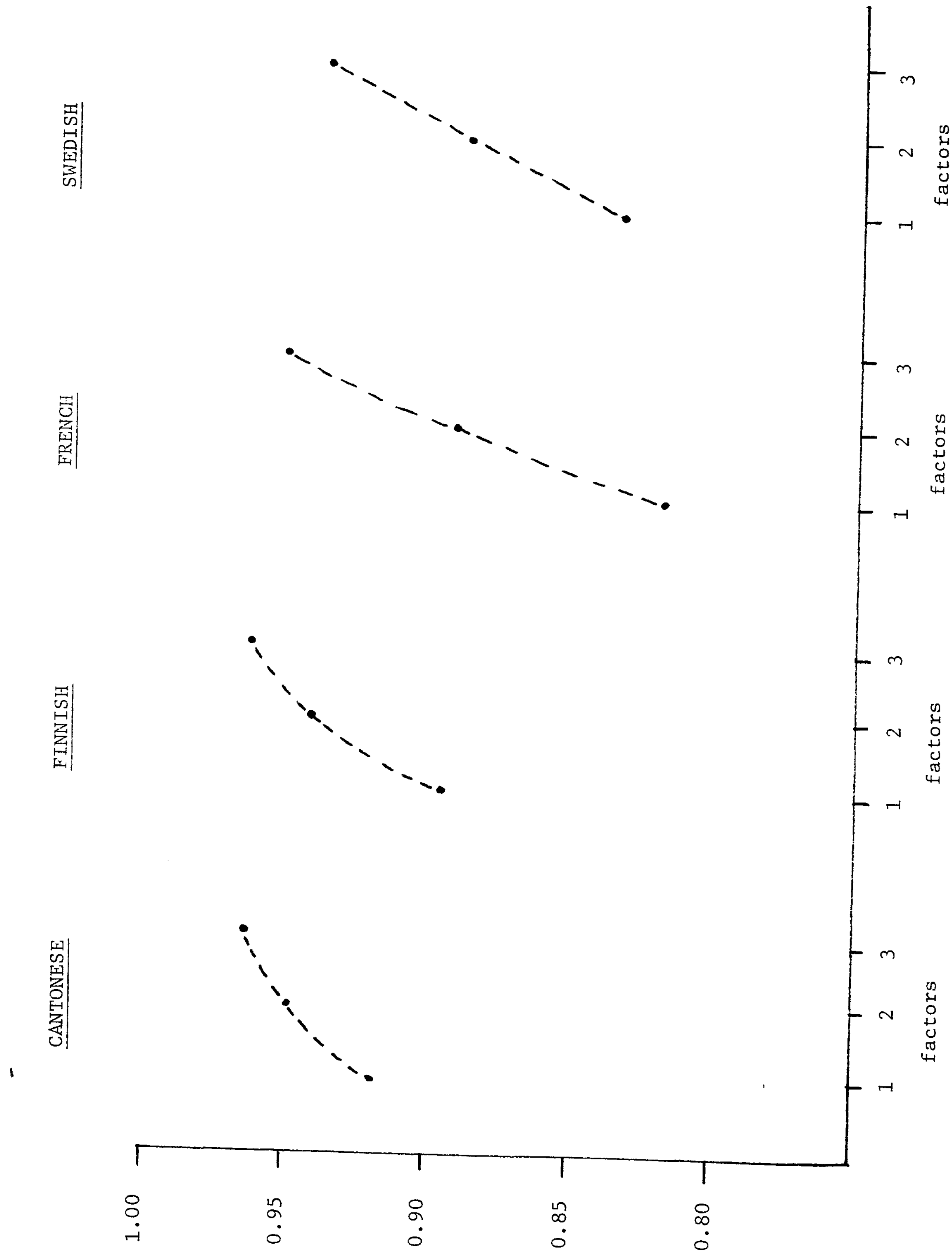
TABLE II: R^2 values from PARAFAC

	CANTONESE	FINNISH	FRENCH	SWEDISH
1 factor solution	0.904	0.894	0.817	0.834
2 factor solution	0.945	0.942	0.891	0.889
3 factor solution	0.964 *	0.963	0.952	0.938
4 factor solution	* +	* +	* +	* +

* = solution did not converge

\+ = no unique solution found

FIGURE 4: R^2 values for solutions from PARAFAC (ordinate) plotted against the number of factors extracted (abcissa). The values are plotted separately for each language.



Goldstein at UCLA. CANON takes as input N data sets (matrices) and produces a single matrix, which represents a least squares approximation to the input data matrices. The new matrix is arrived at one vector at a time, by finding a linear composite of each of the N matrices that correlates most highly with the first canonical vector (Z). This Z vector is chosen so as to maximize the sum of the squared correlations (r_i^2) between itself and each linear composite of each data set; that is, the quantity $\sum_{i=1}^{N} r_i^2$ (Carroll's R^2) is maximized, and is an index of how good the fit is between the original data sets and the matrix produced by CANON. There will thus be n Z-vectors, n R^2 values and n linear composites per data set given that there are n dimensions in the data set of lowest dimensionality.

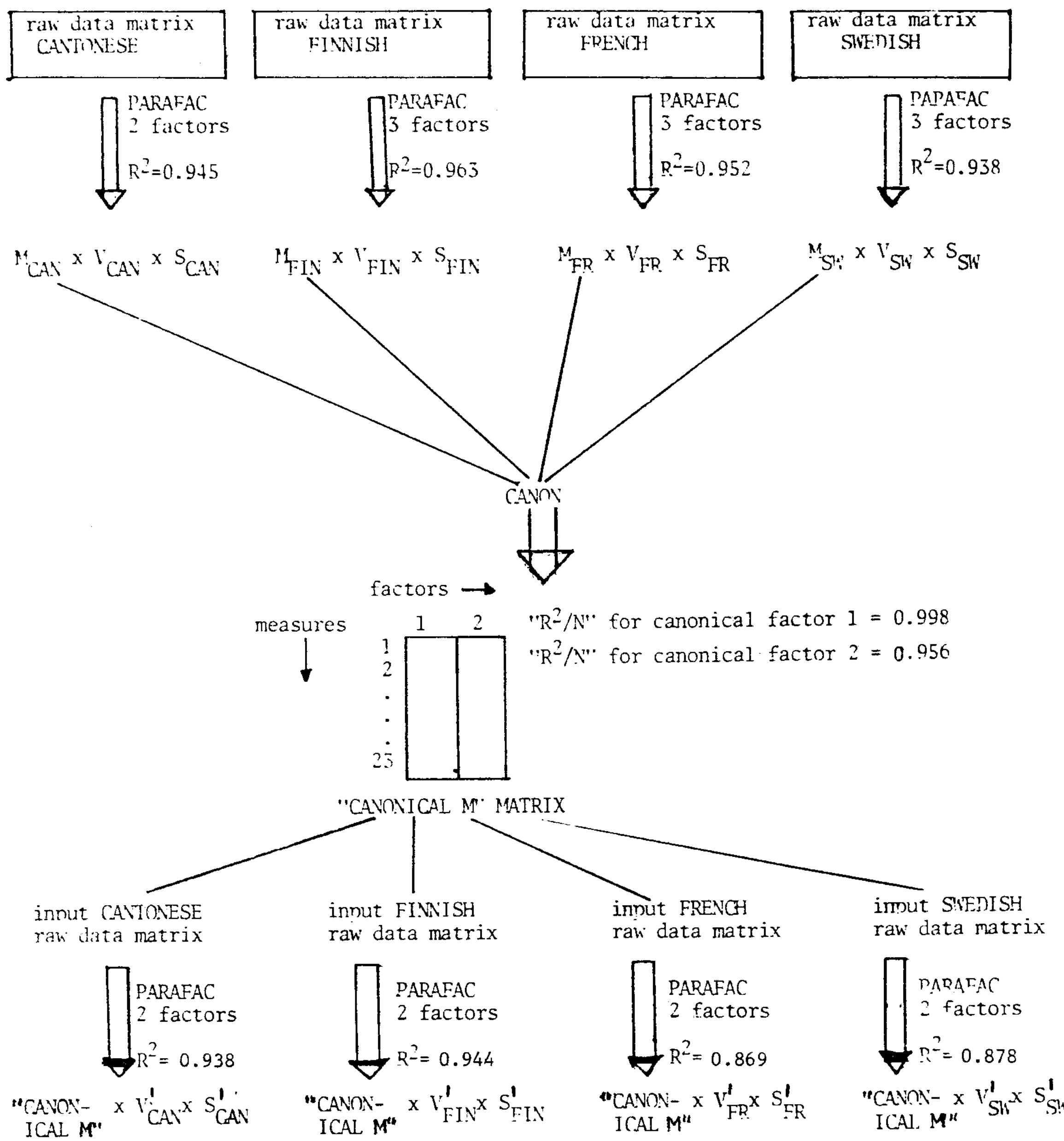
CANON can be applied to the present problem of comparing lip measurements across languages in the following way. If we take the four sets of lip measure loading matrices (M) as input to CANON, we will get out a new matrix (henceforth referred to as the "Canonical M" matrix) that represents a least squares approximation to the four input lip measure loading matrices. If this new matrix shows a good fit to the original four matrices, then we will have succeeded in finding a single coordinate system within which to describe lip position across languages. That is, we will be in a position to compare how vowels in different languages make use of a <u>single</u> set of underlying factors.It should be noted that using CANON in this way is tantamount to rotating the PARAFAC solutions, which technically should lower the fit of the data to the PARAFAC model; however, a high R^2/N will indicate that the original factor spaces do not need to be altered drastically to conform to the Canonical M space.

Returning to the question of choosing between the two and three dimensional PARAFAC solutions, it is now clear that the decision will also be affected by the value of Carroll's R^2/N. Thus, not only should the PARAFAC solutions fulfill the criteria raised earlier, but the solutions chosen must be sufficiently compatible for R^2/N to be high.

This leaves us with three alternatives: (1) 2-factor solutions for all languages, (2) 3-factor solutions for all languages, and (3) a 2-factor solution for Cantonese and 3-factor solutions for the other languages.

Alternative (2) is not feasible because there is no unique solution at three dimensions for Cantonese (see Table II). Alternative (1) is feasible in terms of the criteria of uniqueness, convergence, etc.; however, R^2/N for the first canonical variate (Z vector) is 0.995, while for the second canonical variate it is only 0.547. Alternative (3) fulfills all the criteria: the solutions are unique, they converge, the PARAFAC R^2 is very high in all cases, and Carroll's R^2/N for both canonical variates is very high (0.998 and 0.956, respectively). Thus it was decided to continue the analysis using the results obtained from a Canonical M matrix of two dimensions, derived from one two dimensional space and three three dimensional spaces.

FIGURE 5: Outline of the analysis



How PARAFAC and CANON were applied to the data sets of each language is diagrammed explicitly at the top of Figure 5. The remaining portion of the figure will be explained later, and the reader may refer to it when necessary.

Factor Interpretation

Interpretation of the canonical factors is not completely straightforward, as is often the case with factor analysis, but we can get a fairly good idea of what these factors represent by examining the ordering of the measure loadings in the Canonical M space (see Figure 6). With regard to factor 1, at one end of the scale we find measures from point B (the fixed point on the upper teeth) to points C and D (the corner of the mouth and the most forward point of contact between the lips, respectively). At the other end of the scale we find measure 24 (the area of the mouth opening in the front view), and extremely far down at this end we find measure 23 (the horizontal opening in the front view). Looking at the loadings along canonical factor 2 we find at one end measures 1, 3, and 5 (protrusion from the teeth), while at the other end of the scale are measures 20, 21, and 22 (all of which represent vertical opening in the front view), as well as measure 24 (the front view area). Measure 23, the horizontal opening, appears in the middle of the scale, so we expect its contribution to canonical factor 2 to be small.

To summarize, canonical factor 1 can be said to be related mainly to horizontal opening (with some component of lip protrusion as well), while canonical factor 2 appears to be directly proportional to lip protrusion as measured from the teeth, and inversely proportional to vertical opening between the lips.

Perhaps an easier way to see the effects of factors 1 and 2 is illustrated in Figures 7 and 8. In the side view, the solid line represents a neutral lip shape, given a fixed point B. This neutral lip shape is based on the raw data measurements taken for the Cantonese vowel [ɛ], as produced by one speaker. The dashed lines represent deviations from the neutral lip shape, calculated by adding (C′-F′) and subtracting (C′′-F′′) values of particular canonical lip measure loadings (i.e., those loadings plotted in Figure 4; measure loadings 1,2,3,and 4 were used). These loadings were first normalized by PARAFAC (to give a mean squared loading of 1) and were then multiplied by a constant. By adding and subtracting these factor loadings, we see the effect of increases and decreases in the factors themselves. The same procedure was used to see the effect of the factors looking at the front view of the lips, using measure loadings 21 and 23. Note that the deviations drawn do not represent actual reconstructed lip shapes, and these diagrams are to be taken only as illustrations of the relative <u>effects</u> of canonical factors 1 and 2. We can see, for example, that for canonical factor 1 (Figure 7) that a positive value of this factor represents a very slight increase in protrusion, combined with a large <u>decrease</u> in horizontal opening, as we saw by examining the Canonical M space itself. Similarly, a large positive value of canonical factor 2 (Figure 8) represents a large

FIGURE 6: CANONICAL M MATRIX

(Lip measure loadings on canonical factors 1 and 2; true loadings x 10^{-2})

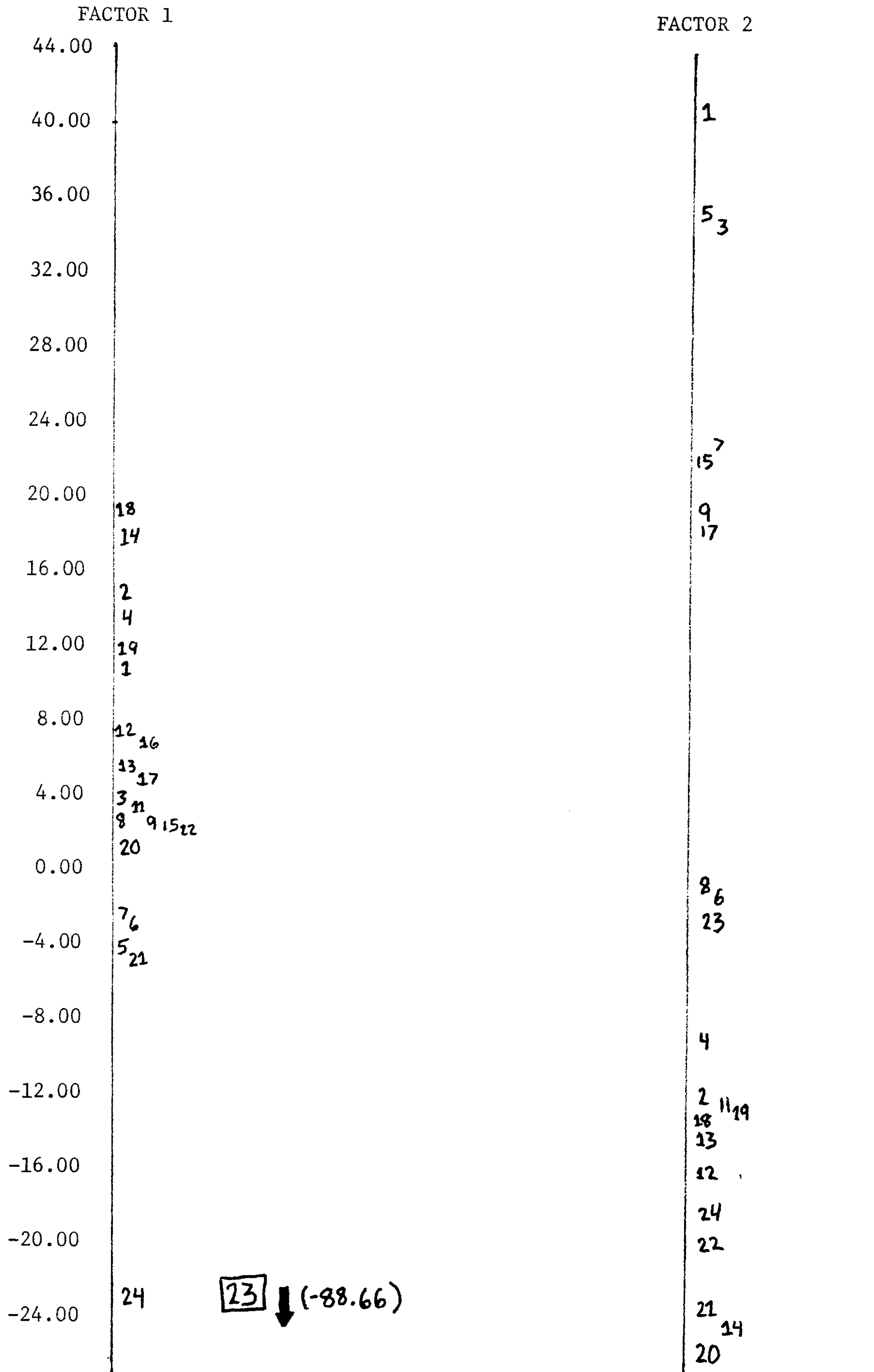


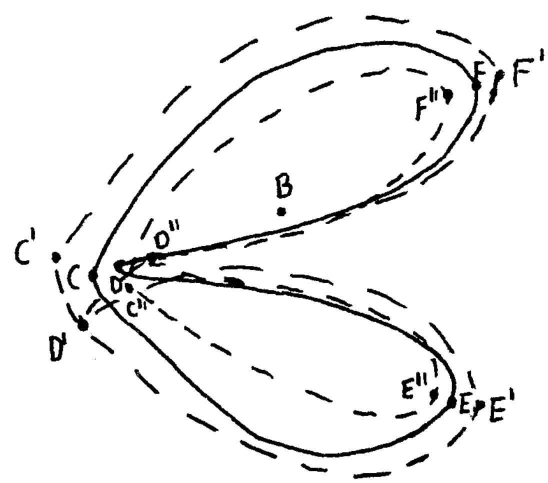
FIGURE 7: Effects of increases and decreases in canonical factor 1. The neutral lip shape is taken from raw data measurements for the Cantonese vowel [ɛ] (scale multiplied by 2). The dashed lines represent deviations from the neutral lip shape.

——— = neutral (C - J)

- - - - = increase in factor 1 (C' - J')

- - - - = decrease in factor 1 (C" - J")

SIDE VIEW OF LIPS



FRONT VIEW OF LIPS

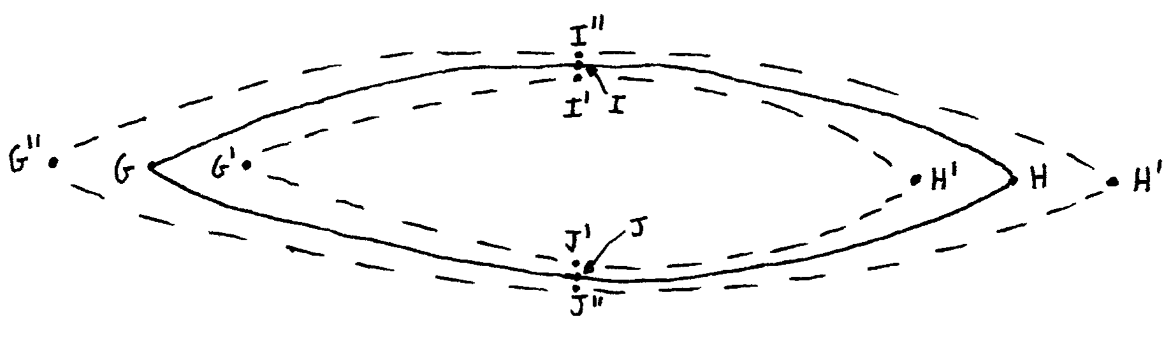


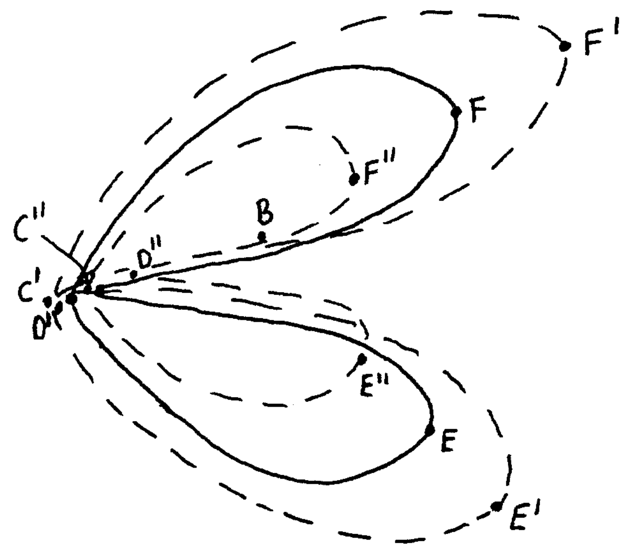
FIGURE 8: Effects of increases and decreases in canonical factor 2. The neutral lip shape is taken from raw data measurements for the Cantonese vowel [ε] (scale multiplied by 2). The dashed lines represent deviations from the neutral lip shape.

——— = neutral (C - J)

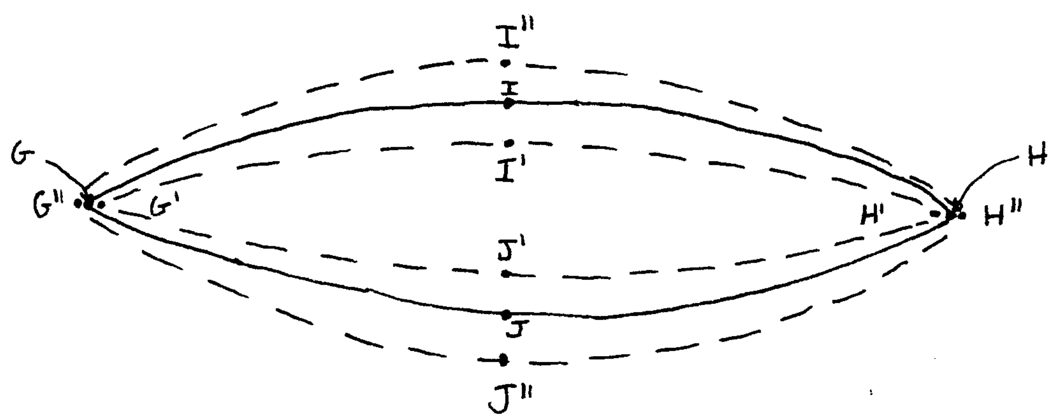
- - - - = increase in factor 2 (C' - J')

- - - - = decrease in factor 2 (C" - C")

SIDE VIEW OF LIPS



FRONT VIEW OF LIPS



increase in the amount of protrusion (from point B to points E and F), combined with a substantial decrease in vertical opening.

Results and Discussion: Similar vowels across languages

We may now turn to the first question raised at the beginning of this investigation; that is, we can ask whether or not the vowels in these four languages differ with respect to their use of canonical factors 1 and 2. In order to answer this question, PARAFAC was run a second time on each of the four sets of measurement data, but this time the M matrix was constrained to be the Canonical M matrix for all languages, as illustrated at the bottom of Figure 5. Note that because of this constraint the original PARAFAC measure loading matrices have been altered, and therefore the product V x S (vowel loadings x speaker loadings) is now also altered to conform to the new common factor space. However, because the data set for each language is different, V x S is still language specific.

The loadings in the vowel space along canonical factor 1 (horizontal opening) were multiplied by the mean speaker loading ($\bar{S}$) for each language. The value of the product vowel loading x $\bar{S}$ can be thought of as an average loading for a given vowel along canonical factor 1, and these loadings can now be compared across languages. They are plotted for each language in Figure 9. Here we observe: (1) in all languages, this factor separates traditionally labelled "rounded" from "unrounded" vowels; as would be expected, rounded vowels show a low degree of horizontal opening while unrounded vowels show a high degree, (2) in all languages we find the subgroupings (a) high, round (b) mid, round (c) unround; in Swedish we find the high rounded front vowel /y/ grouped with the mid rounded vowels, indicating a greater degree of horizontal opening for this vowel (note also that the Swedish [ɑ] is, in fact, a somewhat rounded vowel, and so its location on the scale is not surprising), (3) so-called lax rounded vowels show greater horizontal opening than their tense counterparts, in general, and (4) while the overall ordering of vowels is similar from language to language along canonical factor 1, in some languages vowels within particular subgroups (e.g., high, round) are more closely tied to one another than in other languages. For example, in Finnish, there is a good deal of separation between /u/, /y/ and /ø/, while in French these vowels lie close together on a scale of horizontal opening.

Figure 10 shows the vowel x $\bar{S}$ loadings along canonical factor 2 (protrusion/vertical opening), and we find a pattern similar to the factor 1 vowel space. As with factor 1, rounded and unrounded vowels are separated by factor 2, and in general there is an ordering, at least among rounded vowels, corresponding to vowel height, with so-called lax vowels being further down the scale than their tense counterparts. Although the scale is compressed relative to the factor 1 vowel space, there are even more differences between languages in the relative spacing of vowels within the subgroups "round" and "unround" along the scale of protrusion/vertical opening than we saw with the scale of horizontal opening. The results with regard to Swedish are particularly interesting in this light; note that /ʉ/ and

FIGURE 9: Vowel loadings x $\bar{S}$ from PARAFAC along canonical factor 1 (horizontal opening). $\bar{S}$ = the mean of the speaker loadings for a given language.

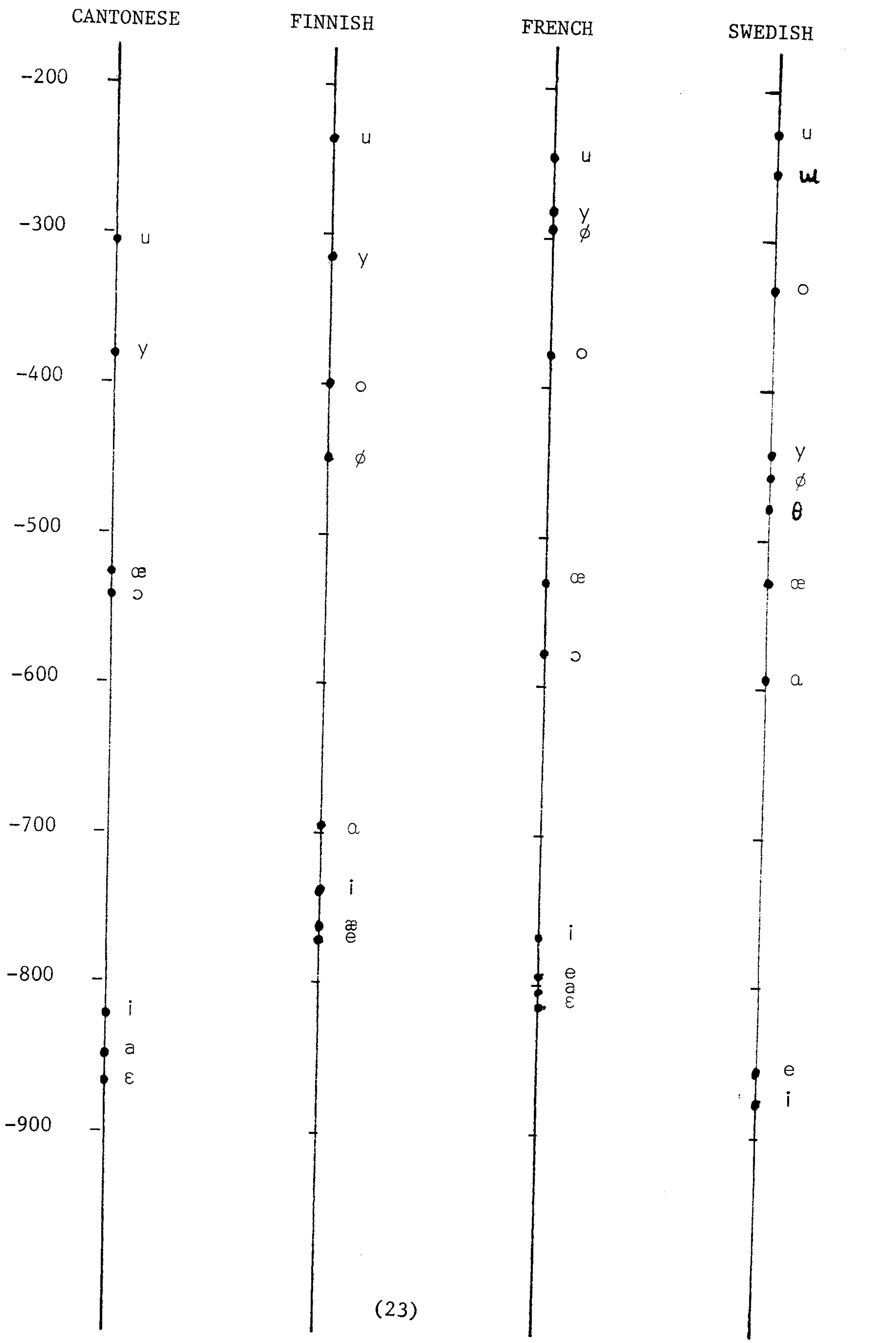


FIGURE 10: Vowel loadings x $\bar{S}$ from PARAFAC along canonical factor 2 (protrusion/vertical opening). $\bar{S}$ = the mean of the speaker loadings for a given language.

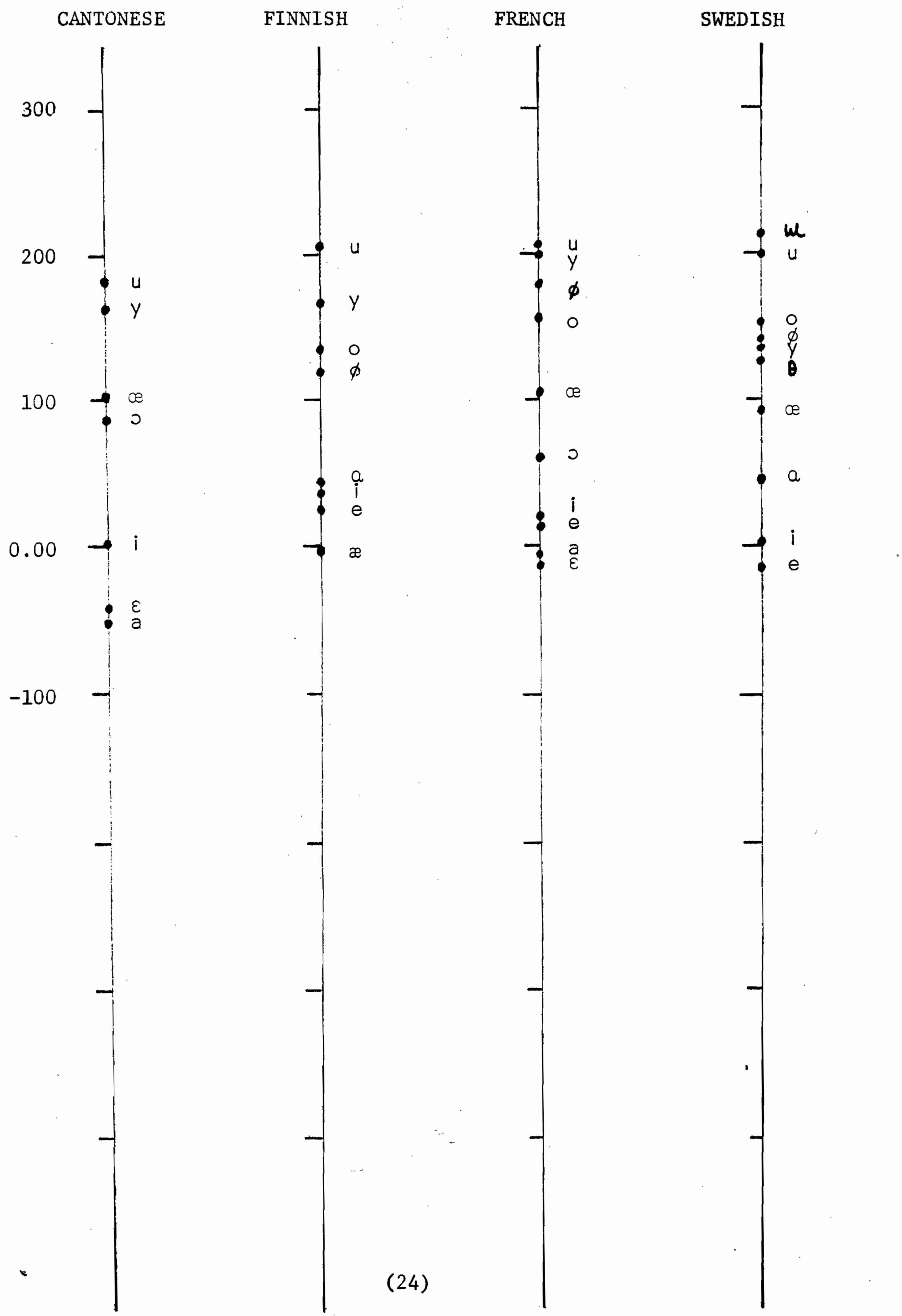


TABLE III: Hierarchical ordering of languages for similar vowels along canonical factor 1 (<u>horizontal</u> <u>opening</u>). The languages are shown in order of decreasing loadings along this factor.

/i/ :	Swedish >> Cantonese >> French > Finnish
/e/ :	Swedish >> French ≃ Finnish
/ɛ/ :	Cantonese >> French
/a/ :	Cantonese >> French
/ɑ/ :	Finnish >> Swedish
/ɔ/ :	French > Cantonese
/o/ :	Finnish > French > Swedish
/u/ :	Cantonese >> French ≃ Finnish ≃ Swedish
/y/ :	Swedish >> Cantonese >> Finnish > French
/ø/ :	Swedish ≃ Finnish >> French
/œ/ :	Swedish ≃ French ≃ Cantonese

>> = difference greater than or equal to 50

\> = difference less than 50 and greater than or equal to 25

≃ = difference less than 25

TABLE IV: Hierarchical ordering of languages for similar vowels along canonical factor 2 (protrusion/vertical opening). The languages are shown in order of decreasing loadings along this factor.

/i/ :	Finnish > French > Swedish ≃ Cantonese
/e/ :	Finnish ≃ French >> Swedish
/ɛ/ :	French >> Cantonese
/a/ :	French >> Cantonese
/ɑ/ :	Swedish ≃ Finnish
/ɔ/ :	Cantonese > French
/o/ :	French ≃ Swedish > Finnish
/u/ :	French ≃ Finnish ≃ Swedish > Cantonese
/y/ :	French >> Finnish ≃ Cantonese >> Swedish
/ø/ :	French >> Swedish > Finnish
/œ/ :	French ≃ Cantonese ≃ Swedish

\>> = difference greater than or equal to 25

\> = difference less than 25 and greater than or equal to 12.5

≃ = difference less than 12.5

/u/ are grouped very close together at one end of the scale (indicating a large value for the ratio protrusion/vertical opening), while the high front rounded vowel /y/ is located much closer to the mid front rounded vowel /ø/. This grouping of /ʉ/ and /u/ together is just what we would expect from the traditional description of inrounding and outrounding in Swedish vowels. (It is interesting to note, however, that one of the Swedish inrounded vowels, which I have transcribed as /ʉ/ as in <u>hus</u> 'house' (which at an earlier stage of Swedish was presumably a back vowel) is synchronically a front vowel (at least acoustically), not a back or mixed vowel -- cf. Sweet's description of inrounding and outrounding). In addition, the more recent work done by McAllister, Lubker and Carlson mentioned earlier has shown that while there is lip movement in the posterior-anterior direction in the production of /ʉ/ and /u/, there is in addition substantial compression of the lips in the superior-inferior (i.e., vertical) direction. Given vertical compression of the lips and hence a very small vertical opening, we would expect to find these two vowels grouped together at an extreme end of the scale; similarly, we would not expect to find the outrounded vowels /y/ and /ø/ to be located at either extreme, but to be grouped together towards the middle of the scale, which is exactly what we observe.

To summarize: we see some similar patterns across languages with regard to separation of rounded and unrounded vowels, and with regard to the relative ordering of vowels within particular subgroups. However, we also see differences relating especially to the relative spacing of vowels within these subgroups, along both factors 1 and 2. Tables III and IV reorganize the data presented in Figures 9 and 10 in terms of hierarchies along the dimensions of horizontal opening and protrusion/ vertical opening.

<u>Analysis</u>: Acoustically identical vowels across languages

We would now like to briefly examine the second question raised at the beginning of this study: that is, we would like to know if the kinds of differences that have been observed among these four vowel systems with respect to canonical factors 1 and 2 would be present if the vowels within these systems were acoustically identical. In order to determine the relationship between acoustic vowel quality and lip position, separate stepwise multiple regression analyses were performed for each language, as shown below:

REGRESSION EQUATION TO PREDICT VOWEL X SPEAKER LOADINGS ($\hat{V}'$ X $\hat{S}'$) ON ONE FACTOR FROM FORMANT FREQUENCY:

$$\hat{V}' \times \hat{S}' = b_1X_1 + b_2X_2 + b_3X_3 + b_4P_1 + b_5P_2 + \ldots + b_{10}P_7 + \text{Constant}$$

where $\hat{V}' \times \hat{S}'$ = predicted product of vowel and speaker loadings on a given factor (dependent variable)

X = independent variable consisting of one of the first three formant frequencies or one of the products and ratios of these formants

P = variable representing variation due to speaker

b = (beta) weighting of a particular variable

Thus, two regression equations were obtained for each language, one per factor. The correlation coefficients (multiple R) ranged between 0.842 and 0.911. The actual regression equations for each language are given in Table V. A set of constant formant frequencies (F1, F2, F3) was introduced into the regression equations, to simulate the vowels /i, e, æ, ɑ, o, u, y, ø/ (data taken from Terbeek, 1977).

The variable P was originally introduced in order to make it possible to later analyze statistically the differences found among the vowel x speaker loadings predicted from formant frequency (i.e., to obtain a scatter plot of predicted vowel x speaker loadings). This is, however, beyond the scope of the present study. Thus, in looking at the results of the regression analyses, the mean beta weight (b) for each variable P was found for each language for each factor, and these constants were used in the equations.

Discussion and Preliminary Results

Even given the mean beta weights, it is unfortunately not clear that one can directly compare the absolute values of the predicted vowel x speaker loadings across languages, as was possible to do with Figures 9 and 10. However, we can still observe differences and similarities in the general patterning of the vowels from one language to another.

Figure 11 shows the vowel x speaker loadings predicted from formant frequency plotted along canonical factor 1, for each language. As can be seen, once again this factor of horizontal opening separates rounded from unrounded vowels. However, this separation is much less distinct in Swedish, where the spacing is relatively even among all vowels within the entire vowel space. There appears to be no common ordering among the unrounded vowels; however, all languages except French order rounded vowels identically along the dimension of horizontal opening. The spacing among the rounded vowels themselves, however, differs from language to language.

Figure 12 shows the predicted vowel x speaker loadings on canonical factor 2 (protrusion/vertical opening). The following observations can be made: (1) all languages show large separation between the unrounded and rounded groups of vowels, (2) no language has the same ordering among the rounded or the unrounded vowels, (3) three languages show even spacing within the rounded group of vowels; the exception is Swedish, where these vowels lie very close together on a dimension of protrusion/vertical opening, and (4) The spacing among the unrounded vowels is relatively even for all languages except Finnish, where these vowels show very little separation at all.

These results indicate that languages differ greatly in the lip gestures they use to make the same acoustic distinctions among vowels. If we look at the ordering of pairs of vowels within the subgroups "rounded" and "unrounded" along both factors 1 and 2, we see that the four languages have only two characteristics in common:

TABLE V: Regression equations for each language. $\bar{b}$ = the mean beta weight for the variable P that extracted speaker variation, and was added to the equations.

CANTONESE

factor 1: (R=0.911)

$$V \times S = -1.039 \times F3 - \frac{0.524 \times F1 \times F2}{10^3} - \frac{5.118 \times 10^6}{F3} + \frac{1.286 \times 10^8}{F1 \times F2} + 4.337 \times 10^3; \quad \bar{b} = -6.605 \times 10^1$$

factor 2: (R=0.890)

$$V \times S = 0.762 \times F1 - 0.256 \times F3 - \frac{0.212 \times F1 \times F2}{10^3} - \frac{2.349 \times 10^3 \times F1}{F3} + 0.735 \times 10^2; \quad \bar{b} = -4.937 \times 10^1$$

FINNISH

factor 1: (R=0.897)

$$V \times S = \frac{-0.412 \times F1 \times F2}{10^3} + \frac{4.468 \times 10^6}{F3} - \frac{2.442 \times 10^{11}}{F1 \times F2 \times F3} + \frac{6.524 \times 10^4 \times F3}{F1 \times F2} - 2.053 \times 10^3; \quad \bar{b} = -2.027 \times 10^2$$

factor 2: (R=0.883)

$$V \times S = \frac{-0.157 \times F1 \times F2}{10^3} + \frac{4.468 \times 10^6}{F3} - \frac{2.442 \times 10^{11}}{F1 \times F2 \times F3} + \frac{2.757 \times 10^4 \times F3}{F1 \times F2} - 4.221 \times 10^2; \quad \bar{b} = -8.915 \times 10^1$$

FRENCH

factor 1: (R=0.891)

$$V \times S = 2.865 \times F1 - 0.528 \times F3 - \frac{1.809 \times F1 \times F2}{10^3} - \frac{5.054 \times 10^6 \times F1}{F2 \times F3} + 1.173 \times 10^3; \quad \bar{b} = 1.284 \times 10^2$$

factor 2: (R=0.885)

$$V \times S = 1.049 \times F1 - 0.195 \times F3 - \frac{0.664 \times F1 \times F2}{10^3} - \frac{1.969 \times 10^6 \times F1}{F2 \times F3} + 7.740 \times 10^2; \quad \bar{b} = 1.767 \times 10^1$$

SWEDISH

factor 1: (R=0.875)

$$V \times S = \frac{-0.515 \times F2 \times F3}{10^4} + \frac{2.586 \times 10^6}{F1 \times F3} - \frac{1.585 \times 10^2 \times F2 \times F3}{F1 \times F3} - 3.611 \times 10^2; \quad \bar{b} = 7.873 \times 10^1$$

factor 2: (R=0.842)

$$V \times S = \frac{-8.91 \times 10^2 \times F1}{F3} + \frac{1.058 \times 10^6}{F3} - \frac{4.233 \times F2 \times F3}{10^4 \times F1} - 1.308 \times 10^2; \quad \bar{b} = -5.513$$

FIGURE 11: Vowel x Speaker loadings predicted from formant frequency along canonical factor 1 (horizontal opening).

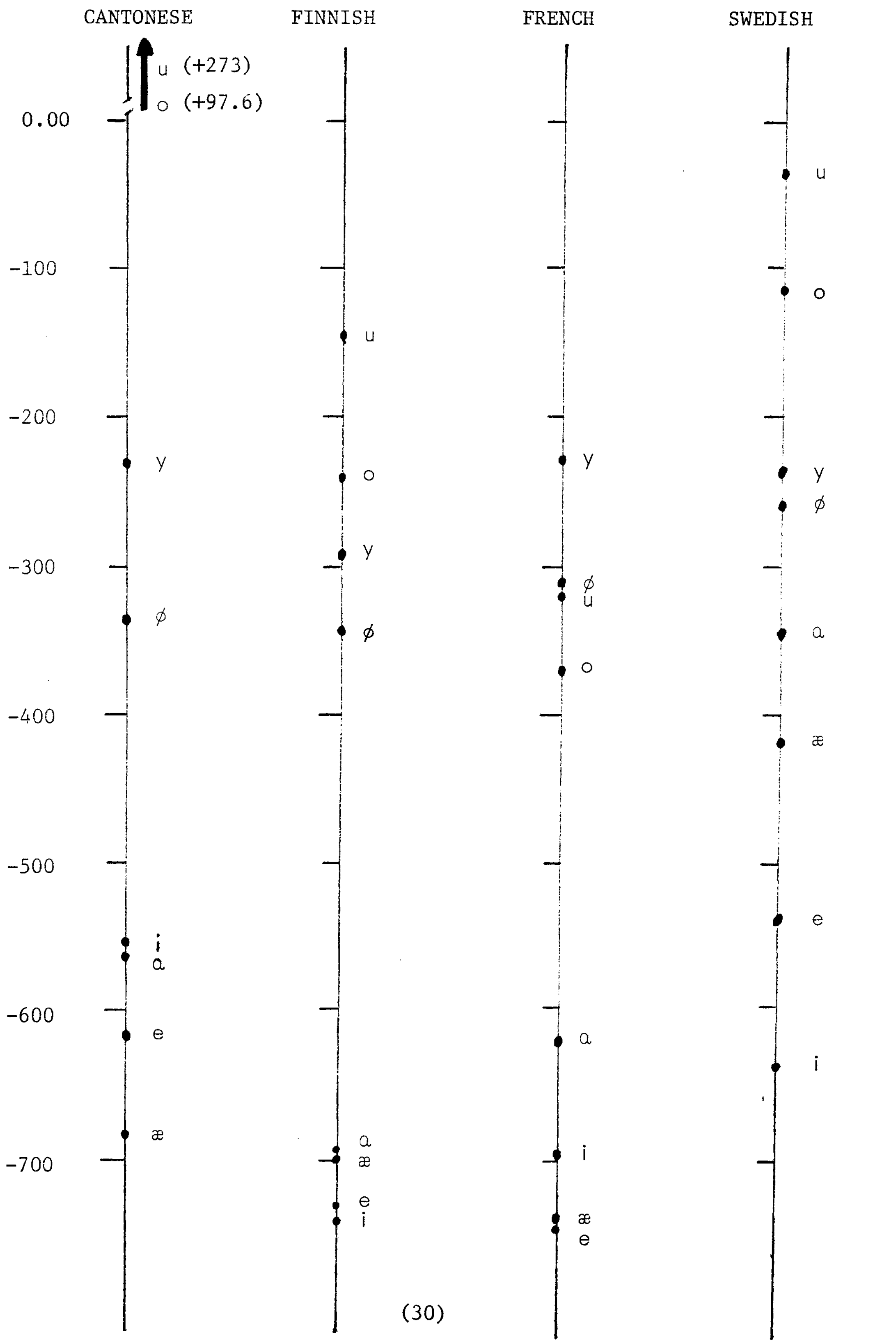


FIGURE 12: Vowel x Speaker loadings predicted from formant frequency along canonical factor 2 (protrusion/vertical opening).

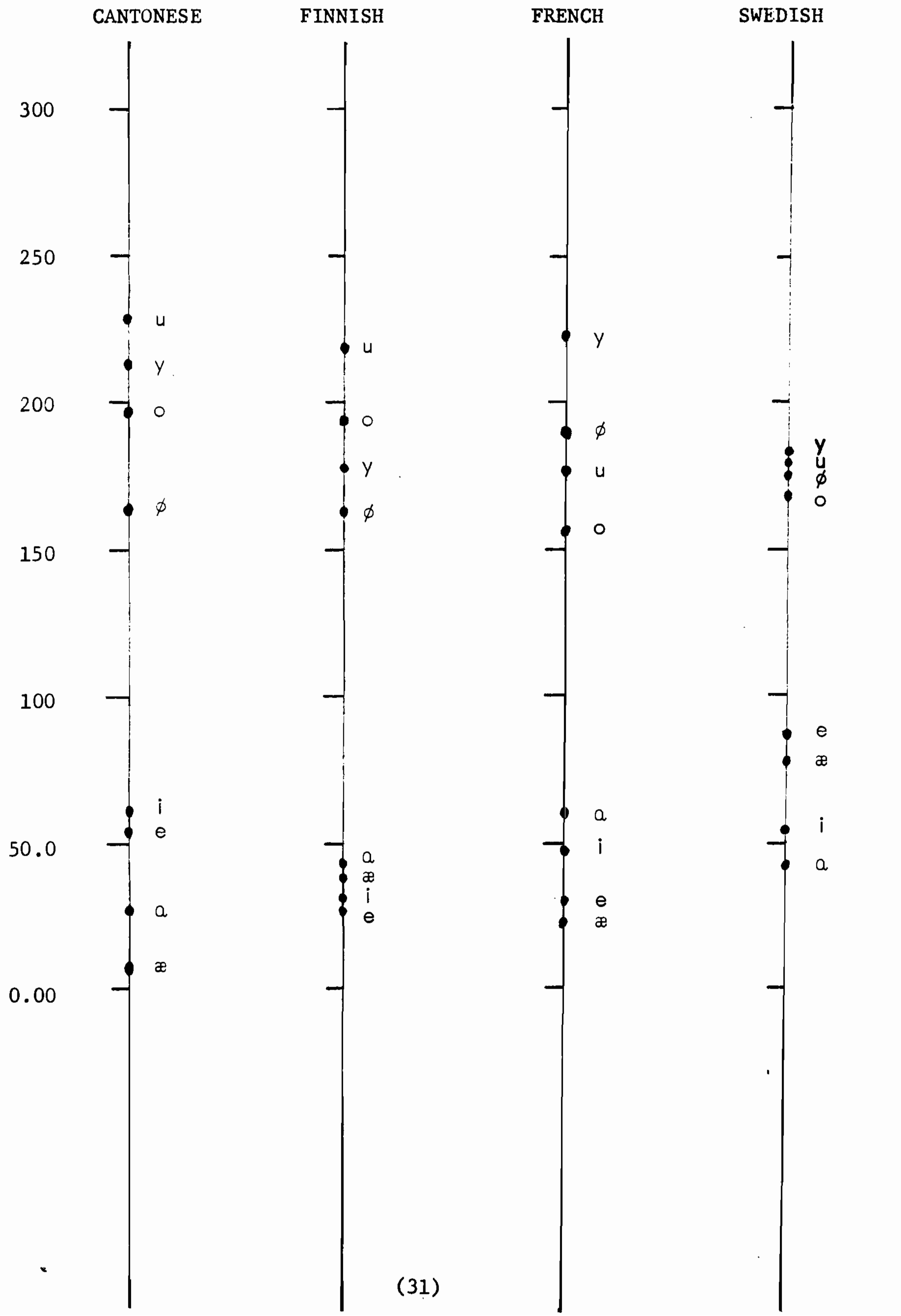


TABLE VI: Relative weightings of factors 1 and 2 used to distinguish pairs of vowels across languages.

+ = relatively higher weighting of a given factor

– = relatively lower weighting of a given factor

		i/e	i/æ	i/ɑ	e/æ	e/ɑ	æ/ɑ	u/y	u/ø	u/o	y/ø	y/o	ø/o
CANTONESE	factor 1	+ –	+ –	+ –	+ –	– +	– +	+ –	+ –	+ –	+ –	– +	– +
	factor 2	+ –	+ –	+ –	+ –	+ –	– +	+ –	+ –	+ –	+ –	+ –	– +
FINNISH	factor 1	– +	– +	– +	– +	– +	– +	+ –	+ –	+ –	+ –	– +	– +
	factor 2	+ –	– +	– +	– +	– +	– +	+ –	+ –	+ –	+ –	– +	– +
FRENCH	factor 1	+ –	+ –	– +	– +	– +	– +	– +	– +	+ –	+ –	+ –	+ –
	factor 2	+ –	+ –	– +	+ –	– +	– +	– +	– +	+ –	+ –	+ –	+ –
SWEDISH	factor 1	– +	– +	– +	– +	– +	– +	+ –	+ –	+ –	+ –	– +	– +
	factor 2	– +	– +	+ –	+ –	+ –	+ –	– +	+ –	+ –	+ –	+ –	+ –

(1) To distinguish between [y] and [ø], all languages use relatively higher values of both factors 1 and 2 for [y], and relatively lower values of both factors 1 and 2 for [ø].

(2) To distinguish between [u] and [o], all languages use relatively higher values of both factors 1 and 2 for [u], and relatively lower values of both factors 1 and 2 for [o].

Otherwise, the relative amounts (weightings) of factors 1 and 2 used to distinguish pairs of vowels differ greatly from language to language. A comparison of these weightings for 12 pairs of vowels is given in Table VI.

IV. Conclusions

Two questions were asked at the beginning of this study: (1) if we compare the lip gestures used in the production of similar vowels across languages, will we find differences in the use of these gestures? and (2) if we find differences in lip positions among similar vowels in different languages, to what extent do these reflect the fact that the vowels in these systems are not acoustically identical?

I have investigated these questions by observing the general patterning of vowels in four languages along two dimensions of lip position: horizontal opening and protrusion/vertical opening. With regard to the first question, we have observed that traditionally labeled "rounded" and "unrounded" vowels are separated by these dimensions, and that there are similar patterns with regard to the relative ordering of vowels within the subgroups "rounded" and "unrounded". However, we have also observed differences relating especially to the relative spacing of vowels within the particular subgroups, and exceptions to the general similarities in ordering within subgroups were found.

As for the second question, we have observed numerous differences in (1) the ordering of vowels within subgroups and in (2) the relative spacing of vowels within subgroups, as well as within the entire vowel space itself. We have also seen that while languages use similar relative weightings of factors 1 and 2 to distinguish between the vowel pairs [y]/[ø] and [u]/[o], in general, languages make use of quite different lip gestures to make the same acoustic distinctions within pairs of vowels. These results tentatively indicate that the relationship between acoustic vowel quality and lip position is, in fact, language- specific, and hence the differences found among similar vowels with respect to the parameters horizontal opening and protrusion/vertical opening are perhaps not soley due to the fact that the vowels within these systems are different acoustically. However, a closer investigation of the relationship between vowel quality and lip position in these languages is needed before an explanation can be provided for the kinds of differences that have been observed among the acoustically identical vowels.

Appendix I: Word Lists

1. Cantonese (all words on mid level tone)

si:	ku:
sɛ:	sy:
sa:	kœ:
sɔ:	

2. Finnish

si:	so:
se:	su:
sæ:	sy:
sɑ:	sø:

3. French

si	so
se	su
sɛl	sy
sa	sø
sɔl	sœl

4. Swedish

hi:t	hɯ:t
he:t	hɵ t:
hɑ:t	hy:
ho:	hø:
hu:t	hœt:

Note: the vowel [ɯ] is sometimes transcribed as [ʉ]; however, [ɯ] is a front, rather than a central vowel.

REFERENCES

Abry, C., L.J. Boë, M. Gentil, R. Descout and P. Graillot. 1979. La geometrie des levres en francais - protrusion vocalique et protrusion consonantique. 10èmes Journées D'Étude sur la Parole.

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

Fromkin, V. 1964. Lip positions in American English vowels. Language and Speech 7, 215-225.

Fromkin, V. 1965. Some phonetic specifications of linguistic units: an electromyographic investigation. UCLA Working Papers in Phonetics 3.

Fujimura, O. 1961. Bilabial stops and nasal consonants: a motion picture study and its acoustical implications. Journal of Speech and Hearing Research 4, 233-247.

Harshman, R. 1970. Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics 16.

Harshman, R., P. Ladefoged and L. Goldstein. 1977. Factor analysis of tongue shapes. JASA 62, 693-707.

Henderson, E.J.A. 1971. A Selection from the Writings of Henry Sweet. Oxford Univ. Press. London.

Jones, D. 1960. An Outline of English Phonetics. 9th ed. Cambridge.

Ladefoged, P., R. Harshman, L. Goldstein and L. Rice. 1978. Generating vocal tract shapes from formant frequencies. JASA 64, 1027-1035.

Linker, W. 1978. Lip position and formant frequency in American English vowels. UCLA Working Papers in Phonetics 41, 20-25.

McAllister, R. 1978. Some phonetic correlates of the tense-lax feature in Swedish rounded vowels. Journal of Phonetics (forthcoming).

McAllister, R., J. Lubker and J. Carlson. 1974. An EMG study of some characteristics of the Swedish rounded vowels. Journal of Phonetics 2, 267-278.

Stevens, K.N. and A.S. House. 1955. Development of a quantitative description of vowel articulation. JASA 27, 484-493.

Terbeek, D. 1977. A cross-language multidimensional scaling study of vowel perception. UCLA Working Papers in Phonetics 37.

Why Phonology isn't "Natural"*

Stephen R. Anderson
U.C.L.A.

0. Introduction. Phonological research in recent years has been dominated by the notion that the theories with greatest *a priori* plausibility would be ones that were in some sense *natural*. Of course, the senses that have been given to the notion of naturalness have been many and varied, and by no means always mutually compatible. They range from the evident (but not immediately very helpful) fact that anything that occurs in nature is natural to much more precise claims about detailed aspects of phonological description.

The most persuasive appeals to naturalness, however, have been based on the claim that the properties of a linguistic theory ought to follow in some way from the independently established properties of other systems within which language is embedded. Thus, Donegan & Stampe (1979:126) develop the "basic thesis [...] that the living sound patterns of languages, in their development in each individual as well as in their evolution over the centuries, are governed by forces implicit in human vocalization and perception." Such a premise, in their view, leads to "a *natural* theory [...] in that it presents language (specifically the phonological aspect of language) as a natural reflection of the needs, capacities, and world of its users [...] It is a natural theory also in the sense that it is intended to *explain* its subject matter, to show that it follows naturally from the nature of things" (*Ibid*, p. 127). A consequence of this view of course, as Donegan & Stampe emphasize, is the claim that the subject matter of linguistics (and of phonology, in particular) consists exactly of those aspects of language which are amenable to the establishment of connections with other domains. "In the case of natural phonology this means everything that language owes to the fact that it is *spoken*." (*Ibid*, p. 128)

In one form or another, this view that phonological constructs (rules, representations, organizing principles of grammars, and the like) ought to have fairly direct interpretations (or "explanations") in terms that are not limited to the structural properties of language alone has directed much research and determined the consensus of the field with regard to many theoretical issues in the years since the publication of Chomsky & Halle, 1968. Connections with such domains as acoustic phonetics, articulatory physiology, the study of memory and cognition, general principles of concept formation, and the like have been adduced (or required) in support both of general principles and of analytic accounts of particular languages.

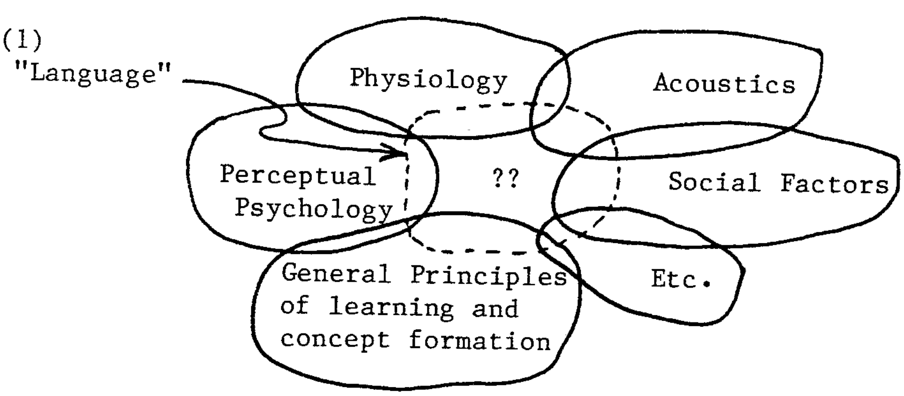
In addition, it has been suggested that the "psychological reality" of a phonological description ought *necessarily* to be susceptible to confirmation through experimental tests involving general psychological tasks not specific to normal language mastery and use. Some indeed have gone so far as to suggest that any aspect of a description which cannot be given such a foundation in another domain, or at least validated through observations that do not rely specifically on the system of the language, is simply metaphysics, and inadmissable in the scientific study of language.

One way of assessing this foundational issue in the study of language is the following. Essentially all linguists would agree that the totality of human verbal activity ("Language" in the widest possible sense) can be approached as resulting from the interaction of a number of conceptually distinct domains. Contributions are made to determining the observable shape of various aspects of Language by such features of the world as the physiology of the human vocal tract, the physical acoustics of sound transmission in air, structural properties of the auditory system, the nature of (and limitations on) memory, general notions of well-formedness of concepts, principles of inductive learning and generalization, and many others.

Now the interesting property which these domains have in common, from our point of view, is that they can all be studied (in principle) independently of one another, and to a great extent independently of the particular problem of their role in determining the nature of Language. We can study the shape and musculature of the vocal tract, for example, without needing to know that it is used to make sounds as well as to eat and to breathe; we can study our memory of all sorts of emotions, impressions, tasks, concepts and structures other than just those whose form is specifically given in a language, etc. It is in fact the results of such general, non-language-dependent studies that we refer to when we speak of the contributions made to our understanding of Language by psychologists, physiologists, physicists, and others.

The question of the "naturalness" of phonology can then be posed in the following (somewhat untraditional) way: after we have "factored out" or otherwise taken account of whatever in Language can be adequately treated as aspects of more general problems, not specific to Language per se, how much is left over? In terms of the diagram in figure (1), that is, assuming that the total area inside the dotted inner circle represents "Language" in the broad sense, while that contained in each of the other, partially overlapping circles represents facts specific to some general domain, how much (if anything) is left in the middle (the boolean intersection of the properties of Language and the complement-sets of all of the other relevant domains)?

(1)



On the view that phonology ought to be "natural" in the sense suggested above, we should expect a particular answer to this question: as far as the scientific study of Language is concerned, nothing at all remains in the area of "??". Many linguists - and by no means only those of this century - have pursued the possibility that virtually all aspects of Language fall within one or another of the solid circles in diagram (1): that is, they are accessible from one or another more general perspective (as special cases of the functioning of the upper respiratory tract, pragmatic strategies for successful social interaction, a generalized learning strategy, etc.).

These researchers put a great deal of faith in the possibility of experimental validation of all of the details of linguistic structure. If it is not possible, given some proposed aspect of the structure of Language in general or of a particular language, either a) to show that it follows directly from the other properties of some more general system; or b) to devise some language-external psychological task to verify it, this is sometimes felt to establish a case <u>ex silentio</u> against the proposed analysis. Clearly, of course, if all aspects of Language (or at least those that are scientifically interesting) are simply special cases of more general capacities, then it ought to be possible to find approaches to them from outside Language <u>per se</u>.

Other linguists, however, have taken a rather different view. Within the tradition that sees Language as a unique human capacity (or "mental organ" in the terms of Chomsky, 1980 and elsewhere), it is possible to suggest that there are at least some components of this capacity that are particular to Language itself, and hence not necessarily studiable or explicable directly as special cases of other systems. Among the intersecting domains that contribute to the facts and properties of "Language", that is, we may find a particular language faculty which is not reducible to features of other kinds (at least, within the limits of present knowledge in such areas as neurology, brain chemistry, the genetic control of development, etc.). It is exactly this area, one might claim, that ought to occupy the central concern of linguists if they wish to arrive at an adequate conception of the essential and special nature of human Language.

The validation of such a claim - that there is some essentially <u>linguistic</u> faculty not necessarily observable or even deducible purely in extra-linguistic terms - would clearly have to be indirect. It would have to rest not on the establishment of direct links between the posited language faculty and some set of observation statements, but rather on an inference: insofar as the properties of the proposed system achieve some measure of success in facilitating and revealing regular connections among phenomena, and bring orderliness and coherence to our understanding of Language, we are justified in attributing these properties to the system of language.

As long as we understand the terms of validation of hypotheses about specifically linguistic structure in this way, it is clearly an empirical issue to determine whether or not a uniquely linguistic faculty should be attributed to the human mind. Nonetheless, when the empirical issue is posed as it most often is - namely, as whether we ought to attribute "existence" to a structure which may in principle not be susceptible to observational confirmation - it seems to be a matter of purest mysticism, and those who would maintain its truth are sometimes seen as speculative metaphysicians in the great tradition.

Rather than posing the problem of the "existence" of theoretical entities in such strictly literal terms, however, it is perhaps better to contrast the two views above not directly as an empirical issue, but rather as a choice between research strategies: should we or should we not limit the terms and constructs of linguistic theory to elements that can be given an extra-linguistic foundation? When the question is put in this way, the burden of proof appears to shift. Indeed, it seems thoroughly mystical to maintain that every aspect of language has an explanatory basis in some other domain, given our rather limited success in actually finding secure explanations of this sort for any aspects of language. If we confine our attention to those features of Language (and languages) for which we have (or can plausibly claim to be close to) exhaustive accounts from an extra-linguistic perspective, we will be reduced to a remarkably small class of objects of study (despite optimistic claims to the contrary on the part of some proponents of "natural" phonologies who have recognized this problem).

We are clearly unlikely to establish one or the other of the two possible research programs suggested above in an absolutely firm and conclusive way: too much inevitably remains speculative and programmatic on either view. The best we can hope to do is to sharpen the issues between them to a point where it is possible to ask which appears to give the greater hope of success in understanding the nature of Language.

When we ask the general question of whether we should proceed on the assumption that linguistics must be "natural" in the sense of being entirely reducible to phenomena that can be observed and explained from outside the system itself, there are a number of ways in which we can make the issues clearer within the domain of phonology. Some of these are:

a) Should a descriptive framework for phonologicl analysis (primarily, a universal system of distinctive features) be limited to a set associated directly with measurable acoustic, articulatory, and/or auditory properties? Or might it also include some properties for which the evidence is sometimes (or perhaps always) indirect or inferential rather than observational?

b) Should the set of rules admitted in the grammars of particular languages be limited to formal statements of surface regularities that have a demonstrable phonetic basis ("natural rules")? Or might grammars contain as a central part rules that are not grounded (synchronically, at least) in such phonetic explanations?

c) Should the relationship between underlying and surface representations of linguistic elements be limited to ones which can be directly induced from the corpus of forms and alternations available to the child, and thus "learnable" in some general sense? Or might phonological theory admit of more abstract representations, justified inferentially or by reference to other sorts of evidence?

On a rigorously naturalist view of phonological structure, particular choices are indicated in each of these cases. The first alternative presented under each of a) and b) above would ensure that the inventory of phonological parameters and rules could be reduced to the independently studiable properties of the articulatory, acoustic, and/or auditory systems both in form and in content; an affirmative answer to question c) would ensure that the relation between underlying and surface forms followed more or less directly from principles of inductive learning. Constraints on (and thus essential properties of) the system of language would then be shown to be reducible to those of other, not specifically linguistic, domains; and the research program that required such reduction by denying the existence of specifically linguistic structure and capacities would be made more plausible.

The contention of this paper, on the other hand, is that none of the three choices posed above can be limited to the first alternative given without limiting the scope of enquiry in linguistics in arbitrary and undesirable ways. In each of the areas considered here (the character of representations, the substantive content of rules, and the nature of the relation between phonological and phonetic representations), it will be argued below that it is necessary to assume a considerable degree of independence between linguistic principles proper and the principles that obtain in those extralinguistic domains that appear to underlie them. The basis of this independence, it will be claimed, is the fact that as an autonomous cognitive faculty (insofar as our current understanding of various ancillary areas allows us to see), the "mental organ" central to our knowledge of a language (and by extension, to language learning or the development of such knowledge) is subject to its own organizing principles that cannot be reduced entirely to other domains. Of course, it would seem to follow that if there <u>are</u> such essentially linguistic principles, it is the primary business of linguists to uncover and explore them - rather than to attempt <u>a priori</u> to rule out theoretical accounts that are not based on "natural" (i.e., extralinguistic) principles.

Of course, if we wish to find and explore a cognitive faculty essential to the nature of Language, we have to have a way of looking for it. We often have no way of knowing <u>a priori</u> with respect to a given linguistic fact whether or not it is to be attributed to some structure or capacity of man or his world that is not specifically linguistic. This is because many such extra-linguistic facts obviously <u>do</u> play a role in Language, and also because it is entirely possible that some formal properties of the language system (or of particular languages) may bear a significant resemblance to the properties of some such other domain without yet being exhaustively attributable to it. The situation here is somewhat analogous to that in which we might be placed on having the opportunity for the first time to watch television. If the figures on the screen appear to exhibit the properties of people engaged in (relatively!) ordinary human activity, we might be tempted to suggest that a television is really a box in which miniature people are imprisoned, and thus that the principles of the observed image can be reduced to the general ones of human behavior and of objects in the real world. It is only when we see images that are inconsistent with the normal behavior of humans and objects in the real world that we are forced to reevaluate this position, and to conclude that the organizing properties of a television image may be quite different from those of the behavior of small people imprisoned in a box - though often the observed results give us little if any grounds for distinguishing the two accounts.

Pursuing this point, we see that the sort of fact which would lead us to reassess our first, naive view of the nature of television would be images that could not be reconciled with or explained by it. Similarly, if we wanted to study organizing principles proper to television, it would presumably be this sort of fact we would concentrate on, since as long as we are watching a picture of people doing normal human things, we have no real way of being sure whether what we see at a given point is due to the particular nature of the television system, or to the extrinsic properties of the people or other objects being photographed.

Analogously, in linguistics, if we want to find and explore principles essential to language, we have to look for facts whose nature would appear to demand an appeal to the linguistic faculty. The principle involved is actually one which is familiar from historical linguistics. In attempting to reconstruct an ancestral system for a family of languages, the most valuable data are forms that are arbitrary or unmotivated from the point of view of the synchronic systems of the daughter languages. Whatever can be given an explanation internal to these latter is of little value for the purposes of reconstruction, precisely because we do not _need_ to attribute it to the ancestral proto-language in order to account for its appearance. It is exactly what does not have such a synchronic basis that we attribute (_faute de mieux_) to earlier forms of the language.

Similarly, in considering areas of phonological structure as mentioned above, we can only determine that some property is to be attributed to the essential nature of language if it does not seem to have an account in more general terms. Of course, it might well be that some essentially linguistic structure might "mimic" properties of another domain; indeed, there are good reasons to expect this, as we will note below. Nonetheless, if we wish to uncover specifically linguistic properties, we must look for those which appear inexplicable in other terms. On this view, it is still very much part of the business of phonologists to look for "phonetic explanations" of phonological phenomena, but not in order to justify the traditional hope that all phenomena of interest can be exhaustively reduced in this way. Rather, just as when syntacticians look for pragmatic accounts of aspects of sentence structure, the reason is to determine what sorts of facts the linguistic system proper is _not_ responsible for: to isolate the core of features whose arbitrariness from other points of view makes them a secure basis for assessing the properties of the language faculty itself.

The sections below will consider each of the questions posed above, and attempt to show that there are linguistic facts which motivate a negative answer in each case. As a consequence of the conclusion that much of linguistic structure cannot in fact be reduced to the study of other properties of the mind or the world (at least in the present state of our knowledge), the plausibility of a research program that posits distinctively linguistic entities, structures, and principles will be enhanced. On such a program, we can eliminate from the study of what is natural in phonology much of the _a priori_ flavor which often characterizes such research today and leads to a degree of arbitrariness and sterility[1]. In fact, we are left with the task of discovering not what can be reduced to other fields, but what occurs in nature.

1. _The nature of distinctive features_. Probably the most basic notion in phonology is that of a _representation_ of an utterance, and (given the program of phonology since at least the appearance of Chomsky & Halle, 1965) the central consideration in determining the character of these representations is the choice of a set of (distinctive) features. In determining the set of features to be recognized by linguistic theory, there are a number of considerations that must be balanced against one another.

Within the history of this issue, the positions that have been taken can be differentiated on the basis of the role assigned to directly observational phonetic properties in the phonological system. A fully abstract view, like that of Hjelmslev, takes the line that only phonological information (basically, data about contrastive function) can play any role in the choice of distinctive elements, and these should not even (in principle) have a direct interpretation in terms of articulatory, acoustic, etc. mechanisms. At the opposite extreme, a fully concrete view like that of Lieberman (1970) appears to require that the properties of the articulatory and auditory system be derived first, without necessary reference to anything beyond the inventory of muscles or muscle groups, property detectors in the peripheral auditory system, etc.; and then if some one of these characteristics matches a linguistic difference, it is available for use as a feature. Most linguists have taken a position somewhere between these two, admitting both "phonological" and "phonetic" evidence in arriving at an inventory of features. The balance between the two sorts of considerations, however, has often been a rather uneasy one.

A "natural" theory in the sense discussed in the preceding section would necessarily take a rather concrete line on this issue. If everything in language is rooted in other domains, it must be the case that the features available for use could only be some subset of an independently established inventory taken from other domains, perhaps as proposed by Lieberman. On the other hand, if the system of language is (in part) based on a unique language faculty, it might be the case that at least some of the distinctions admitted by the system for phonological purposes fail to correspond directly to any measurable phonetic property from another domain. As we will see below, this might turn out to be the case even in a theory in which most of the features are provided with rather direct correlates in an articulatory or auditory framework.

In the sections below, we consider the relation between the external observables of phonetics and considerations internal to that system of a language which we call phonology. We ask first whether a system of features can provide an adequate basis for the description of actual sound patterns if it has the property that all distinctions it makes correspond to overt phonetic differences; we then go on to inquire whether a feature system based on phonological properties can be expected to serve as the basis for a direct interpretation in phonetic, externally measurable terms.

1.1 <u>The phonological adequacy of a "phonetic" feature set</u>. When we ask whether all of the distinctions provided by a feature system should have observable phonetic correlates, the issue involved is illustrated by the problem of labiovelar (and potentially other multiply-articulated) consonants. Following Chomsky & Halle (1968), arguments were given by Anderson (1976a) to the effect that these elements (such as the stops [k͡p, g͡b] and the semi-vowel [w]) ought to be provided with two distinct descriptions by an adequate feature system: either as labialized velar consonants, or as velarized labials. Despite the fact that the two descriptions differ in (at least) one feature, however, there is no necessary phonetic difference between them. The evidence that leads us to classify a given segment containing more than one articulatory constriction of (approximately) the same degree in one way or the other need not be measurable phonetic evidence of any kind, but rather may be entirely phonological, arising out of the behavior of the segment involved in the sound pattern of the language. The evidence provided for this assertion was basically of two sorts.

On the one hand, it was suggested that consideration of the phonological inventories of various languages led to a more natural treatment of labio-velars in one way in some languages, and in the opposite way in others. This is, in essence, an appeal to the "hole in the pattern" arguments familiar from structural linguistics. When we note that Temne, e.g., has a /k/ but no /g/, and a /g͡b/ but no /k͡p/, we are tempted to assign /g͡b/ to the velar series (with concomitant labialization); while the fact that Efik has both a /k/ and a /kʷ/ (in addition to its /k͡p/) but no /p/ leads us to suggest that its /k͡p/ ought in fact to be treated as a (velarized) labial.

On the other hand, it was also suggested that the phonological behavior of labiovelars with respect to phonological rules differs from language to language. In Kpelle, for example (among others), nasal consonants preceding labio-velars assimilate to these, becoming velars -thus suggesting that the following segments themselves are primarily velars. In a number of other languages, however, labiovelar segments (particularly [w]) lose their velar component (or shift it to some position further forward) before certain (typically, front) vowels - suggesting that the labial component of these segments is somehow more basic. Again, the conclusion reached was that these phonological differences between languages corresponded to a difference in the character of their labio-velar consonants, but not, apparently, to one with measurable phonetic consequences.

Ohala, in various places (cf. Ohala & Lorentz, 1977; Ohala, 1979a,b) has quite correctly criticized the nature and quality of the evidence that has been presented for this proposition concerning the analysis of labiovelars. He notes (cf. Ohala & Lorentz, op. cit.) that the "hole in the pattern" argument is either unhelpful or actually suspect in a great many cases. For instance, there is no apparent statistical tendency for labiovelars to appear predominantly in languages where there is a motivated gap (either in the labial or in the velar series) for them to fill. Furthermore, in some languages, such as Efik (cited above), the "hole in the pattern" argument suggests one conclusion (here, that /k͡p/ is a labial), while the argument from phonological rules (in this case, nasal assimilation) leads in the opposite direction (in Efik, to the conclusion that /k͡p/ is primarily a velar consonant).

Ohala is clearly correct in maintaining that arguments from the internal structure of segment inventories can at best be suggestive, and cannot possibly support the weight of a conclusion as fundamental as that concerning the dual phonological character of phonetically identical labiovelar segments. There may well be some tendency for languages to deploy relatively symmetrical segment inventories (for instance, by exploiting an extant contrast before introducing a new one to characterize a new segment), but the undoubted existence of severely assymmetrical inventories makes it clear that this is at best a tendency, of little evidential value.

In fact, the seductive character of this type of argument within the framework of Chomsky & Halle (1968) arises from the observation that the system of features, arrived at without regard to the problem of multiply-articulated segments, in fact provides just the two possibilities considered above for these latter. Both labial and velar segments exist, and both labialization and velarization in various degrees exist; it ought to follow that maximally velarized labials and maximally labialized velars exist, and these converge on the same phonetic segment type. Given that result, which implies that in many particular cases it will be extremely difficult if not impossible to determine the distinctive character of a given sound from strictly external facts, one is naturally tempted to overvalue the weight of what evidence does present itself: here, the tendency to symmetry in phonological inventories. The difficulty of making the decision in particular cases, however, does not in itself convert such tendencies into rigorous predictors.

Ohala concludes that the essential lack of evidence from the character of segment inventories for a dichotomy in phonetically identical labiovelars is itself evidence against the feature system that predicted such a distinction. On his view, "labiovelar" is simply another value of the traditional parameter "point of articulation", and there is no reason to characterize these segments as either labials or velars (to the exclusion, in each case, of the other). Naturally, on this view, no such decision is necessary (or possible): perhaps a desireable result, if it can be maintained, since it restricts the class of possible segment types displayed by natural languages.

Ohala further supports his claim about the unitary (rather than ambiguous) character of labiovelars by arguing that their phonological properties can all be accounted for by reference to phonetic properties alone. He notes, for example, that (regardless of the thrust of other potential evidence in a given language), labiovelars essentially always behave as velars for the purpose of nasal assimilation. There is, he observes, a straightforward explanation for this which makes appeal only to acoustic factors, and is quite independent of the phonological status of the segments involved. Consider the diagrams in figure (2) below (adapted from Ohala, 1979a):

(2) a. [m] b. [ŋ] c. [m͡ŋ]

These diagrams represent the acoustically relevant articulatory properties of nasals at labial, velar, and labiovelar position, respectively. In each case, we find two relevant cavities. One of these, the mouth, is a closed cavity that contributes to the vocal tract transfer function an antiresonance whose center frequency is determined by its size. In the labial case, the closure at the lips makes the closed cavity relatively large. In the case of velars, on the other hand, the closure at this position virtually removes the closed second cavity from acoustic relevance. When we now consider the case of labiovelars (e.g., (2c)), we see that the velar closure has the acoustic consequence that the labial one is essentially irrelevant, at least during the articulation of the nasal itself. Its presence might be revealed during the transition between the consonant and an adjacent vowel, but this is perhaps a less prominent cue for nasals than for stop consonants[2]. In consequence, the difference between a purely velar nasal and a labiovelar one is practically inaudible, and it is entirely reasonable to expect that the difference between the two would be neutralized as far as the output of a comparatively "phonetic" assimilatory process like nasal-assimilation.

Explanations in a similar vein, from either auditory or articulatory properties of the sounds involved, are provided for a number of other features of the phonological behavior of labiovelars. While there are some problems with the degree of predictive force of these arguments, they are generally persuasive (at least in outline) with respect to the facts they attempt to explain, and we can conclude that a number of apparently phonological properties of these segments might be accounted for in phonetic terms alone. From this, Ohala asserts that it follows that it is unnecessary (and probably misleading) to talk about the phonological identity of labiovelars apart from their phonetic content: "the behavior of speech sounds is better understood by reference to system-external factors than system-internal factors." (Ohala, 1979a:46).

Accepting the (general) validity of Ohala's account of particular cases, there is an alternative conclusion possible along the lines suggested in the previous section. We can allow, for example, that the behavior of labiovelars in nasal-assimilation processes may be explicable in terms that are not specifically linguistic: what this shows however, is that facts of this sort may be irrelevant to determining the character of the specifically linguistic phonological system. Of course, if it were demonstrable that <u>all</u> of the properties of the behavior of all speech sounds had this character, we might conclude that in this domain at least, no such system need be posited at all. By contrast, if we found some properties that were apparently not susceptible to such an explanation the import of Ohala's argument would be not to eliminate appeal to this system, but rather to sharpen it by clarifying, along lines already sketched, just what sort of evidence is necessarily revealing of its character.

In this case, then, we would conclude that the behavior of nasals consonants under assimilation before [k͡p] (or other labiovelars) is not, as had been claimed, criterial for determining the phonological behavior of these segments - exactly because there is an account of this behavior that is essentially external to the proper domain of phonology. What is relevant, thus, is not a fact like that of nasal assimilation in Yoruba, but rather some system in which such an externally based account is not available.

In fact, however, such examples have been cited in the literature. Anderson (1976b) cites in this connection the system of consonantal alternations in Fula, a West Atlantic language. In this language, underlying initial continuant consonants in stems are replaced by the corresponding stops or prenasalized consonants under specifiable conditions[3]. What is of interest to us in this example is the behavior of stems beginning with /w/ (the only labiovelar consonant in this language). In fact, some such stems show an alternation with labial consonants, while others show alternations with velars:

(3) a. war "kill" / (bar)[4]/ mbar
    b. war "come" / (gar) / ŋgar

Such stems as that of (3a) show an absolutely consistent alternation with labial consonants in all related forms, derivational and inflectional.

Stems like (3b), in contrast, show equally consistent alternations with velars in all cases. We can also note that no other consonant in the Fula system has the same property: namely, that of alternating with stops and prenasals at different points of articulation, depending on the root itself and not on the phonological environment of the segment affected. With the single exception of stems beginning with basic /w/, in fact, the gradation system involved is quite straightforward as concerns the question of what alternates with what.

On the basis of these facts, then, we would like to distinguish between two classes of stems beginning with /w/ - those that alternate with labials and those that alternate with velars. While it is always possible that the distinction is simply an arbitrary morphological one (this "solution" is of course available for any alternation at all), the internal coherence of the gradation system, combined with the range of positions in which a /w/ can alternate with one or the other of the two stops (cf. Anderson, 1976b for details and discussion) suggest that it is not an arbitrary morphological property of whole stems that is involved, but rather an idiosyncratic distinction between two sorts of /w/: call them $/w/_1$ and $/w/_2$.

What is noteworthy about the difference between $/w/_1$ and $/w/_2$ is the fact that it is exactly a <u>labial-velar</u> consonant which is involved. That is, the consonant type in question is one which involves articulations both in the labial and in the velar regions - and some instances of the segment alternate (exclusively) with labials, while others alternate with velars. The relation between these facts and the phonetic character of the segment is surely not accidental.

Of course, the feature system of Chomsky & Halle (1968), as argued by them and by Anderson (1976a) provides exactly the distinction which is required: $/w/_1$ (which alternates with labials) can be characterized as a velarized labial, while $/w/_2$ (which alternates with velars) can be treated as a labialized velar. The two segments thus share all features except (on one interpretation of the system) [$\pm$Anterior]: $/w/_1$ is [+Anterior], while $/w/_2$ is [-Anterior]. As noted above, however, these two segment types may well converge on exactly the same articulatory configuration. In fact, this is precisely the case in Fula: there is no measurable phonetic difference between instances of $/w/_1$ and $/w/_2$.

If we require that all differences in feature characterization correspond to observable phonetic distinctions, the description of Fula labiovelar /w/ which is directly motivated by phonological considerations would not be available to us. This fact suggests that the abstract phonological character of a sound, its role in the systematic sound pattern of a language, may not in general be reducible to its phonetic content alone, contrary to what Ohala suggests. Of course, in order to arrive at this conclusion, it does not suffice simply to look for cases in which the same segment displays different behavior under different circumstances. Ohala has shown convincingly that many apparent examples of this sort of evidence in fact fail to substantiate the claim at issue. It is important to ascertain whether there are also cases in which such a difference in behavior is not amenable to a similar extra-phonological account. The fact that there are such other examples, where "phonetic" explanations appear to be unavailable, suggests that Ohala's conclusion does not generalize completely.

In fact, Ohala & Lorentz (1977) mention the existence of such residual examples, but suggest that these also have a non-phonological account: they argue that in all such cases, the apparent distinction involved has a historical origin. In Fula, in fact, the instances of /w/ that alternate with labials and those that alternate with velars derive historically from bilabial and velar fricatives, respectively. A historical change has resulted in the merger of these two segment types as /w/, without any residue of their distinct origins beyond the alternation patterns we have been concerned with here. Similar historical accounts are available for some other parallel cases.

To assert that this historical fact exhausts the phenomenon observed, however, and thus eliminates it as an argument for a synchronic phonological distinction without direct phonetic correlates, is to miss the point of the example. The fact of the "etiology" of the dichotomous behavior of Fula /w/ is strictly speaking as irrelevant to its synchronic behavior as the historical changes described by Grimm's law are to an account of the consonantal phonology of English. The fact is that, regardless of where it came from, the distinction between two sorts of /w/ is a synchronic feature of Fula phonology, demanding a synchronic description. It is also, apparently, a rather stable one, showing no tendency to be obliterated as is often the case with historical residues that cannot be integrated into the system of a language at later stages. The proposed distinction between [+Anterior] and [-Anterior] /w/, in addition, is an entirely natural one on the basis of the feature system proposed. Furthermore, it is motivated entirely by synchronic considerations having to do with the sound pattern of the language. The synchronic security of the opposition here within the system of the language, and the character of the feature system on which it rests, are thus mutually reinforcing.

It is not enough to provide an etymology for a fact to remove it from consideration: one must also provide a description of it which reflects its role in a synchronic _état de langue_. In this case (and in several others), the descriptive account which is apparently required provides us with the crucial evidence we were seeking that phonological distinctions do not always correspond directly to phonetic observables. While this possibility has only been illustrated for the single case of [+Anterior] vs. [-Anterior] labio-velars, it is potentially more general: various aspects of suprasegmental structure (such as syllabification, the organization of syllables into larger units such as feet or stress groups, stress subordination, etc.) immediately suggest themselves as phonologically relevant aspects of sound structure which often cannot be correlated with surface phonetic properties in a straightforward way. There are likely to be other strictly segmental possibilties of this sort as well. A feature system which is directly and exhaustively phonetic in character will not in the general case lead to adequate descriptions of the sound patterns of natural languages.

1.2 _The phonetic adequacy of a "phonological" feature set_. A given phonological representation, we assume, characterizes a restricted class of possible phonetic realizations, and we concluded in the previous section that in some cases two (or more) of these classes may overlap. This, of course, results in a phonetic form with more than one representation. If it is indeed a genuine possibility that the relation between representations in a feature system and their potential realizations is not one-to-one, we must raise the question of what, indeed, the feature system represents.

The notion of distinctive feature originates in the work of Bloomfield and Jakobson, who intended by this to designate those phonetic properties that distinguish contrasting utterances in a given language. A case can be made that (at least originally, in the work of Bloomfield and of Prague school linguists such as Trubetzkoy, on whose ideas Jakobson built) the features that characterized phonemic units were phonetic parameters; and thus that a one-to-one relationship ought to exist between representations and (linguistically significant) distinct phonetic segments. As the theory developed, however (particularly in Jakobson's later work), the role of distinctive features in characterizing phonologically unitary classes became more important than their direct phonetic interpretation.

As the system of distinctive features was employed in phonological descriptions of a variety of languages, however, it became clear that modifications were necessary to the set of features proposed by Jakobson and his colleagues. In some cases, this required the replacement of a feature which subsumed several different distinctions (e.g., the feature [Flat], which described rounding, pharyngealization, velarization, and retroflexion) by multiple features corresponding to the separate distinctions involved. As it developed, the feature system came to resemble more and more closely the set of parameters which a descriptive phonetician might posit. Chomsky and Halle (1968) contributed significantly to this growing equation of features with direct phonetic interpretations when they characterized "[t]he total set of features [as] identical with the set of phonetic properties that can in principle be controlled in speech; they represent the phonetic capabilities of man[.]"

If we take this notion quite literally, we are led to a conception of distinctive features as simply a notation for independently established phonetic properties - the particular set of these that are relevant to human language, at any rate. This is, of course, the interpretation which fits most closely into a program of describing phonology as "natural", for on this view the properties of a representation are derived directly from observable external functions. As Fromkin (1979) observes, this "natural" embedding of phonological representations in the study of phonetic substance is supported, in part, by the observation that the basis of classification (the primary phonological role of a system of features) is typically a phonetically natural, non-arbitrary one: segments which behave similarly in the sound pattern of a language generally have something else in common. If the phonological similarities are rooted directly in the phonetic commonality, this property of natural languages would follow immediately.

However, the observation that phonological classes are generally phonetically natural ones does not really serve to establish the conclusion that phonological features can be identified with articulatory, acoustic, or some other sort of parameters. That is, the central phonological function of a set of features is to allow the classification of sound elements along a number of dimensions, so as to reveal their similarities in behavior with respect to the sound pattern of the language. Even if it is indeed the case that sounds which behave similarly have something in common phonetically, we must still ask whether a system of features which is appropriate for phonological analysis can serve directly as a system for phonetic description as well.

Halle & Stevens (1979) observe that there are a number of distinct bases for grouping sounds into classes. One of these is provided by the systems of the phonologies of natural languages; another is provided by the details of articulation, and still others by the study of acoustic and auditory properties of sounds. Quite often, the same class of sounds turns out to be defineable in more than one of these realms. In some cases, however, a class required for phonological purposes may not have satisfactory definitions in other domains. The class of r-sounds, for instance, has a (somewhat tenuous) auditory similarity, but little in the way of an articulatory property in common; the class of dental stops, on the other hand, can be identified with a set of (context dependent) formant transitions in a preceding or following vowel and with a characteristic noise burst, but (disregarding the lack of invariance in the formant transitions) any one or even two of these three acoustic factors may be absent in a particular environment, without making the sound less of a "dental stop." It is thus primarily articulation that this class of segments has in common. The parameter of Stress is particularly hard to give a unitary definition in other terms: under various circumstances, relatively stressed vowels may have greater length, higher intensity, or higher pitch than relatively less stressed vowels - but no one of these by itself can serve as a "definition" of stress, as a vast literature devoted to the subject attests. Conversely, some classes which have a straightforward acoustic basis (e.g., the class of consonants with falling F_2 onset transitions) have no apparent classificatory unity in phonological terms.

A distinctive feature, then, typically represents a parameter of phonological classification which has some corresponding properties in other (articulatory, acoustic, auditory) domains. However, as Ladefoged (1980; Ladefoged & Traill, 1980) has recently stressed, these correlates are by no means invariant across languages, or even within a single language. He points out that elements which are "the same" as far as their phonological categorization is concerned may be quite different in their realization. The simplest cases of this sort are quite traditional. It has often been remarked that, e.g., the high front vowel of one language (say Danish) is not really quite the same as the high front vowel of another (say English) - though the differences are not of a sort to cause a phonologist to deny that both languages have "high front vowels." This matter of phonetic detail, which has been validated statistically for vowel systems in recent work by Disner (1978), suggests that as a minimum, a complete account of particular languages needs to provide for an interpretation of "systematic phonetic" representations along dimensions with multiple phonetic values.

However, the degree of variance within a single category as given by an (apparently) adequate set of phonological features is not limited to such matters of degree. Recent work (some of it summarized by Ladefoged, 1980; Ladefoged & Traill, 1980) has shown that the same feature may be implemented in quite different ways by quite different mechanisms in different languages. Thus, there are at least two articulatorily distinct segment types that are typically called "implosive stops"; at least three different sorts of "(ejective) glottalized" sounds, etc. Some of this variation falls within the scope of what has sometimes been called "compensatory articulations": the existence of this phenomenon (for instance, in speakers who have undergone surgical removal or other disturbance of the organs of speech, as discussed by Drachmann, 1969; and also in normal speakers under unusual conditions, such as speaking while eating, speaking with a bite block maintaining a constant mandibular angle as in the research reported by Lindblom et al., 1979, or performing as a ventriloquist) provides further evidence that the "same" segment type may have multiple realizations.

Just as more than phonetic dimension may be employed as the realization of a given phonological characteristic, so also more than one feature may affect the same phonetic parameter. Distinctions of voicing, aspiration, and glottalization in consonants; and voice quality (breathiness, laryngealization, etc.) in vowels, for example, are described by a set of features of laryngeal activity. Most of these distinctions, at least some of the time, have consequences for the pitch (F_o) of voiced segments. In addition, a great many languages employ distinctions of tone contrastively, and of course these are also realized as differences in pitch. Nonetheless, the two sets of features cannot be identified (for arguments to this effect, cf. Anderson, 1978a), despite the fact that the same phonetic dimension is of great importance to them both.

The relation between (phonologically motivated) distinctive features and phonetic parameters, then, is typically many-many. A distinctive feature can be thought of as a sort of archetypal contrast between sound elements, which can be realized potentially in a number of different ways. The properties on which these contrasts are based, in turn, may also show some overlap between features. All of the realizations of a given feature share at least a family resemblance (in the sense of Wittgenstein); in the simplest case, they have a single defining property, but this is not necessary. As long as the different realizations of a particular feature are not independently contrastive within any one language, they may be regarded as instantiations of the "same" phonological contrast (though of course, it is necessary for phonological theory to provide a way to specify the translation of such generalized oppositions into language-specific, uniquely interpreted articulatory values and the acoustic/auditory cues which allow their recovery).

Just as a given feature may not always have the same articulatory implementation, it may also not always correspond to the same perceptual cue. For speakers of a great many languages, for instance, pitch change is a highly salient perceptual cue to the identification of tones - even if the language does not have independently contrastive contour tone elements (cf. Gandour, 1978). Thus, in some languages, it may well be absolute pitch level that identifies a tone as phonologically [low], while in others (e.g., Yoruba) it may be not simply the value of F_o, but the presence (in some environments) of a pitch drop within the vowel. The relation between auditory cues and distinctive features, given the context-dependence of most such cues, is even more manifestly many-many than that between articulation and feature representation.

Fant (1970) clarifies the nature of the relation between features and external dimensions when he says that "[a] distinctive feature has certain correlates on each stage of the speech communication chain and these correlates are described in terms of various parameters and cues, e.g. formant locations. A distinctive feature is thus a unit of the message ensemble rather than a property of the signal ensemble." The feature itself is thus a property of the linguistic system (the "message"), while its correlates and cues are aspects of speech (the "signal"). The two are not to be identified, despite the fact that their roles in Language are not independent of one another.

We conclude, then, that the system of representation which is central to phonology cannot be reduced to a particular subset of the dimensions revealed by a language-independent study of the mechanisms employed in speech. The system of features which plays a role in phonological analysis is a function of the human language capacity, and is not "natural" in the sense of following directly from other principles. If we wish to study the nature of phonological representations, we obviously cannot neglect phonetic information: phonological contrasts do, in fact, have (non-trivial) phonetic implementations, and as such, reflect much of the same categorization as do strictly phonetic ones. Nonetheless, the system of features which serves the function of phonological classification cannot be identified with an independently established set of phonetic capabilities of the human organism.

2. The phonetic content of phonological rules. We observed in the previous section that there is reason to believe that the phonological behavior of linguistic elements is not exhausted by its phonetic content, and indeed that its phonetic content is not directly predicted by its phonological character either. When we extend our consideration beyond the treatment of individual segments, to include the content of phonological rules as well, the same conclusion holds: if anything, more self-evidently.

The tradition of seeking phonetic explanations for phonological rules dates (in its serious, not purely speculative aspects) at least to the pioneering research of phoneticians like Sweet, Grammont, Rousselot, and others. These researchers discovered that, when we are able to investigate the fine details of articulation, acoustics, and other phonetic domains, we can in many cases propose that individual changes (seen diachronically by e.g. Grammont, or synchronically, as by most recent workers) have their basis directly in principles governing the operation of the organs or the sounds of speech.

We observe that many languages substitute velar nasals for dental ones when a velar stop follows, for example. It is easy to propose that this follows from a general principle of "ease of articulation": it is easier to articulate adjacent consonants with substantial articulatory similarity at the same place in the mouth than at different places. Such an articulatorily based account is the sort of thing we have in mind when we speak of a phonetic explanation for a particular rule: an appeal to general principles in another domain (here, articulation) which grounds the content of the rule. Similarly, in the discussion of nasal assimilation before labiovelars in section 1.1 above, we appealed to acoustic facts to explain the fact that nasals in this position appear as velars rather than as labials: acoustically, labiovelar nasals and velar nasals differ only in their onset and offset transitions. If the acoustic properties of the consonant itself are of more importance as an auditory cue to the identification of place of articulation in nasals than the transitions, this would (it has been suggested) explain the particular phonological behavior observed.

If we take seriously the program of reducing phonological rules to such phonetic explanations, however, it seems we must inevitably be somewhat disappointed at the comprehensiveness of the results obtained. Indeed, when we examine practically any phonological fact in detail, we find a certain amount of significant arbitrariness that does not appear in any serious sense to be reducible to a mechanical explanation.

For illustration, let us take the example of the length of syllabic nuclei in English words, considered as a function of a following consonant. It is of course well known that there is a close dependence between such vowel length and the identity of the consonant that follows, and the phenomenon has been studied in considerable detail by a number of investigators[5]. The best-known result in this area is probably the following: syllabic nuclei in English (as well as other languages) are significantly longer before a following voiced obstruent than before a voiceless one.

A large number of explanations for this undoubted result have been proposed, based on articulatory, acoustic, and auditory factors. It is not proposed here to review these; as there cannot be said to be any particular consensus about what the correct explanation for the effect is, there is no real point to summarizing one or another account. Nonetheless, since the effect is observable not only across a number of styles of English that have been studied, as well as in a great variety of other (genetically and areally unrelated) languages, we will accept the general view that some sort of mechanical explanation is, in principle, available.

The interesting point, however, is that whatever this phonetic explanation may be, it seems pointless to hope that it will completely account for the phenomenon of vowel lengthening in English. This is because it is clear that what happens in English is, at least in part, a fact about English as an individual language, and not a necessary property of the vocal tract, the character of vowel sounds, the organization of the peripheral auditory system, or the like. Cross-linguistic studies (such as that of Chen, 1970) show clearly that the effect is not at all limited to English. However, the degree of lengthening induced by voiced obstruents (or, perhaps, shortening due to voiceless ones: the difference is immaterial here) is qualitatively different in English than in other languages studied. One can thus maintain that there is a universal, necessary basis for a vowel lengthening effect, but that English makes use of something else: either an idiosyncratically exaggerated version of the general rule, or a distinctly English rule in addition. The relative contributions of the general effect and the English-particular component have been analyzed and particularly clearly elucidated by Lovins (1979). In addition, if the general effect is due in some way to the physiology and functioning of the larynx, the demonstration by Gandour, et al. (1980) that essentially the same facts obtain in esophageal speech as in normal laryngeal speech would support the claim that the observed distribution of length is due at least in part to facts particular to English.

Of course, we can always hold out hope that whatever the explanation for the universal effect turns out to be, it will be such that some other, independently motivated property of English will suffice to predict the difference between this language and others. Nonetheless, this hope appears at the moment to be entirely speculative: a question to pose in research, without any concrete indication of where an answer might lie. It would seem much more reasonable to maintain that there is a rule of lengthening which is simply a fact about English: a part of the abstract, non-mechanical cognitive structure that represents our competence in the language, rather than being "wired in" in some more general system not specific to (this) language.

This result is further strengthened by the apparent fact that some languages (e.g., Saudi Arabic[6]) do not in fact display vowel length differences dependent on following obstruent voicing, contrary to the undoubted general tendency. If this is indeed correct, it would appear that whatever external account can be given for the effect is neither necessary nor sufficient as an exhaustive description of its language-particular instantiations.

The example of vowel lengthening is by no means isolated. Another instance of the same general sort is furnished by the facts of pitch variation in vowels following obstruent consonants. As is quite well known (cf. Hombert, 1978 for a survey of the facts), the pitch (F_o) of vowels following a voiced obstruent shows a general tendency to be lowered, while F_o is generally raised after a voiceless consonant. This effect can be observed in a wide variety of languages, both tonal and non-tonal in character, and there is little doubt that it is due to a feature of articulatory dynamics (cf. Anderson, 1978a for some discussion of the issues involved). It is apparently the case that in some languages, this effect has been sufficiently exaggerated to become the basis of a genuine tonal contrast: we can see this development, particularly common in the languages of the Sino-Tibetan family, as similar in character to the language-particular exaggeration of vowel-lengthening we find in English. On the other hand, when we look across languages, we find that in some languages (e.g., Yoruba: cf. Hombert, 1978) the duration of the perturbation caused by prevocalic consonants on F_o of a following vowel is reduced by a considerable degree in comparison with languages like English. Thus, like the effect of vowel lengthening before voiced obstruents, the perturbation of F_o after an obstruent consonant is a phonetically explicable effect which can still be present to a greater or a lesser degree than usual on a language-particular basis.

These examples are particularly instructive, because they show the two faces of phonological facts. On the one hand, we find that a great many phonological rules are tantalizingly close to some sort of phonetic explanation; but on the other, we find that when we try to pin them down in such terms, they have evidently been transformed into something which is no longer just "functional phonetics." An adequate account of the nature of phonological structure will have to deal with both of these aspects of the phenomena: it must provide some basis for the apparent connection between phonetic and phonological effects, but must also recognize the essential independence of the latter domain from rigid deterministic control by the former. In the following sections, we consider first the nature of "phonetic explanation" (the basis of phonological rules in phonetic substance). We then go on to consider particular examples indicative of the distinctly phonological (or if we think of language as an autonomous cognitive system, distinctly linguistic) principles operative in sound systems.

2.1 _Phonetic explanation and the evolution of rules_. The example of vowel lengthening in English considered above illustrates the general point that even where a phonological process in a given language instantiates something which we believe to be connected in some way with an explanation from outside of language proper, the language-particular rule cannot be identified with the extra-linguistic phenomenon. Of course, most examples show considerably greater differences between the rule in question and the scope of proposed explanations than the simple matter of degree which characterizes this example.

Consider, for example, the assimilation of place of articulation in Icelandic velar consonants to that of a following vowel. In this language, as described by Einarsson (1945) and confirmed by the cineradiographic stufy of Pétursson (1974), the consonants /k/ and /g/ are articulated as velars (written here as [k^h] for voiceless, aspirated /k/; and [k] for voiceless, lenis /g/) when followed by consonants or most phonetic back vowels:

(4)[7] a. kaka "cake (n.sg.)" [k^{h}a:k^{h}a]
b. kúla "bullet" [k^{h}u:la]
c. kría "arctic tern" [k^{h}ri:ja]
d. gata "street" [ka:t^{h}a]
e. gómur "palate" [kou:mÜr]
f. glíma "wrestle" [kli:ma]

When followed by most front vowels, however, these velars are replaced by fronted velars (palatals, written here as [c^h] and [c] for the voiceless aspirated and lenis voiceless segments, respectively):

(5) a. kílo "kilogram" [c^{h}i:lou]
b. keila "cone" [c^{h}ei:la]
c. gifta "marry" [cIfta]
d. Geysir "name of famous hot spring" [cei:sIr]

This fact is of course parallel to that observed in many (indeed, most) languages: velars are front next to front vowels, and back elsewhere. However, in Icelandic the situation is not quite so simple as that. On the one hand, there are certain phonetically front vowels ([Ü] and [ö]) before which back, rather than front velars occur; and on the other hand, there is one diphthong with a back-vowel nucleus before which front velars occur:

(6) a. kunna "know (how to)" [k^{h}Ün:a]
b. gufa "steam" [kÜ:va]
c. köku "cake (obl.sg.)" [k^{h}ö:k^{h}Ü]
d. götu "street (obl.sg.)" [kö:tÜ]
e. kær "dear, beloved" [c^{h}ai:r]
f. gæfa "good luck" [cai:va]

When we inquire as to the possible motivation for the facts presented in (6), we cannot find it in the phonetic character of the vowels involved. The vowels [Ü] and [ö] are clearly front - as front as [e] and [I], for example, as shown by Pétursson's (1974) study - while the nucleus of the diphthong [ai] is equally clearly back - as back as other instances of [a]. While it is indeed appealing to see facts such as those of velar assimilation to adjacent vowels as based on an essentially articulatory motivation, it is apparent here that this account would not suffice to explain the facts, since in crucial places (as shown in (6)), they contradict it.

From a diachronic point of view, however, these facts are by no means arbitrary. The rule of velar fronting has apparently existed in the language for quite some time; the vowels relevant to the facts in (6), however, have only had their present qualities for the past several centuries (a short time by the standards of Icelandic diachrony). The two phonetic front vowels which appear with back velars were previously phonetic [u] and [ɔ], and were subject to a change that fronted them; while the diphthong [ai] was previously (as indicated still in the orthography) [æ:], and has only recently become a diphthong with a back nucleus.

It would appear to be the case, then, that the Icelandic velar fronting rule applies before a class of vowels that were historically front (though synchronically, the class cannot be given this definition). Note that, just as with the case of Fula /w/, it does not suffice to make this observation to make the facts go away: a synchronic grammar of the language must contain a rule which represents them. Consider the following forms of the verb <u>kala</u> "freeze, become frostbitten" for example:

(7) a. (ég) kel ([c^{h}el̥]) 1st sg. pres. indic.
    b. (vi) kölum ([k^{h}ö:lüm]) 1st pl. pres. indic.
    c. (ég) kól ([k^{h}ou:l]) 1st sg. pret. indic.
    d. (ég) k li ([c^{h}ai:li]) 1st sg. pret. subjunctive

In the infinitive, this stem shows /k/ followed by a back vowel, and thus [k^h]. Most of the present indicative forms, however, involve fronting of this vowel from /a/ to /e/; since the latter is a front vowel, /k/ is replaced by [c^h] before it in forms like (7a). In the first person plural, however, the phonological rule of <u>u</u>-Umlaut shifts the stem vowel to [ö]. While phonetically front, [ö] does not condition velar fronting, and thus we get a back velar [k^h] in forms like (7b). In the preterite, the Ablaut class of this verb requires replacement of the basic /a/ of the stem by /ó/: since this is a back vowel, forms like (7c) show a back velar [k^h]. The preterite subjunctive, however, shifts the stem /ó/ to /æ/. Although this is not phonetically a front vowel, it nevertheless (as we saw above) conditions velar fronting, and thus (7d) shows a front velar [c^h]. The velar fronting rule, in the form we would be led to posit from the distributional facts in (4), (5), and (6) is thus closely connected with other alternations in the language, and must be taken to be a phonological rule.

If it is a phonological rule, however, we cannot take it to be one which would be completely "explained" by the phonetic facts of assimilation to the point of articulation of a following vowel, since this would incorrectly describe the facts we have been discussing. We might attempt to save the notion that the velar fronting rule is phonetically explicable by some such move as that of Anderson (1969), where it was assumed that [ü] and [ö] were underlyingly back vowels, and that the diphthong [ai] went through an intermediate stage as a front vowel. If we then posit rules which essentially recapitulate the historical changes, and apply these after the rule of velar fronting, we could (on these assumptions) state the latter in a phonetically natural way.

While this saves the phonetic naturalness of the statement of the rule, it does not improve its degree of connection with the facts of articulation: if the rule applies to representations which are abstract, rather than to actually pronounced forms, it is not clear how constraints on pronunciation can usefully be invoked to explain its form and content. Nonetheless, we might retreat from the strongest possible constraint (that rules must always be phonetically interpreted ones) to the view that "naturalness" holds as a constraint on the statement of rules whether or not they apply to actual surface forms. The constraint of phonetic naturalness would then be a formal part of the theory of grammar, rather than an actual connection between alternations and, say, articulatory dynamics.

This view too, though, has serious weaknesses: we can note first of all that (at least in this example) there is absolutely no evidence for the intermediate representations involved or the rules recapitulating the history except the desire to make velar fronting a phonetically natural rule. Other than that, there is no reason not to take [ü] to be underlying /ü/, for example, or [ai] to be underlying /ai/. Furthermore, other examples which we will see below do not even admit of this (completely ad hoc) alternative, as they involve rules which cannot be stated as phonetically natural at any point in a derivation.

We must conclude, therefore, that the class of possible phonological rules in natural languages includes at least some that are apparently not "natural" in the sense of being synchronically subject to phonetic (or, in more general terms, external) explanation. Some of these rules, indeed, are among the ones that seem at first glance quite close to phonetic motivation.

This example is instructive because it is entirely typical. When we study the evolution over time of a given rule, we find it to be almost a general tendency of sound changes that they very often have the effect of separating a rule from its original phonetic motivation. While there is little reason to doubt that Icelandic once had a rule of velar fonting which was amenable to external phonetic explanation (at least in outline), subsequent changes (the fronting of some vowels, and the replacement of others by diphthongs with back nuclei) have resulted in a state of affairs in which this is no longer true. In fact, if we were to attempt to impose phonetic naturalness as a defining condition on the class of objects we study in phonology, we will not be left with very much.

There is certainly no doubt, however, that a theory of possible (or plausible) phonological rules should reflect in some way the phonetic content of alternations and processes. This is because far too many rules encountered in far too many languages are sufficiently close to phonetic "coherence" for us to believe that this is simply due to chance. On the other hand, languages do, with somewhat depressing regularity, contain unnatural rules as well, rules whose phonetic motivation is not at all apparent synchronically, if indeed they have one. Indeed, as observed in a number of articles by various scholars (notably, by Bach & Harms, 1972) apparently natural rules show a disconcerting tendency to be replaced by "crazy" rules, as in the case of the Icelandic velar fronting rule considered above. Other particularly persuasive cases of the evolution of once-natural processes into phonetically arbitrary ones have been presented in recent papers by Hallberg (1978, 1980).

The mechanisms by which such change in the phonetically motivated character of a rule takes place are those that are proper to the study of the internal development of grammars: changes in linguistic competence, motivated by properties of the system of Language rather than by the necessities of performance, and showing no particular tendency to be constrained by the latter. These are the types of restructuring which various scholars have called "rule inversion", "telescoping", "generalization", and the like, and their principles are ones which are appropriately studied in terms of precise formal statements of the content of grammars, not in terms of the exigencies of articulation, audition, memory, etc. Examples of the operation and consequences of such restructurings will be discussed below in section 2.2.

If there are principles of this sort which operate within the system of Language, we will of course only be able to study them if we take seriously the specifically linguistic nature of the cognitive domain involved, rather than trying to reduce it to some other. And yet we must also ask how this can be done in a way that preserves the other consideration above, namely making allowance for the undoubted influence that extra-linguistic factors do have in shaping the nature of Language.

An appropriate path to the resolution of this paradox is suggested (in outline) in the work of Baudouin de Courtenay[8], one of the pioneers of modern phonology. According to Baudouin, the role of phonetic factors in language is due to the fact that they condition a vast range of low-level, more or less mechanically determined variation in the speech signal: what he called anthropophonic effects in language. These generally operate outside of our recognition, and more importantly, without being determined by the grammar of a particular language. We might attribute the mechanical effect of vowel lengthening before voiced obstruents which is observed in most languages to this domain, for example, or the rise in pitch following voiceless obstruents and the corresponding lowering following voiced ones.

Of course, such low-level, mechanically determined variation in the speech signal is not completely unrelated to the specific, grammatically determined properties of a given language. A particular language may indeed take up some such property and employ it linguistically (generally increasing the magnitude of the effect in the process). This may happen in several ways: for instance, as we have noted, some languages apparently make use of a potentially mechanical aspect of variability to establish or reinforce a contrast, or to provide a cue for some linguistically significant difference between forms. When English exaggerates the degree of necessary vowel lengthening, we can see this as a linguistically determined variation on the phonetically given possibility - perhaps employed to reinforce the language's voicing contrast by providing an alternate auditory cue, but in any event incorporated into the system of the language rather than being any longer simply mechanical. Similarly, when a language develops tonal contrasts out of the originally mechanical pitch variation accompanying a (possibly lost) voicing distinction in obstruents, the same conversion of mechanically conditioned variation into linguistic significance can be seen.

This evolution of mechanical, "anthropophonic" variation into grammatically conditioned, "psychophonetic" variation is what has been called (e.g., by Hyman, 1976 in a valuable study of the role of this process in the development of tonal contrasts) the phonologization of phonetic variation. When a language comes to treat some articulatory, auditory, or other property of speech as being systematically determined - possibly as the basis of a contrast, but possibly just as a linguistically controlled aspect of pronunciation in the particular language or dialect, we can consider that the property in question has been phonologized. Phonologization is most noticeable when it takes the form of conditioned exaggeration of an otherwise mechanical effect, but we can recognize the same development when a general effect is suppressed: in Saudi Arabic, for example, when mechanical vowel lengthening is apparently prevented under linguistic control (possibly for reasons related to the language's independent use of contrastive length, but not necessarily, as apparently mechanical length is present to some degree even in some languages with length contrasts: cf. Lehiste, 1970), or in Yoruba when the amount of perturbation caused by prevocalic consonants in F_o in a following vowel is limited by comparison with a language like English.

The notion of phonologization, then, serves to resolve (or perhaps, rather, to recognize) the paradoxical role of extralinguistic factors in the system of language. On the one hand, as long as some property has not been brought under the control of the grammar of a language, it is subject to determination by non-linguistic constraints or tendencies. On the other hand, the effects of these latter may be explicitly recognized, systematized, and taken over by the grammar if they are phonologized. If we recognize that phonologized effects can overrule merely mechanical tendencies, we can see that everything that is natural need not necessarily be absolutely, deterministically pervasive; and if we recognize the phonologization of "anthropophonic" variation as the raw material giving rise to phonological rules in the grammars of particular languages, as is suggested in Baudouin's program, we can see that rules will tend to express phonetically natural processes.

When a rule is phonologized, however, it is important to recognize that its status has changed: even though it may have originated in the exigencies of articulatory dynamics, for example, when it is incorporated under the control of the cognitive system which is at the heart of language, these factors no longer limit or prescribe its content. The motivations for subsequent evolution of such a process are quite different, and internal to the system of language as discussed above. These, too, were largely outlined by Baudouin: a phonetic difference, once mechanical but now linguistically determined may become the basis of a contrast; the rules distributing some such property in linguistic forms may become opaque through the accretion of other, subsequent rules, and may change in content; rules whose phonological conditions have become sufficiently opaque may be reanalyzed as conditioned by morphological factors (i.e., they may be "morphologized"); and even morphologically conditioned alternations, eventually, may through similar developments be reduced to the status of mere lexical correspondences[9].

The essential step on this path for our purposes, however, is the first one: when a process is phonologized, it becomes in an important sense phonetically arbitrary, even though it may continue to reflect a phonetically natural content. Phonetic explanation thus serves as a sort of constraint on the entry of processes into the system, since (at least many, if not perhaps all) rules originate in the raw material of phonetically natural variation. Once phonologized, however, their essential character is radically altered. When we recognize that much of phonology (and phonological change) is <u>phonological</u> - not phonetic - and thus rooted in the system of language, we can reach a better understanding of the processes involved, and (perhaps paradoxically) of the role of phonetic motivation in language.

2.2 *The interplay of phonetic and non-phonetic "explanations".* Most work in phonological theory that has sought to provide explanatory accounts of observed phenomena (rather than just descriptions) can be divided into two (unfortunately, often mutually exclusive) categories: attempts to find the basis of the facts in extralinguistic domains of phonetic substance; and studies of the internal properties of particular formal systems of description. On the program sketched thus far, however, neither of these approaches can be expected to be completely successful by itself. Only by recognizing both the external phonetic basis of much linguistic variation and the internal linguistic principles that govern its systematization in a grammar can we hope to understand why things are as we find them.

2.2.1 *Compensatory lengthening.* An example of a process type which resists explanation solely in either phonetic or formal terms is the class of rules called compensatory lengthening, studied in some detail recently by deChene & Anderson (1979). These can be characterized roughly as rules - either synchronic alternations or diachronic changes - by which $XVCY > X\bar{V}Y$ (i.e., where a sequence of vowel plus consonant is replaced by a single long vowel. Rules of this sort are one of the best-known classes in the study of phonology and of historical change, and can be attested from a wide variety of languages. This fact has led phoneticians and phonologists to consider compensatory lengthening processes a natural type, and to assume that a unitary phonetic explanation should exist for them.

The usual assumption has been that an explanation invoking some sort of "conservation of timing" could be found. On this view, the basic process at work is the loss of the consonant: it is assumed that, when this segment is reduced and disappears, the vowel is simply prolonged so as to preserve the overall temporal properties of the syllable. The segmental content of the syllable is thus reduced, but the "mora structure" remains.

Given the general lack of understanding we have of the detailed mechanisms of timing in speech, such an explanation must be considered at best programmatic rather than substantive. Its explanatory force would appear to rest essentially on the assumption that a unit like the syllable constitutes a relatively invariant element in the temporal organization of speech production (at least in all those languages in which compensatory lengthening processes are observed). This is certainly not an uncontroversial assumption: the hypothesis of temporal isochrony in speech production has been investigated in some detail in recent years, with somewhat less than comforting results for those who would invoke it in explanation of facts like compensatory lengthening. It appears at present that there is, in some languages, an isochrony effect in speech perception (cf. Lehiste, 1977 for discussion), but a) this apparently has little if any correlate in production; and b) the elements of isochronous structure are often stress groups rather than individual syllables.

Some more direct evidence bearing on the "conservation of timing" account of compensatory lengthening was suggested in a paper by Lehiste (1971). Using a paradigm developed by Kozhevnikov & Chistovich (1965), she reports a pilot study in which she studied temporal compensation within and across syllable boundaries when the same material is repeated over and over. She found some apparent evidence that, for instance, when the vowel of a test syllable is produced slightly longer or shorter in one repetition than another, the duration of syllable final consonants shows a negative correlation with this feature. Subsequent research has not in general given much support to this result, however, and the experimental technique itself has been seriously questioned by other investigators. For example, the use of a negative correlation procedure tends to magnify minor measurement errors in such a way as to exaggerate the compensation effect. Furthermore, the sort of correlation found is also observed between linguistic units between which there is no reason to believe a significant relation exists. For a review of some of the issues involved in this controversy, cf. Fowler, 1977. While it is possible that subsequent research will ultimately validate the notion of syllable-internal duration compensation, it must be concluded that at the present this is a largely speculative idea.

Even if we were ultimately to confirm such a phonetic effect, there are a certain number of apparent peculiarities of observed compensatory lengthening rules for which we would still need to provide an explanation:

A) Apparently, only the loss of a consonant immediately adjacent to a given vowel results in its compensatory lengthening. If the effect is a direct consequence of syllable-internal phonetic timing conservation, we would expect that changes such as $XVC_1C_2 > X\bar{V}C_1$ or $XVC_1C_2 > XV\bar{C}_1$ (i.e., cases in which a syllable final consonant cluster is reduced by the loss of a non-initial element, with resultant lengthening either of the nuclear vowel or of the remaining elements of the cluster) would occur as well.

B) When a language has a rule deleting <u>any</u> consonant (specified only by major class features), this apparently does not lead to compensatory lengthening. Thus, Finnish obstruent clusters are simplified as in (8):

(8) a. /kiitoks/ > <u>kiitos</u> "book (n.sg.)" (gen. <u>kiitoksen</u>, part. /kiitoks+ta/ > <u>kiitosta</u>)
    b. /lapsi/ > <u>lapsi</u> "child (n.sg.)" (gen. <u>lapsen</u>, part. /laps+ta/ > <u>lasta</u>)

This reduction of three obstruents to two medially or two to one finally does not lead to compensatory lengthening (else we would get forms like *<u>kiitoos</u>, *<u>laasta</u>). It is apparently the case that compensatory lengthening only arises when some particular set of consonants, defined in terms of an articulatory category rather than simply major class features, is lost.

C) Apparently, compensatory lengthening does not arise unless a language already has distinctively long vowels and/or diphthongs. These elements may have arisen from sequences of simple vowels (/...V+V.../ > /...$\bar{V}$.../); from the loss of intervocalic consonants (/...VCV.../ > /...VV.../ > /...$\bar{V}$.../); or from some other source; but apparently languages do not develop a new length contrast where there were no long syllabic nuclei before solely through the operation of compensatory lengthening. DeChene & Anderson (1979) document a number of consonant-loss processes in the history of French, for example, some of which lead to compensatory lengthening of vowels and some of which do not. The important point is that those which do not result in long vowels all antedate the independent development of contrastive long vowels from other sources; those which do lead to compensatorily lengthened vowels all post-date this development.

All of these properties of compensatory lengthening would appear to remain arbitrary on the view which attributes the phenomenon to a purely phonetic timing effect. There is no reason, that is, why such an effect should discriminate among consonant-loss processes on the basis of such system-internal considerations as those adduced above: if the effect is genuinely phonetic in character, we would expect it to be operative wherever consonants are lost within a syllable (subject, perhaps, to differences between languages with respect to the elementary units of timing - but there is no apparent reason to believe that "syllable-timed" languages differ from "stress-timed" ones with regard to the possibility of compensatory lengthening, insofar as we can divide languages into these two types).

A hint that compensatory lengthening is not, perhaps, a unitary phenomenon after all is furnished by the fact that in many cases, apparently unmotivated quality changes are associated with its results. Consider the following developments in the Swiss German dialect of Visperterminen (cf. Wipf, 1910 for details), for example[10]:

(9) a. an > ã: (e.g., sã:ft "Sanft") [before /f/]
b. an > ãi (e.g., bãix "Bank") [before /s,š,x/]
c. en > ẽi (e.g., dẽixu "denken")
d. in > ĩ: (e.g., fĩ:šter "finster")
e. un > ü̃: (e.g., xü̃:št "Kunst")
f. ün > ĩ: (e.g., frnĩ:ftig "vernünftig")

In this dialect, nasals have evidently been lost before following spirants, resulting in long or diphthongized (nasal) vowels. The developments in (9a,d) would be considered a straightforward instance of compensatory lengthening, but (9b,c) show that the process of interest involves not simply loss of the nasal, but its conversion to a semivowel: /VnS/>/ṼiS/. If we assume that this development was a general one, we have an account of (9e): /unS/>/ũiS/, with subsequent reduction of /ũi/ to /ü̃:/. This conjecture is supported by the isolated form [xuišt] "kommst", in which for some reason (possibly the fact that nasality appears to have been lost as well in this form) the monophthongization of /ui/ to [ü:] failed to occur. Similarly, for (9f), we can posit /ünS/>/ü̃iS/ with reduction of /ü̃i/ to /ĩ:/. In (9a), the diphthong /ãi/ is seen to monophthongize to [ã:] before /f/, but not (as seen in e.g. 9b) before other fricatives. Wipf does not give any examples of the development of <u>en</u> before /f/; we cannot therefore tell whether (9c) represents the only possible outcome for <u>en</u> or whether a similar monophthongized form would exist in this environment.

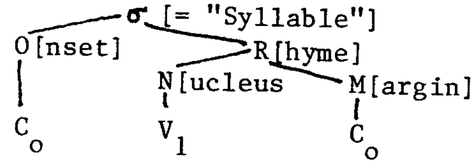
The point of this example is that the appearance of long vowels as a result of the loss of a nasal consonant is evidently not a simple matter of timing here. Rather, the nasal consonant is first reduced to a semivowel, and the resulting vowel plus semivowel sequences are then further reduced or reinterpreted. This ultimately leads to long vowels or to diphthongs depending (<u>inter alia</u>) on how different in quality the two elements are.

On the basis of examples like this, it was suggested by deChene & Anderson (1979) that perhaps compensatory lengthening is <u>always</u> the outcome of a similar development: not a unitary phonetically based process, but rather a combination of a) the weakening of some post-vocalic[11] consonant to a semivowel; b) the reinterpretation of the resultant vowel plus semivowel sequence as a diphthong; and c) the monophthongization of these diphthongs, resulting in a long vowel. Of course, in some languages step <u>c</u> of this development may either not take place (as in examples (9b,c) above), or it may lead to concomitant shifts in quality (as in (9d,e) above).

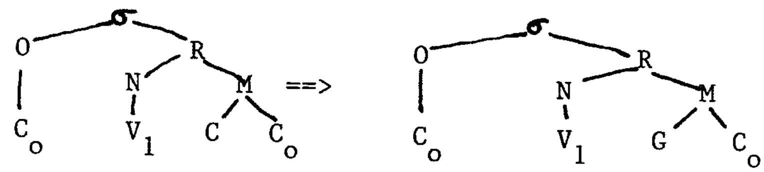
This view renders the peculiarities of compensatory lengthening noted above immediately comprehensible. The fact that it is exactly an immediately adjacent consonant which gives rise to "compensatory length" follows from the fact that unless it is adjacent to a vowel, the reduction of a consonant to a semivowel will not result in a (potential) diphthong, and hence will not furnish the conditions necessary for monophthongization to a long vowel. Similarly, we can find abundant precedent for individual, phonetically natural processes converting various consonant types into glides in syllable-codas (e.g., stops may become [w,j]; spirants may become [h]; etc. - cf. deChene and Anderson, 1979, for further discussion), but no such phonetically natural evolution which is independent of other articulatory properties of the segments involved. On the other hand, we do find "indiscriminate" processes reducing the complexity of consonant clusters in certain positions by deletion. This complex of facts accounts for the observation that when a particular consonant or coherent class of consonants is reduced, "compensatory length" may result (since the reduction may take place through conversion to a glide), but when all consonants are reduced in a given position, it does not (since here reduction is by deletion, rather than weakening).

To understand the role of point C noted above, let us first formulate the proposed changes which ultimately lead to "compensatory lengthening" in a somewhat more precise fashion. Recall that we claimed above that these involved essential reference to syllable structure - since it is asserted to be (phonetically natural) syllable-final weakening processes that furnish the first step of the proposed chain of explanation. Accordingly, we need to be able to describe the operation of processes sensitive to (intra-)syllabic structure.

We assume here (in line with arguments presented in Anderson, 1974, and also by a number of other scholars in recent years) that the organization of an utterance into syllables is a relevant aspect of phonological representation (in addition to its organization into segments). Syllables, in turn, can be assigned internal constituent structure: we assume that this consists of the familiar division of syllables into an onset (O) and a rhyme (R). The rhyme, in turn, can be regarded as consisting of a nucleus (N) and a margin (M). The latter is traditionally named the "coda"; we choose "margin" for convenience in labelling constituents. We assume that this structure is universal, though of course languages may differ in the specific material they allow to be assigned to any of the syllable's constituents. Some, for example, require that the onset be filled; some do not allow any material in the margin; etc. It is apparently the case that while either the onset or the margin may be absent, no language allows syllables without any nucleus constituent, though the nuclei of some syllables may contain no distinctive phonological material, as suggested for the Northwest Caucasian languages in Anderson, 1978b. The feature [±Syllabic] of Chomsky & Halle, 1968 and Anderson, 1974 can be regarded as equivalent to one aspect of intra-syllabic position: "[+syllabic]" segments are precisely those which are elements of N, while "[-syllabic]" segments are those (perhaps identical in all other features) which are members either of O or of M. The posited constituent structure can be represented as (10) below[12]:

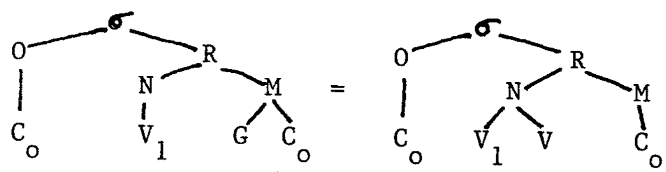
(10) 

In these terms, the first step leading to a rule of "compensatory lengthening" is some process of syllable-final weakening, representable schematically as (11):

(11) 

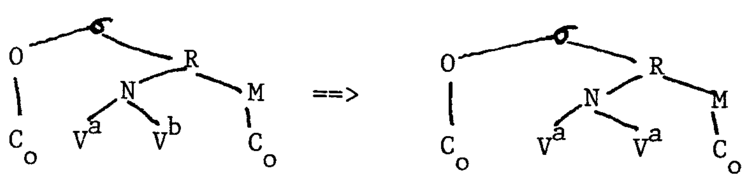
That is, a consonant belonging to the margin is reduced to a glide (G).

As a result of this weakening, a sequence of vowel plus glide is produced. If this glide is (or becomes) semivocalic in character, it is potentially ambiguous in structural position: it may be treated as a consonant, belonging to the margin, or it may be regarded as the second element of a diphthong, and thus as a member of the nucleus. On this second possibility, the output of the process in (11) is re-interpreted structurally as (12):

(12) 

This reinterpretation depends on (inter alia) the fact that semivowels and vowels have the same features, differing only in structural position (corresponding to a difference only in "[±syllabic]").

The reinterpretation in (12) results in a diphthong; the final stage in the development of "compensatory length" is a nucleus-internal assimilation of vowel quality (i.e., "monophthongization"):

(13) 

On the assumption (cf. Anderson, 1974 for some discussion and references) that long vowels are to be represented as sequences of identical vowel elements (similar to "moras") forming a single constituent, the result of the change in (13) is a long vowel of uniform quality.

The sequence of developments involving (11), (12), and (13) can be regarded as providing a "phonetic explanation" of the phenomenon of compensatory lengthening. This is because (11), syllable-final weakening, is generally thought to be tied in an explanatory way to extralinguistic properties of syllable-final articulatory dynamics, while (13) - the only other overt change involved - is a straightforward assimilation, again usually conceded to be relatable to articulatory effects. These "explanations", while far from firm in character, nonetheless seem more satisfactory than the essentially speculative account based on "conservation of timing" considered above.

We have already seen that this view allows for a straightforward incorporation of properties A and B above as constraints on "compensatory lengthening" processes. To see that property C is also accounted for, we turn to the reinterpretation represented in (12). Clearly, if we want to represent monophthongization as "nucleus-internal quality assimilation" (a rather natural step), this reinterpretation is an essential bridge between the processes of (11) and (13). However, (12) rests on the premise that a sequence of vowel plus semivowel is in fact structurally ambiguous. This ambiguity obviously exists only in a language which allows for complex syllable nuclei in the first place: if a language is subject to the constraint that all of its nuclei contain exactly one vocalic element (equivalent to the exclusion from its vowel inventory of long vowels or structural diphthongs), a sequence of two (potentially) vocalic elements within the same syllable is not ambiguous, and the conditions for (12) are not met. We can suggest, then, that on the view which takes (11) and (13) as essential parts of the development of "compensatory length", the limitation in C above follows naturally, since it is clearly equivalent to the possibility of the intermediate step (12). This further confirms the correctness of this "explanation" over that predicated on simple "conservation of timing."

Another account of "compensatory lengthening" processes has been given in a recent paper by Ingria (1980). This treatment contrasts with the traditional one, in that instead of seeking an explanation for the phenomenon in properties of phonetic substance, Ingria approaches it strictly as a formal problem. He attempts, that is, to show that the properties of compensatory lengthening rules follow directly from the formal expression of deletion rules within a framework recognizing syllabic structure and incorporating a particular view of the nature of phonological length.

Ingria's treatment shares some properties with that given above. In particular, he recognizes the significance of the adjacency of the deleted segment to the lengthened vowel (point A above); and also the crucial role played by language-particular syllable structure rules. According to this view, compensatory lengthening follows automatically when a post-nuclear consonant is deleted: the preceding vowel, by general convention, is automatically associated with the structural position of the deleted segment, and (on the view that a long vowel is precisely one associated with two structural positions rather than just one) this results in derived length.

Without going into detail, we claim that, while this view is more satisfactory than the traditional one, its exclusive focus on the synchronic expression of compensatory lengthening rules as the source of explanatory principles results in an account which is incomplete at best. It would appear difficult on this view to incorporate compensatory lengthening as a function of the loss of preceding consonants (cf. deChene & Anderson, 1979); to differentiate general from specific consonant loss processes (point B above); or to incorporate quality changes that are concomitant with compensatory lengthening (as in the Swiss German example discussed above). If we are correct about the interrelation of phonetic effects, phonological restructurings and historical change, it is highly unlikely that a comprehensive account of a phenomenon such as compensatory lengthening can be achieved through a strictly synchronic study of such rules: much of what determines their character is to be found not in their formal expression, but in their originary processes.

Now it is important to be clear that the sequence of developments (11)-(12)-(13) which we envision is to be considered as a historical chain of events, which may well not be reflected in an isomorphic fashion in individual synchronic grammars after the entire sequence has been completed. If a language has a synchronic alternation between, say, /VC+V/ on the one hand, and /$\bar{V}$+C/ and /$\bar{V}$#/ on the other, there is no reason (*ceteris paribus*) not to formulate this simply as the phonological rule in (14), rather than as a sequence of three rules:

(14) $$VC \text{ ---> } \bar{V} \ / \ \underline{\quad} \begin{Bmatrix} +C \\ \# \end{Bmatrix}$$

This rule (which could presumably also be formulated in terms of syllabic structure in most cases), is the synchronic residue of the sequence of diachronic changes schematized in (11)-(12)-(13). It exemplifies the type of restructuring often called "rule telescoping", by which a sequence of processes A--->B followed by B--->C (in the absence of independent evidence for the intermediate stage B) is represented by the single rule A--->C. If the explanation proposed here (and supported in more detail in deChene & Anderson, 1979) is essentially correct, rules of deletion with "compensatory lengthening" will always have the character of telescoped rules. If we were to seek to account for the nature of rules like (14) either in phonetic terms (by connecting them with independently established properties of some system external to language) or in purely formal terms (by explaining their particular properties as a consequence of general conventions stated in terms of their expression in a formal system), we would in an important sense be bound to fail - for the "explanation" of such a rule lies in a sequence of developments, each of which has its own motivation. Rule (14) is thus *not* a "natural rule", in the sense of one which can be tied *per se* to a phonetic motivation or which can be shown to follow from the possibilities inherent in a particular formalism.

This interplay is an example of the role of phonetic and non-phonetic factors in "explaining" phonological rules. Phonetic explanations in the traditional sense apply to the raw material of language change, which may become phonologized and thus represented in the grammar; each such rule is then subject to phonological interpretation and restructuring in ways that do not depend on such phonetic connections. One of the principles of phonology which operates in Language (as distinct from speech) is that of rule telescoping. To understand the kinds of phonological rules that occur in languages, we have to recognize that phonetic change and its constraints are only a part of the story. Once we accept this, it is apparent that rules do not have to have unitary phonetic explanations in order to be subject to systematic study. This is fortunate, since rules of the type (14) apparently do not have such an explanation in a direct sense.

2.2.2 <u>The quantity and quality of mid vowels in Breton</u>. For another example of the difficulties that arise if we insist on a "natural" phonetic basis for synchronic phonological rules, we turn to a set of interrelated processes in Breton. We discuss this example in considerable detail, in order to illustrate the sort of historical development that can be expected to divorce phonological rules from their original phonetic motivation. As virtually all descriptions of the relevant dialects[13] have stressed, there are three parameters which are closely connected with one another in this language: a) manner of articulation for consonants; b) length in vowels; and c) the quality of mid vowels. As we will see below, the historical developments which have led to the modern system are all of the sort which we would normally assume to have a "phonetic explanation"; however, various restructurings of the system which have taken place have led to a state of affairs in which at least some of the rules no longer have "natural" expressions. Of course, the existence and sources of "crazy" rules with natural antecedents has been generally acknowledged at least since the work of Bach & Harms (1972). The interest of this case for our purposes, however, lies in the fact that the interactions among the rules involved are such that it is not possible (even if it were desirable) to maintain a "natural" formulation by building an artificial recapitulation of the historical sequence of events into the synchronic account.

We begin by summarizing the facts as described by Falc'hun (1951). In addition to a number of diphthongs and the nasal vowels [ẽ, ö̃, õ, ã] (which will by and large not concern us below), there are three high vowels [i,ü,u], one low vowel [a], and a total of nine mid vowels that potentially contrast. Of the mid vowels, we find three each in the front unrounded, front rounded, and back unrounded categories; in each series, we find a "high-mid", a "true mid", and a "low-mid" vowel (in the terminology of Jackson, 1967). These can be represented[14] as in (15):

(15)	[-back,-round]	[-back,+round]	[+back,+round]
high-mid:	[é]	[ö́]	[ó]
true mid:	[è]	[ö̀]	[ò]
low-mid:	[ê]	[ö̂]	[ô]

Most of these elements (including the high and low vowels, as well as the mid vowels) can appear either short or long depending on stress and the surrounding consonants; the rules governing the distribution of length will be discussed below. Stress normally falls on the penultimate syllable (except in monosyllables, of course), though there are a few polysyllables with (synchronically) exceptional final stress.

The consonant system is rather complex. We ignore here the set of distinctions in word-initial position, since the phenomena of interest to us concern the interaction of stressed vowels and post-vocalic consonants. We can divide the consonants in (post-stress) intervocalic position into two classes: obstruents (stops and fricatives) and sonorants (semi-vowels, nasals, and liquids). The obstruents can in turn be divided into two classes, on the basis of the opposition between <u>fortis</u> ([+tense]) and <u>lenis</u> ([-tense]) segments. Fortis consonants are generally voiceless, while lenis ones are voiced in intervocalic position; in addition, the fortis series are significantly longer in closure or frication period (as determined instrumentally by Falc'hun, 1951). Intervocalic fortis consonants are described by Jackson (1967) as ambisyllabic, while lenis consonants are said to be syllabified only with the following vowel. Furthermore, Falc'hun (1951) shows that there is a considerably greater build-up of intra-oral pressure in the fortis series than in the lenis. We represent the segments involved as follows:

(16)

	Fortis	Lenis
Stops:	p,t,k	b,d,g
Fricatives:	f,s,š,x	f̌[15],v,z,ž,h[16]

In absolute final position, the two series are neutralized as a series of lenis, voiceless segments: that is, as segments having the duration and intra-oral pressure characteristics of the lenis series, but without voicing.

Turning now to the sonorants, we observe a parallel opposition between fortis and lenis sound types, except that there is no concomitant voicing difference here (the sonorants all being voiced in most positions). In intervocalic (post-stress) position, the semivowels [j, ẅ, w] are always fortis. Among the liquids and nasals, we find an opposition between fortis and lenis sounds only for [l, r, n]:

(17)

	Fortis	Lenis
semi-vowels:	j, ẅ, w	
liquids:	l:, r:, Ĭ[17]	l, r
nasals:	m(:), n:, ñ[17]	n

The palatal sonorants ([ñ, Ĭ]) and the labial nasal ([m]) are always fortis. Unlike the situation with the obstruents, the [±tense] opposition is not generally neutralized for sonorants in final position. Only /r:/ is replaced by its lenis correspondent /r/ finally; the other fortis sonorant consonants remain distinct.

Having surveyed the sound system of the language, we can now proceed to the facts of interest: the rules governing vowel quantity and the quality of mid vowels. The features which most concern us are those which characterize stressed syllables; in unstressed syllables, all vowels are short, and the quality of mid vowels is basically invariant (they appear as "true mid" vowels here). A low-level phonetic process may assimilate the exact degree of height in an unstressed mid vowel to that of a mid vowel in an adjacent syllable (cf. Anderson, 1974 for discussion of this rule), but this process will not concern us here.

We turn first to the rules governing the distribution of vowel quantity. These affect all vowels equally (high and low as well as mid). In unstressed syllables, all vowels are short. In stressed syllables, vowels are long under any of three conditions: a) when final (e.g., [ti:] "house", [bró:] "country"); b) when followed immediately by another vowel in hiatus (e.g., [bi:ãn] "small", [kó:ãn] "supper"); and c) when followed by a lenis consonant (e.g., [pa:gãn] "pagan", [i:bil] "peg", [ró:žu] "wheels", [bé:lêr] "cress", [ré:d] "necessary", [kã:n] "song", [bé:r] "spit"). As these examples show, the lenis consonant in question may be either an obstruent or a sonorant, and it may be either medial or final.

In the remaining possible environment (i.e., before a fortis consonant), vowels appear short (e.g., [lakafe] "would put", [pakãn] "I catch", [tèn:a] "to draw", [tam:al] "to blame", [kãn:] "white". Again, the fortis consonant in question may be either an obstruent or a sonorant, either medial or final - though as noted above, all obstruents in final position are lenis, and thus the environment before a final fortis obstruent is not attested.

On the basis of the facts discussed above, we might suggest that quantity is governed by the following rule (assuming that all vowels are basically [-long]):

$$(18)\quad \begin{bmatrix} \text{+syll} \\ \text{+stress} \end{bmatrix} \text{--->} [\text{+long}] \;/\; \underline{\quad\quad} \begin{Bmatrix} \# \\ [\text{+syll}] \\ [\text{-tense}] \end{Bmatrix}$$

In unstressed syllables and before tense consonants, of course, vowels will remain short.

Let us now consider the rules which determine the quality of mid vowels (as "high-mid", "true mid", or "low-mid"). We turn first to the distribution of "low-mid" vowels. In stressed syllables these are generally found (to the exclusion of other mid vowel qualities) in the position before certain consonants ([x, h, r:]) and certain clusters ([r] plus consonant; [l] plus [x] or [h]). Examples include: [sêxa] "driest", [sê:ha] "to dry", [pôr:od] "rut", [sŏrt] "sort", [dêlhêr][18] "to hold." This is evidently a fairly natural class of environments, at least in part. One might propose that at an earlier point in the history of the language, [h] was a voiced velar spirant (which is quite uncontroversial; [h] clearly derives from velar stops by lenition in a great many cases), and that lenis [r] was dental while fortis [r] (including [r] in clusters) was uvular (essentially the state of affairs in some modern Franco-Provençal dialects: cf. Bjerrome, 1957). The environment for the lowering of mid vowels would then be before [+back] consonants. It is not clear that such a natural rule can be posited for the modern language, however, and we will assume simply that the relevant environments can be listed:

$$(19)\quad \begin{bmatrix} \text{+syll} \\ \text{-high} \\ \text{-low} \\ \text{+stress} \end{bmatrix} \text{--->} \text{"lower mid"} \;/\; \underline{\quad\quad} \begin{Bmatrix} ([\text{l}])\ [\text{x, h, r:}] \\ [\text{r}]\,[\text{-syll}] \end{Bmatrix}$$

The presence of such a rule in the grammar is necessary (independently of the fact that it expresses a generalization about the language), because it is responsible for alternations: e.g., [kémé:ret] "taken", from /kemer/ "take", but [kemêrfe] "would take" with the conditional desinence /fe/; [bè:r] "short", from /ber:/ with reduction of /r:/ to [r] in final position, but [bêr:a] "shortest", with /r:/ preserved because non-final.

Turning now to the distribution of the "high-mid" and "true mid" vowels, we find that the general trend is quite similar to that of the distribution of length. In the majority of cases, that is, we find stressed "high-mid" vowels in final position (e.g., [bró:] "country", [mé:] "1st sg. pronoun"), or before another vowel in hiatus (e.g., [kó:ad] "wood", [ké:ar] "town"), or before a lenis consonant (e.g., [skø:l] "ladder", [logó:den] "mouse"). Stressed vowels in the remaining environment, namely before a fortis consonant, typically have "true mid" quality (e.g., [kønøta] "to collect wood", [šadèn:et] "chained", [kòl:] "to lose"). One might be tempted, therefore, to posit a rule like (20):

$$(20)\ \begin{bmatrix}+\text{syll}\\+\text{stress}\\-\text{high}\\-\text{low}\end{bmatrix} \text{---> "high-mid"} \ / \ \_\_\_\_ \begin{Bmatrix}\#\\ [+\text{syll}]\\ [-\text{tense}]\end{Bmatrix}$$

This rule, of course, assumes that all mid vowels have "true mid" quality underlyingly, and that this is preserved in unstressed syllables and before fortis consonants. Rule (20), however, misses an important generalization: apparently, mid vowels have "high-mid" quality in the same environments as those in which they are lengthened. The connection between these two properties is quite a natural one: in language after language, we find that long vowels have somewhat "tenser", "higher", etc. qualities than do short vowels. While the phonetic basis for this phenomenon is not immediately apparent, we can accept the proposal that a rule like (21) would have such a basis at least in principle:

$$(21)\ \begin{bmatrix}+\text{syll}\\-\text{high}\\-\text{low}\\+\text{long}\end{bmatrix} \text{---> "high-mid"}$$

It is necessary to have such a rule as (21) (or perhaps (20)), just as we needed rule (19), not only insofar as it expresses a true generalization about the language[19], but also because there are numerous alternations which it accounts for: cf. [bró:] "country" vs. [bròju] "countries"; [ré:d] "necessary" vs. [rètôx] "more necessary", [logó:den] "mouse" vs. [logòta] "look for mice", [é:f̌a] "to drink" vs. [èfše] "would drink", etc. Of course, these alternations could be described as well by positing underlying "high-mid" quality, and writing the rule so as to <u>lower</u> this to "true mid" before fortis consonants (or when [-long]); the choice of (20)/(21) at this point is largely arbitrary. Either way of writing the rule would allow us to claim that it is natural because it expresses a natural connection between vowel length and vowel height (or tenseness).

The situation described by rules (20)/(21), or the alternate formulation just suggested, is potentially a phonetically natural one, and there is little doubt that the Breton phenomena under consideration had this form at one point. Jackson (1967) argues that the distribution of quantity described by rule (18) above was a part of the language which is the common ancestor of Welsh, Cornish and Breton (i.e., prior to the fifth century A.D. at the latest), a claim which is supported by the fact that Welsh shows virtually the same rule. The rules for mid-vowel quality, according to Jackson, were established in approximately the eleventh century, as shown by the fact that they treat all vowels which merged as /e/, /o/, or /ø/ prior to that date in the same way. This chronology, together with the apparent phonetic basis for the facts, gives us no reason to doubt that when the quality alternations entered the language, it was in the form of a rule like (21), applying to the result of the already established lengthening rule (18).

Unfortunately, subsequent developments affected the language in such a way as to make this simple system no longer adequate. For example, a number of different sound changes subsequent to the establishment of the mid vowel quality rule in the eleventh century produced new instances of mid vowels which did not behave as this rule would predict:

A) The diphthong /ae/ was simplified to [e] in most areas. This new [e] had the quality [è] (i.e., it was a "true-mid" vowel) which was not subject to adjustment by rule (21), though it was subject to lengthening by (18). This results in some long "true-mid" vowels in the modern language, in words like [bè:leg] "priest" (compare [bé:lêr] "cress"). Words of this sort would have to be treated as exceptions to rule (21).

B) In some positions, the original glide /w/ was vocalized to [o]. This development could have the effect of placing a mid vowel in position for lengthening, in which case rule (18) applied - but the quality of the resulting vowel was not affected by rule (21), resulting in further instances of long [è:] in the modern language in words like [dè:ro] "oaks" (from earlier /derw/). These words, also, would have to be treated as exceptions to (21) in the modern language (though not to (18)).

C) Some consonant clusters were simplified as well, again creating the conditions for lengthening of preceding mid vowels. This also yields long mid vowels in the modern language, in words like [bè:red] "cemetery" (from earlier /bezred/); these too would have to be treated as exceptions to rule (21).

As a result of these various developments, modern Breton must distinguish between alternating and non-alternating mid vowels: that is, between vowels which are subject to rule (21) and those that are not. The most straightforward way to do this would be to represent only the non-alternating vowels as underlying "true-mid" in quality, treating the alternating ones as underlying "high-mid" vowels. If we do that, we need only replace rule (21) with (22):

(22) $\begin{bmatrix} +\text{syll} \\ -\text{high} \\ -\text{low} \\ -\text{long} \end{bmatrix}$ ---> "true mid"

This rule merges underlying "high-mid" vowels with underlying "true mid" ones when the conditions for lengthening by rule (18) are not met (i.e., when the vowel is followed by a fortis consonant or is not stressed). As such, it does not yet result in a phonetically "unnatural" rule in the phonology. These developments and their attendant restructuring, however, do have the effect of breaking the tight link between length and the quality of mid vowels, since this quality is now distinctive in at least some lexical environments.

Rather more damaging to the connection between length and quality, however, were the consequences of the development of consonants in final position. As noted above, in final position all obstruents are represented by lenis forms, whether they are underlyingly fortis or lenis. In absolute final position, these lenis consonants are all voiceless[20]. Before these final obstruents, rule (18) applies - since they are in fact lenis. However, the quality of preceding mid vowels is not altered: i.e., whatever rule is responsible for determining quality applies to yield a lower vowel if the following consonant is underlyingly fortis. Using [d̥], etc., to represent the final voiceless lenis obstruents, we get forms such as [tɔ̀:g̥] "hat" (from /tok/; cf. [tɔ̀ku] "hats"), [lɔ̀:b̥] "(he) strikes" (from /lop/; cf. infinitve [lɔ̀pa] "to strike"), [prɔ̀:b̥] "clean" (cf. diminutive [prɔ̀pig̥]), etc. The lengthening of vowels before these (derived) lenis consonants is by no means restricted to mid vowels, either: cf. [du:g̥] "duke", pl. [duked̥]; [ka:z̥] "(he) sends", vs. [kasan] "I send", etc. As a result of the operation of this neutralization, forms like [ka:z̥] "cat", from /kaz/, and [ka:z̥] "(he) sends", from /kas/, are completely homophonous. It is only in the case of words with mid vowels that the surface form is unambiguous, for here the differential operation of the vowel quality rule indicates whether the final consonant was underlyingly fortis or lenis.

The neutralization of final obstruents as voiceless lenis is characteristic of absolute final position; in phrase internal position, the character of a following segment influences the exact form which a word-final obstruent takes. When an obstruent-final word is immediately followed by a word beginning with a sonorant, the obstruent is still neutralized to a lenis segment, but here (again, regardless of the underlying value the segment may have for the feature [±tense]) the resulting consonant is voiced. Thus, [pa:b ü:el] "noble pope" (from /pab/; cf. [pa:bed̥] "popes") and [dü:g ü:el] "noble duke" (from /dük/; cf. [düked̥] "dukes") both show voiced lenis stops in spite of the fact that the underlying final consonant of /dük/ is a fortis segment. Again, as in absolute final position, a mid vowel shows the quality that would be expected in light of the underlying form of the following consonant, not its surface form: cf. [tɔ̀:g ü:el] "high hat" (from /tok/, as shown by the plural [tɔ̀ku]). Word-final obstruents, then, are uniformly neutralized as lenis elements, which may be either voiced or voiceless depending on the environment in which they appear[21].

The only segments which preserve the fortis/lenis opposition in final position, then, are the sonorants [n]/[n:] and [l]/[l:] (cf. pairs like [ébé:n] - with exceptional final stress - "the other of two (fem.)" vs. [é bɛ̀n:] "his head", [mé:l] "honey" vs. [mɛ̀l:] "ball (glossed as "ring" by Falc'hun, 1951)"). The segments [m(:)], [ñ], and [Ī] are always fortis, in final as well as medial position. The remaining consonantal sonorant pair ([r]/[r:]), however, is neutralized as (voiced) lenis [r] finally (cf. [ka:r] "cart", from /kar:/ as shown by the plural [kir:i], and [ga:r] "leg", from /gar/ as shown by [ga:reg̥] "one with large legs"). Again, the difference between underlying fortis and lenis /r(:)/ is revealed by the quality of a preceding mid vowel. We might expect a word like /ber:/ "short" or /tor:/ "broken" to show a surface vowel of "low-mid" quality, since (medial) fortis /r:/ is preceded by such vowels as a result of the operation of rule (19). This is not what we find, however. Instead, /ber:/ is surface [bɛ̀:r] (compare [bé:r] "spit", from /ber/) and /tor:/ is surface [tɔ̀:r] (compare [dó:r] "door", from /dor/). Evidently, the surface vowel is determined by the fact that the underlying final segment is [+tense], while its length is determined by the fact that the surface vowel is [-tense]. The derivation of e.g. [bè:r] must thus be as follows:

(23) underlying form:	/b e r:/	
"high-/true" mid rule:	è	
r: --> r /___#:	r	
rule (19):	- - -	(not applicable)
rule (18):	è:	
surface form:	[bè:r]	

The essential point of this derivation is the observation that the rule determining "high-mid" vs. "true mid" quality must precede the neutralization of final /r:/ and /r/ as [r], while the rule determining vowel length (rule (18)) must follow it. This is the same conclusion we reached above implicitly for the rule neutralizing final obstruents as lenis: the "high-mid"/"true mid" rule must precede the neutralization, while the length rule must follow it.

To complete the picture of mid-vowel quality in Breton, we can also consider the facts that obtain before consonant clusters. In medial position, the facts are as follows: before clusters of sonorant plus obstruent or obstruent plus obstruent, stressed vowels are short and mid vowels have their "true mid" quality (or "low-mid", if the cluster is one which causes rule (19) to apply). Examples include [amzêr] "weather", [èmban:] "publish", [ènvel] "to name", [mantrus] "astonishing", [mèstrèz] "mistress", etc. Before medial clusters of obstruent plus sonorant, however, stressed vowels are long and mid vowels have "high-mid" quality: cf. [a:brant] "eyebrow", [lé:grest] "lobster", etc. This distribution follows from the rules we have already established, if we observe that in the case of the "lengthening" obstruent plus sonorant clusters, the obstruent is always a lenis one (by virtue of constraints on possible clusters, resulting in part from the historical operation of lenition). Thus, if the rules for length and mid-vowel quality predict long vowels and "high-mid" quality before lenis consonants, they will apply here correctly. Clusters of obstruent plus obstruent, however, are always made up of fortis consonants; and if we assume further that sonorant consonants preceding obstruents are similarly fortis in articulation, we reach the result that the clusters before which vowels show up short and with "true mid" quality are those which begin with a fortis segment - again, just what is necessary to make the independently established quantity and quality rules yield the correct results. We conclude, then, that no modification of the generalizations already given is necessary to account for medial clusters.

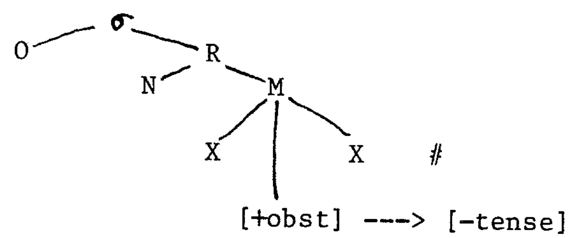
When we turn to final clusters, the picture is superficially a bit more complex. The clusters which can occur consist of a stop, preceded by a liquid or /s/ and/or followed by a liquid (/l/ or /r/). Before clusters of liquid plus stop, we find a short vowel (e.g., [lard] "fat"), while before clusters of stop plus liquid we find a long vowel (e.g., [sü:gr] "sugar"). As we observed before, liquids must be assumed to be fortis when followed by obstruents in order to account for medial clusters; in order to account for the short vowel in [lard], we need only assume that this applies to final ones as well, and that the neither the rule laxing final /r:/ to /r/ (since it is not final) nor the rule laxing final obstruents alters this state of affairs in final /rd/. As we see from the phonetic representation of [sü:g̥r̥], obstruents in final clusters undergo laxing even if followed by another consonant. The final laxing rule, thus, must be written as (24):

(24) [+obstruent] ---> [-tense] / ____ ([-syll]) #

This rule will correctly lax the /k/ in /sükr/, allowing the length rule to apply correctly here also.

The last case is that of final obstruent clusters. In forms like [mè:str] "master", [tǒ:st] "close", etc. we see that these result in preceding long vowels, with "true mid" quality. To account for the first fact, we need only assume that the entire cluster is lenis: this will allow the length rule to apply correctly. To achieve this result, we need to extend the scope of rule (24) yet further. If we do this by allowing for another optional final consonant, however, we might expect a convention of disjunctive ordering to block the correct application of the rule to the entire cluster. We can avoid this difficulty by formulating rule (25) in terms of syllable structure:

(25)



This rule has the effect of laxing an obstruent anywhere in the margin of a word-final syllable[22]. Assuming that the elements /sk/ in e.g. /pesk/ "fish" are underlyingly fortis, we derive the surface form [pè:z̥g̥] by first determining "true mid" quality for the vowel from the following fortis consonants, then applying rule (25) to lax them, and finally applying rule (18) to lengthen the vowel. Again we see that the vowel quality rule must precede final laxing (25), while the quantity rule (18) must follow it.

One other class of final consonants undergoes yet another development. In stressed monosyllables, final clusters of /r/ or /l/ plus velars generally undergo a rule of epenthesis, yielding forms like [ê:rax] "snow". The fact that the vowel in this form is of "low-mid" quality, which we would expect before a cluster of /r/ plus consonant, suggests that the rule (19) applies before epenthesis. The fact that the vowel is long, however, suggests that vowel length follows epenthesis (since epenthesis eliminates the cluster, and thus prevents /r/ from becoming [+tense] in preconsonantal position).

All of the data surveyed above converge on a single general conclusion: that despite the fact that the mid-vowel quality rules undoubtedly originated as rules sensitive to the difference between long and short vowels, in the modern language the quality rules must apply on the basis of underlying representations, while the quantity rule applies only after several intervening rules have operated. That is, quality must be assigned on the basis of whether the vowel is followed in underlying representation by a fortis or a lenis consonant (or, for rule (19), by one of a given set of segments and clusters); subsequently rules such as (25) (final laxing), epenthesis, /r:/--->[r]/__#, etc. apply; and only then can quantity be assigned. While mid-vowel quality thus originally depended in a natural way on vowel length (itself perhaps naturally based on syllable shape), it now depends on the articulatory properties of the following consonant in a way that apparently divorces it from its phonetic motivation. There is no particular reason to expect a correlation between vowel height and the property of tenseness in a following consonant, except as mediated by length.

The historical development of this state of affairs is quite straightforward. Originally, vowel length depended on syllable structure (or directly on the following consonant, perhaps as in English: cf. the discussion in section 2.0 above). Subsequently, a quality difference appeared in the mid vowels, related to quantity in a natural way. After this was established, however, other developments (sketched above) introduced some opacity into the relationship between quantity and quality. Since the developments surveyed above did not lead to any opacity in the relation between quantity and surface phonetic values, the quantity rule continued to apply to essentially surface forms. Quality, on the other hand, where predictable was dependent on underlying properties of segments, and so the rule was restructured so as to be stated as something like the following:

$$(26)\quad \begin{bmatrix} +\text{stress} \\ +\text{syll} \\ -\text{high} \\ -\text{low} \end{bmatrix} \text{---> ``true mid''} \;/\; \underline{\quad\quad} \begin{Bmatrix} \# \\ [+\text{syll}] \\ [+\text{tense}] \end{Bmatrix}$$

This rule assumes that potentially alternating mid vowels are represented basically with "high-mid" quality, while non-alternating ones are basically "true mid." It represents a sort of telescoping of the original quantity rule with the inverted form of the original quality rule, but has become quite independent of these (natural, phonetically motivated) processes. Furthermore, as we have taken pains to show in the analysis above, the interaction of this rule with the rest of the phonology makes it impossible to restore the naturalness of this rule by simply building in the history: the lengthening rule, which would make the quality rule a natural one, must follow it rather than preceding it.

This is exactly the sort of rule we sought: one which it is synchronically impossible (as far as we know) to found directly in the properties of some other domain external to language. Rather, its form and content result from the operation of principles of organization which are specifically characteristic of Language, rather than speech.

3. <u>The relation between underlying and surface forms</u>. Since the publication of Chomsky & Halle (1968), a good deal of work in phonological theory has been directed by the perception that this description (and thus the program that underlies it) is based on an unrealistically abstract concept of the relation between underlying phonological form and observable surface phonetics. It has generally been felt that one way to correct the perceived excesses of this "abstract phonology" is to build into phonological theory some set of constraints on the kind of relation that can exist between phonological and phonetic representations. Such constraints might take a number of tactical forms: they might be posited as limitations on the kinds

of effects phonological rules can have or the operations they can carry out, or on the kinds of underlying forms that can play a role in grammars, or directly as a requirement that underlying forms can differ from observed surface forms only in specific, limited ways. All of these approaches have been taken at one time or another in the development of phonological theory.

The issue of naturalness arises in connection with such proposed constraints on the form of grammars when it is suggested that they ought to have their origin in independent properties of systems (not necessarily specific to language) that are involved in language-learning: that they ought to follow in some direct way from the kinds of evidence that are available to language learners, e.g., or from the use they are (presumably) able to make of this evidence.

The proposition that characteristics of the learning system ought to be reflected in the structure of grammars has often been taken for granted. In one sense, this is obvious: short of positing telepathic communication or divine inspiration, it is hard to imagine how some aspect of a particular language could be learned if there were no evidence for it available to the learner. Of course, many aspects of language structure ("Universal grammar") are undoubtedly inherent in human cognitive organization (for a recent discussion of this issue, cf. Chomsky, 1980), and thus do not have to be learned. We are not concerned here with substantive or organizational principles common to all languages, however, but rather with the specific content of the grammar of a particular language. Though a child may not have to learn the principle of the cycle or of limitations on extraction in syntax (or the deeper principles from which these and other properties of grammars follow), he does have to learn the specific lexical items of his language and their representations. Thus, if a grammar is proposed in which (at least some of) these representations are assigned (non-universal) properties for which there does not appear to be an evidential base in plausible experience, we have reason to doubt the significance of the grammar. If a theory has the consequence in the general case that it allows for a wide range of such (apparently) unlearnable grammars, it is not unreasonable to ask how it could be constrained so as to avoid this consequence. One way to achieve this end would be to stipulate directly that the only grammars permitted by linguistic theory are ones that satisfy (in some way yet to be specified) the requirement of "learnability."

The importation into linguistic theory of constraints derived from general learning strategies is thus not unmotivated, but neither is it unavoidable. That is, if we observe that the data normally available to the language learner have a given characteristic which would have the consequence that certain representations are unlearnable, we might constrain the theory directly so that the only grammars it allows are those consistent with the evidential limitation in question. Alternatively, we might leave the theory of language per se unconstrained, allowing both for the systems that can be learned and also for those which apparently cannot. The fact that grammars in this latter class are not in fact learned, insofar as it is true, would then be attributed not to the nature of Language, but rather to the interaction of the cognitive system particular to Language with others that are not so specific. This approach is a natural one on the "interacting systems" approach sketched in section 0 above, but of course it is a completely incoherent one on the view that the properties of language are all to be attributed to other domains, and involve no distinctive language faculty at all.

We may now ask whether the issue posed here has any real consequences for linguistic theory. After all, whether a given property of observed grammars is to be attributed to the language faculty or to some other system does not absolve the linguist from recognizing it and formulating as accurate as possible a characterization of it. However, if linguists are centrally interested in the nature of Language, they ought to be interested in going beyond such a characterization (wherever possible) to determine whether or not the property in question is a linguistic one. In order to do this, a situation of the following sort would be necessary. Suppose we have isolated a property of normal language-acquisition data which appears to place limitations on the class of grammars that can be learned. If this property is reflected in the theory of Language, grammars not having it ought to be absolutely unlearnable (at least as anything like normal languages). Now suppose we can find some non-standard circumstance in which the apparent limitation in question could be overcome, removing the impediment to learning. If the original constraint was in fact appropriate, this ought to make no difference: the grammars newly made "learnable" should still be inaccessible. If, on the other hand, the constraint in question was strictly a property of the learning system, and not an aspect of the nature of Language (not, hence, a part of universal grammar), the wider class of grammars ought to be available under these changed circumstances.

It appears that evidence of the sort just sketched can indeed be found in some cases, and that it supports the notion that constraints on the linguistic system per se are in principle independent of such "functional" explanations from other domains. This is the opposite side of the coin from the argument given by Halle (1979) that where functional explanations seem adequate to the facts of language, they are nonetheless unnecessary in many instances, since the same observed limitations follow from independently necessary "formal" principles that have no such general external motivation (i.e., from organizational principles specific to the nature of Language). Together, these two lines of argument establish the claim that (since available functional explanations are apparently neither necessary nor sufficient to account for the properties of Language) the language faculty is governed by an organization essentially independent of other domains.

To support this claim, we will examine a sort of proposed constraint (due originally to work by Kiparsky) on possible relations between underlying and surface forms. We will see that (though difficult to formulate in precise terms), such a constraint has a plausible relation to the nature of the language learning task, and thus a potential functional explanation. We will then go on to examine circumstances in which limitations on evidence are reduced, as suggested above, and conclude from the consequences that the proposed constraint is not in fact a constraint on grammars but on learning strategies.

3.1 Kiparsky's alternation condition and its descendents. In an attempt to remedy the apparently extravagant excesses of abstractness permitted by the phonological theory of Chomsky & Halle (1968), Kiparsky made a series of proposals (treated most extensively in Kiparsky, 1973, although the discussion has continued since then) concerning the types of rules and of representations that ought to be allowed in grammars. These included the "alternation condition" - a prohibition (perhaps only relative to available alternatives) of rules of absolute neutralization - and the proposal that some rules (described by Kiparsky as "non-automatic neutralization rules") should only be allowed to apply to forms that are "derived", in the sense of including relevant material not all present in the underlying representation of some single formative.

These proposals constituted an important and widely accepted step in the development of the theory, and stimulated an extensive discussion and a number of alternative proposals. Their attractiveness can perhaps be said to center on the consequence that forms which do not show alternations must be assigned underlying representations which do not differ (except in matters of "phonetic detail") from their surface forms. When this notion is extended from the class of absolutely non-alternating forms to more "local" instances (i.e., to forms which do not show some specific alternation, though they may involve others), it is apparently equivalent to some sort of principle along the lines of "What you see is what you get." We assume, that is, that the language-learner does not hypothesize an underlying form distinct from the surface form he hears unless he has some positive evidence to support doing so; if the essential form of such evidence is alternations, it ought to follow that abstract relations between underlying and surface forms (i.e., ones not supported by the evidence of alternations), such as are prohibited by Kiparsky's constraints, should not be possible in language.

An exact formulation of the constraint in question has proven quite difficult to arrive at. Consider the proposal that "non-automatic neutralization rules only apply to derived forms." This is intended to have the consequence that a rule never applies entirely internal to a single formative, except in the case where some other external influence has resulted in a (relevant) change in its representation. The basis of this is the feeling that, where the conditions for applying some rule are all entirely present in the lexical representation of the formative, the consequences of such an application ought to be incorporated directly into the underlying representation. Allowing a phonological form which will always be converted into some (distinct) surface shape would seem to require an illicit inference on the part of the language-learner, one for which he would have no evidence in the properties of the word to which it is to be applied. It would seem reasonable to prohibit the construction of grammars which rest essentially on such inferences.

There are, however, cases which show that this formulation does not have the desired consequences. One such example involves a rule in the phonology of Faroese (cf. Anderson, 1974, and references cited there for fuller discussion and motivation of the rules in question; cf. also Hallberg, 1980). Essentially, this rule deletes the consonants /g/ and /d/ when they stand in intervocalic position; the resulting hiatus is generally broken up (depending on the vowels in question) by the insertion of a semi-vowel. For example, the stem /hug/ "thought" can be followed by the nominative singular ending /+ur/; the resulting form /hug+ur/ undergoes /g,d/-deletion to give /hu+ur/, to which glide-insertion applies to give (ultimately) [hu:wUr]. The presence of a /g/ in this stem is confirmed by the related verb <u>hugsa</u> [huksa] "(to) think". The deletion rule is clearly neutralizing (it neutralizes stems with /g,d/ and those without); it is also "non-automatic", in that it has exceptions in a few words like <u>sýnagoga</u> [sUynago:ga] "synagogue". It ought, therefore to be subject to the constraint in question.

If we take this step, however, we will be unable to derive the alternation in e.g. fagur/fagran [fa:vUr]/[fagran] "beautiful (m.nom.sg./m.acc.sg.)", from the stem /fagur/. The form fagran involves the ending /+an/; in the resultant /fagur+an/, a rule of syncope applies to delete the vowel /u/. As a result, the /g/ in fagran is not intervocalic, and is not deleted. In fagur, no ending is added; the /g/ is thus intervocalic and deletes (to be replaced in hiatus by [v], the eventual product of glide-insertion). The problem here is that the application of /g,d/-deletion is entirely internal to the formative /fagur/, which is unchanged from the underlying representation and thus in no relevant sense a "derived" form. The constraint in question would thus block deletion here, but that would lead to the incorrect result *[fa:gUr].

It is easy to see what characterizes this example: despite the fact that /g,d/-deletion must apply entirely internal to the unchanged representation of a single formative (/fagur/), there are other forms in the language (e.g., fagran) which show that the /g/ should be present in this item, despite its non-appearance in the surface form of fagur. There is evidence, that is, in the form of an alternation between /g/ and /∅/ for positing the "abstract" representation /fagur/ for the form fagur. This example makes it clear that it will be difficult to formulate the desired restriction on underlying forms indirectly in the form of a constraint on the applicability of certain rules (although Kiparsky provides excellent evidence that some such constaint must apparently be operative in at least some cases).

3.2 Evidence against the linguistic charcter of the condition. The fact that a constraint of the sort discussed in section 3.1 has not yet been successfully formulated, however, does not eliminate the appeal of attempts to make it impossible to incorporate into a grammar features for which the evidence available to the learner is generally thought to be inadequate. Regardless of the exact formulation of the relevant constraint, the question still arises: is it appropriate to look for such a principle as an element of linguistic theory? Or is there evidence which would show us that the apparent constraint results in fact from the interaction of the properties of Language with those of other domains?

3.2.1 Evidence from "spelling pronunciations". When we consider the issue seriously, indeed, we can see that there is a class of prima facie exceptions to such a requirement that phonological representations are always inducible from the data available to the learner in the form of the phonetic shapes of linguistic elements, together with their patterns of alternation. This evidence is typified by spelling pronunciations, usually (and, we would assert, incorrectly) written off as linguistically irrelevant behavior on the part of a small fraction of the native speakers of a few natural languages.

Consider the rule of vowel reduction in English, for example. This rule results in the neutralization of (most of) the vowel qualities that appear in syllables under stress to a mid central unrounded vowel varying considerably in exact quality, usually described as "schwa", in completely unstressed syllables. This rule is responsible for a large number of alternations in formatives that appear under more than one stress pattern: consider, for example, derive vs. derivation, phonology vs. phonological and a wealth of others.

In those cases where we have an actual alternation in vowel quality between some full vowel and schwa, it is clear that we attribute the full vowel to the underlying form, and reduce it to schwa if it does not receive stress. What of those words in which some vowel is <u>always</u> unstressed, however? In the absence of any alternation, it would seem that a principle of the type we have just been discussing requires us to posit an underlying schwa, since we could have no direct evidence for any particular vowel quality other than this. This conclusion, since it reduces what would otherwise be a considerable problem of non-uniqueness in representations, has generally been accepted.

However, when we are called upon to create a new form on the basis of one of these words with non-alternating schwa, this often results in a shift of stress onto the syllable with schwa. Suppose we have a friend by the name of Fulton, for example. The second syllable of this name is always (for many speakers) unstressed, and thus schwa; and we would therefore (on the lines sketched above) represent it underlying with this neutral quality. If our friend establishes a political party, however, and egotistically dubs it the Fultonians, we have no hesitation in pronouncing this as [fUltownijənz], rather than [fUltuwnijənz], [fUltInijənz], etc. This suggests that we have in fact assigned the second syllable of Fulton the representation /o/, rather than schwa.

Of course, the explanation for the facility with which we can reconstruct a full vowel quality for a given non-alternating schwa (at least in many, if not most cases) is not far to seek: the word is spelled with an <u>o</u>, not a <u>u</u>, <u>i</u>, etc. Accordingly, a form such as [fUltownijən] is generally simply characterized as a spelling pronunciation. Such pronunciations are usually thought of as a derivative phenomenon peculiar to the speech of literate people in a minority of the world's linguistic communities, of little if any interest for linguistic theory. It is important, however, to ask just what spelling pronunciations signify.

It is possible that our ability to construct spelling pronunciations is strictly speaking non-linguistic in character, similar to some other sorts of behavior which, while resembling language, are obviously based on rather different capacities than normal language use. For example, most of us can train ourselves to read a phonetic alphabet (like IPA), and to translate a text in this system into reasonably fluent vocalization. If the text happens to have been constructed on the basis of a language which we can speak, what results may be phonetically identical (or nearly so) with an utterance in that language: perhaps one which we might produce spontaneously. There is little reason to believe on that account, however, that our production of the text under these conditions involves essential reference to our competence in our native language. Reading phonetic symbols, that is, is not really linguistic behavior in a central sense.

If spelling pronunications had a similar character, there would be no reason (for linguists) to pay any special attention to them. To conclude this, however, is to do something of an injustice to the facts. It is sometimes suggested, for example, that spelling pronunciations are irrelevant because they are primarily a phenomenon of adult speech: obviously, preliterate children (and thus the bulk of the language-learning population) do not have access to orthographic information in constructing their grammars. Does this mean, then, that adult language isn't "really" language at all, and that only the behavior of children who have not yet completely learned their native language (in the sense of having acquired all of the vocabulary items and other idiosyncracies of it which they will eventually learn) can be considered relevant to the study of language? Simply to state this position is to render it untenable. Obviously the particular process by which very young children go about constructing a grammar is of particular interest for linguistic study, but there is no great qualitative step between learning lexical items by hearing them spoken and learning them in other ways (e.g., by reading them): certainly there is no obvious basis for denying *a priori* that what we learn after the first dramatic steps of acquisition is still part of our language.

Spelling pronunciations, in particular, seem to have much the same character as other instances of the acquisition of lexical knowledge. The case of vowel reduction is a typical example: note that, in saying that perhaps we internalize our friend Fulton's name with an underlying /o/ (which always is reduced to schwa, prior to his entry into politics), we are not describing a completely arbitrary, unsystematic fact. Indeed, the rule of vowel reduction in English which supports this representation is well motivated by a host of cases in which alternations are available to support it; all we are doing is adjusting the lexical representation of a single item in accordance with this system, exactly as we would if we arrived at the same conclusion by hearing the name of his new party *before* learning to spell his name. Typically, we do not find spelling pronunciations which are completely unrelated to the phonological system of the language: we do not, for example, hear people replacing [fIzIkəl] by [p^{h}IzIkəl], since there is no support in English for a process relating [f] and [p^h]. There is, thus, no reason to deny a linguistic character to (at least many) spelling pronunciations, and indeed to do so would be to leave unexplained the close relation between them and the rest of the system of the language.

This is not to deny that some spelling pronunciations are genuinely sporadic, unsystematic phenomena. Even these may be significant: where a speaker apparently produces a "spelling pronunciation" which does not reflect the (normative) correct spelling of a word, it may still support the claim that he has internalized some representation that is not equivalent to the surface form, perhaps on the basis of the general possibilities inherent in the orthographic-phonological correspondences of the language. There may well still be a residue of cases in which "spelling pronunciations" are a strictly conscious phenomenon, not related to any systematic rules of the language; these, of course, would be of little linguistic interest.

Nonetheless, there is a substantial body of factual evidence from this source which has important consequences for the nature of the limitations that the language faculty as such imposes on possible grammars. In order to understand the weight of this evidence, we must of course separate those facts of spelling pronunciations and related phenomena which we attribute to language use from those which represent other sorts of behavior. It remains to be seen how this can be accomplished in general, though certain central cases seem clear. Before such a task can be undertaken, however, it is necessary to take seriously the possibility that this area has linguistic significance.

It would seem to be a mistake, then, to write off as irrelevant the facts of "mere" spelling pronunciations. Rather, they suggest that in arriving at the underlying form of a word, we are able to make use of whatever information is available (including, but not limited to, the patterns of alternation in which the word participates). It seems plausible that our initial construction of the phonology of our language is based on these alternation patterns, but (if we learn to write our language) we may later learn that there are certain (semi-)systematic correspondences between the significant sound properties of lexical items and their representation in the orthographic system. On the basis of this knowledge, when we encounter a new word for which we have not yet the full range of information that is (potentially) available from patterns of alternation in which the word could appear, we may make use of the clues given by the orthography in deciding what its sound properties (and thus its phonological representation) should be. This in turn suggests that the possibilities for phonological systems are not limited to ones that can be completely inferred from the properties of a given corpus (e.g., the forms available to a child language-learner). But if this is the case, it is inappropriate to attribute to phonological theory (and thereby to the system of Language) a constraint whose plausibility derives entirely from the limitations of the normal case in comparatively early language acquisition.

3.2.2 Evidence from dialect contact. To show that such transcendence of the limitations inherent in the paradigm case of first language learning is not limited to the (perhaps artificial) circumstances characteristic of highly literate societies, and thus provide some perspective on the point at issue, we can consider some facts which come from a community in which literacy is not a relevant consideration. Again, we can find reason to believe that data going beyond that available in normal first-language learning at an early stage is nonetheless potentially useable in grammar construction. The Damagaram dialect of Hausa spoken in Niger[23] has undergone a change by which the /f/ of standard Hausa has been replaced by a labialized laryngeal ([h^w]). Standard Hausa does not have [h^w], so this change would affect simply the pronunciation of a single segment type if nothing else were involved. However, the Damagaram dialect has also acquired a rule merging /h^w/ with normal unlabialized /h/ when the segment in question is not followed by a low vowel. Where a stem contains an instance of /h^w/ which may (depending on morphological circumstances) be followed by either a low or a non-low vowel, this results in alternations such as those in (27):

(27)	a. tuhwa/tuhwahii	"clothing (sg./pl.)"
	b. k`ahoo/k`ahwannii	"horn (sg./pl.)"
	c. k`ahwa/k`ahoohii	"foot (sg./pl.)"
	d. littaahi/littattaahwai	"book (sg./pl.)"
	e. tsoohoo/tsoohwaa	"old (masc./fem.)"

Alternations such as those in (27) often give an indication that a particular instance of surface [h] is to be regarded as coming from underlying /h^w/, but such evidence is lacking for many other forms. There are also instances of [h] which do not alternate with [h^w] even before low vowels, as in words like (28):

(28)	a. har (*hwar)	"until"
	b. wahala (*wahwala)	"trouble"

Words like those in (28), which are not at all unusual, must be regarded as having underlying /h/; the words in (27), on the other hand, have underlying

/h^w/. When a word contains [h] in a position where it will always precede a non-low vowel, this [h] will never alternate with [h^w]. On the assumption that a constraint of the sort discussed in section 3.1 is valid, we would expect all such non-alternating [h]'s to be represented by /h/ obligatorily.

Now in fact, although a non-alternating [h] could never provide direct evidence for a derivation from anything other than /h/, there is another, indirect source of evidence available to speakers of Damagaram Hausa. Most speakers of this dialect are exposed frequently (on the radio, in the marketplace, etc.) to speakers of standard Hausa, with which their dialect is mutually comprehensible. Insofar as they understand standard Hausa words containing [f] as corresponding to their own forms with [h^w] (before low vowels) or [h] (before non-low vowels), they have an additional source of data on the underlying form of some instances of surface [h]. If they have heard the corresponding word in standard Hausa with [f], that is, they might infer underlying /h^w/ even though they had never heard a phonetic [h^w] in the word (because its [h] failed to occur in the right environment) - a sort of non-literate analog to spelling pronunciation.

We might well ask, however, how we could ever tell that a particular non-alternating [h] was lexicalized as underlying /h^w/ if it never occurred before a low vowel: even if it is underlying /h^w/, it will always be surface [h]. However, this dialect (in common with other forms of Hausa, though the details differ somewhat) contains another rule which gives us a clue. In fact, short high vowels in non-final syllables are subject to variation in backness depending on the character of surrounding consonants (among other things). In no such environment is the difference between [i] and [u] contrastive. In some cases the two qualities are freely variable (e.g., tina/tuna "think"); in other cases one or the other quality is obligatory (e.g., rihe *ruhe "close, cover"). The case of interest to us, however, arises when the consonant (or consonant cluster) on either side of such a non-final short high vowel contains a rounded segment: under these circumstances [u], and not [i], is obligatory (e.g., ruŋhwa, *riŋhwa "roof"). Interestingly enough, this is true even when the rounded segment in question is itself unrounded by the rule we have already seen converting /h^w/ to [h] before a non-low vowel. Thus, we have in some cases an extra clue to the underlying value of a given consonant even where data derived from alternations is absent.

In at least some examples, the type of evidence we have just suggested is available to support the hypothesis of "spelling-pronunciation" like behavior raised above. The name of the Fula people in standard (Kano) Hausa, for example, is filaanii, with initial [f] and front (not back) quality to the short high vowel of the first syllable. The [f] in this form, of course, will appear as [h] in Damagaram Hausa, since it is followed by a non-low vowel. Since this vowel is non-alternating, the [h] will also be non-alternating - and the constraint of section 3.1 would suggest that it therefore ought to be represented as underlying /h/. However, the actual form is hulaanii, suggesting that the initial segment is really /h^w/ rather than /h/: information that could apparently be derived easily from the corresponding form in the standard language, but not from other sorts of data available in the alternation patterns of the language.

If this is indeed the correct explanation for these facts, we have here another instance of the possibility of speakers' employing information other than the central facts of alternations to go beyond the limitations imposed by a learning situation in which they only have access to this type of data. The fact that they can evidently do so successfully suggests in turn that phonological systems are not directly subject to a constraint of this character. If a speaker (or perhaps all speakers of a given language) can be shown to have a grammar which is consistent with such a limitation on the relation between underlying and surface forms, this need not be because such a condition is an essential part of the nature of Language: rather, it may simply be that their access to evidence to use in constructing their grammar has been limited, and has not included any (relevant) facts that would allow them to transcend these limitations, such as systematic orthographic representations or the properties of cognate forms in another systematically related dialect. The fact that the constraint is a natural consequence of the conditions imposed on learning, that is, does not mean it is a constraint on Language, but only that it is likely to characterize the class of accessible grammars which actual speakers subject in practice to such limitations will in fact construct. When the limitations are removed, the apparent constraint disappears.

4. Conclusion. We have surveyed above a number of questions that bear on the same general issue: to what extent can we expect the properties of Language to reflect those of other domains? In each case, we conclude that the relationship (insofar as it exists) is an indirect one, and that an adequate account of the phonological systems of natural languages must accord a central role to a set of principles that have no direct foundation in extra-linguistic considerations.

In considering the nature of phonological representations (not only the most abstract of these, but in fact any representation conforming to the same overall organizing principles), we concluded that it is not possible to equate the system of distinctive features and other central constructs with concrete physical aspects of the phonetic realizations of utterances. Some phonological distinctions apparently do not always have a direct reflection in phonetic substance, while a system of features which is adequate to the description of sound patterns may yet leave unspecified important components of the sounds themselves. Since phonological representations thus cannot be considered to be isomorphic with phonetic reality, but instead include a significant component which is part of a more abstract cognitive system, they cannot be said to be "natural" in the sense of following from other aspects of the nature of human speakers.

Similarly, when we moved on to consider the character of phonological rules, we saw that there is good reason to believe that these often do not directly reflect processes that have their motivation in the properties of the articulatory, acoustic, or general auditory systems of the human organism. Processes particular to the nature of Language shape the rules which occur in particular languages in ways that are not based on extra-linguistic considerations. An attempt to confine phonology to the study of "natural" rules, or even the expectation that "serious" rules in the phonology of any language ought to be "natural" in their content can only result in the unmotivated exclusion of much of what is most essential to the sound patterns of natural languages.

Finally, when we considered the nature of general formal constraints governing the organization and operation of grammatical systems, we saw that there is no reason to expect that these will directly mirror the properties of other, not specifically linguistic cognitive systems. Thus, although the nature of experiential learning as a general problem imposes certain limitations on the character of the linguistic systems and relations that can be learned under ordinary circumstances, it is not correct to translate these limitations into constraints on the language faculty per se: where the specific conditions of normal learning are not met, the limitations that might be attributed to them disappear.

Thus (at least as phonology is concerned, though there is no reason not to expect this conclusion to extend to syntax and other aspects of Language as well) we see that, in important respects, Language involves principles that are not "governed by forces implicit in human vocalization and perception." Language is not simply a "reflection of the needs, capacities, and world of its users"; it is also the product of a distinctive cognitive faculty, for which we have no reason to expect literal analogs in other human capacities or constraints. In this sense, then, phonology is clearly not "natural."

If we were to interpret this conclusion as showing that the content of linguistic systems is intrinsically arbitrary, or that the study of other domains (such as phonetics) is irrelevant to an understanding of Language (specifically, phonology), this would obviously be thoroughly improper. In each of the areas we have considered above, we have found reason to believe that, although the language faculty is an intrinsically autonomous aspect of human mental organization (at least, so far as we know at present), the specific content of particular linguistic systems is heavily based on other considerations. Thus, although features are not simply abstractions of particular phonetic parameters, they are realized through (perhaps complex combinations of) phonetic dimensions; although phonological rules are by no means limited to "natural processes," they typically originate in the phonologization of such processes, and thus their subsequent restructured forms often relate to such a basis; and while linguistic constraints proper need not incorporate the restrictions of other cognitive structures, nonetheless these other structures, insofar as they are involved in particular instances of the formation of language systems, can be expected to have their own consequences.

In each case, then, we see that the character of a linguistic system depends on the interaction of particular substantive considerations which are not specific to language with an irreducible component which is. We can see how important these external considerations are when we consider the nature of natural languages that are independent of the specific medium of speech: particularly the study of manual languages (cf. Klima & Bellugi, 1979 for an excellent summary of these issues) can yield a valuable perspective on the interplay of "substance-based" and purely linguistic considerations in forming the detailed properties of a language. Obviously we cannot understand actual languages without understanding speech (as well as other relevant aspects of human organization and activity), but no more can we understand it if we confine our attention to this side of its reality.

Much of recent research in linguistics has been polarized between an approach which aims to reduce Language to its substance (thus showing that it is "natural") and an approach which intends to reduce it to its form. Both of these programs of research are motivated by the goal of explaining why languages are as they are. Explanations, however, typically consist of connections between one set of principles and another; and in order for such connections to provide an exhaustive account of a set of phenomena, we have to ensure that the domain in question is a closed one (i.e., not subject to significant influences from without). In the study of Language, however, such a condition is rarely if ever met: properties from a variety of distinct systems, including some that are peculiar to linguistic objects and others that are not are relevant to a comprehensive view of Language. As such, we should not really expect explanations of a strictly predictive sort to be forthcoming. We should recognize that when we can understand (*ex post facto*) why things are as they are, we have advanced our knowledge of them: we should not require in addition that we show that they could not have been otherwise. This notion of the "exegetic adequacy" of a theory of Language (for discussion of which, see Anderson, 1979) seems well worth pursuing, in the face of the apparent fact that Language cannot be satisfactorily reduced either to its form or to its content. If we wish to understand what is in nature, we cannot do so by requiring that it be "natural" in some *a priori* sense.

Footnotes.

[*] This work was supported in part by grant number BNS78-07680 from the National Science Foundation to the UCLA phonetics laboratory. I would like to acknowledge the many helpful ideas and comments I have gotten from colleagues in presenting this material at various stages. Particular thanks (and the usual disclaimers) are owed to: Charles-James Bailey, Vicki Fromkin, Osamu Fujimura, Morris Halle, Richard Janda, Pat Keating, Peter Ladefoged, Will Leben, Mark Liberman, and John Ohala.

[1] In Anderson, 1980, an analogy is drawn between the trend represented by various forms of "natural" generative phonology and the intuitionist response to Whitehead & Russell's *Principia Mathematica* in the history of the philosophy of mathematics. In both cases, the retreat from abstraction which results from limiting the range of entities posited to constructible ones has rather interesting consequences, but must be adjudged a failure as far as providing a secure basis for the construction of the entire field in question.

[2] Actually, this proposition (which is essential to the force of Ohala's explanation of the behavior of nasal assimilation in terms of the steady-state acoustic characteristics of the nasal itself) is somewhat dubious. Such evidence as there is (cf. Zee, 1979, and references cited there) suggests that transitions are more significant perceptually than the steady-state nasal murmur in identifying place of articulation for nasal consonants. Ladefoged (1968) shows that the transitions associated with labiovelars are much more like those of labials than velars; one would therefore expect that the acoustic properties of a labiovelar would tend to result in its identification with labials, rather than with velars as Ohala claims. If there is an external explanation for the assimilatory behavior of nasals before labiovelars, then, it may well be an articulatory one (or from some other domain entirely).

[3] In Anderson, 1976b, it is shown in detail that this analysis is to be preferred to alternatives taking some other consonant series as basic.

[4] The forms (*bar*) and (*gar*) in (3) are parenthesized because they happen only to occur in derivational forms of these verbal stems; the prenasal variants *mbar* and *ngar*, however (which suffice to establish the consonant series with which /w/ alternates in each case), appear within the same inflectional paradigm as the basic form of the stem. Had we chosen nominal stems instead of verbal ones, all three "gradation" variants could easily be attested within inflectional paradigms alone, but this particular minimal pair would not be available. For further information on the operation and function of the consonant alternation system in Fula, cf. Anderson, 1976b.

[5] For a substantial survey of results in this area, cf. Lehiste, 1970.

[6] I am grateful to Pat Keating for pointing this result out to me; the conclusion is based on work carried out by Flege. Related results, including the apparent fact that Saudi Arabic speakers tend to transfer the distribution of length as a function of the following consonant to

English when they learn this language, are reported by Flege & Port, 1980.

7 A few phonetic details irrelevant to this discussion have been simplified in the transcriptions that follow. It should also be noted that before "back" vowels (in the sense of vowels that do not cause velar fronting, as will be discussed below), a contrast exists between front and back velars. In all such cases, however, the front velars can be shown to be due to the presence of a front unrounded semivowel /j/ between the velar and the vowel; this causes velar fronting, and is then absorbed by the resulting palatal consonant. No such contrast is possible before "front" vowels (i.e., those that cause velar fronting). We ignore this set of facts below, since it does not appear to bear on the question of the degree of phonetic motivation of the velar fronting rule.

8 A number of works by this frequently ignored figure in the early development of phonological theory have been published in English translation by Stankiewicz (1972). Baudouin de Courtenay (1895) is a particularly important survey of his views concerning the nature, source, and evolution of phonological alternations and the principles that govern them. It would obviously be anachronistic to attribute to Baudouin a concern with some of the specific issues concerning phonologization, opacity, morphologization, etc. discussed below; indeed, there is no reason to believe that he construed the alternations he discussed in terms of a notion of "rule" directly comparable with that current today. His views prefigure to a remarkable extent, however, much recent discussion of these topics.

9 Lopez (1979) provides a very useful study of the nature of lexical correspondences in the phonology of Brazilian Portugese, and of the difference between these redundancy-rule-like statements and genuine (morpho-)phonological rules.

10 I am indebted to Peter Tiersma for drawing this example to my attention.

11 There are also at least some examples of compensatory lengthening resulting from the deletion of prevocalic consonants (cf. deChene & Anderson, 1979, for discussion). I assume that these follow in an obvious way from an account symmetrical with the one developed below, and do not explicitly treat them below. Other examples of "compensatory lengthening" that are traditionally grouped together with these are those in which length results from the loss of an entire syllable. The explanation below does not presume to cover these (if indeed they represent a unitary process type), but it is not obvious that any alternative has any better success in accomodating them in the same framework with the cases involving loss of a consonant adjacent to the lengthened vowel.

12 In the assumption that internal constituent structure of syllables is appropriately represented by a labelling such as that in (10), I agree with the recent discussion of Safir (1979). It should be noted, however, that much other recent work based on the assumption that syllables themselves have structural significance (e.g., McCarthy, 1979) has proceeded on the view that only labels indicating relative prominence (w[eak] vs. s[trong]) are relevant. There are important

issues of principle here, as noted by Safir (1979) and Monahan (1979), but they are difficult to address conclusively.

13 The description below is limited to the Léon dialect of Breton, and specifically to the subdialect described by Falc'hun (1951). It would also extend with minimal modification to the closely related dialect treated in the classic description by Sommerfelt (1921). The Léon dialect is the primary basis of the literary standard language, but it is important to emphasize that there is a great deal of variation across the Breton speaking area. The dialects centering around Vannes, in particular, differ markedly from the system described here. For a survey of the dialects and their history (though without much of the information that has been published recently on vannetais dialects), cf. Jackson, 1967.

14 We continue to refer simply to "high-mid", "true mid", and "low-mid" qualities below, without taking a stand on how these should be differentiated. Obviously, a complete description might have to take this issue into account.

15 The symbol [f̌] represents a voiced labiodental fricative, differering from [v] in having a substantially higher rate of airflow, and correspondingly less prominent vocal cord vibrations. From the description given by Falc'hun, based on kymograph tracings, it would appear that this segment actually involves the laryngeal configuration usually found in voiced aspirates, as opposed to the normally voiced obstruent configuration found in [v]. While the resulting articulatory type would be essentially isolated in the language (and indeed, quite rare in the languages of the world), there seems substantial unanimity of opinion on the part of a number of independent investigators that something like this is involved.

16 The symbol [h] actually represents a segment which is voiced in most Breton dialects; it serves as the fricative correspondent of [g], or the voiced correspondent of [x] despite its laryngeal (rather than velar) point of articulation.

17 The symbols [ñ] and [l̆] represent palatal nasal and lateral segments, respectively. These generally arise from clusters of dental [n] or [l] plus a semivowel [j]; they are also found in some borrowed French words with "n-mouillé" or "l-mouillé", and in a few dialects can arise through palatalization of /n/ or /l/ after (not before) a high front vowel.

18 The quality of the final syllable [êr] in this form may well be due to the rule mentioned above (discussed in Anderson, 1974) which communicates the quality of a stressed mid vowel to mid vowels in adjacent syllables. However, it is quite generally the case that final unstressed /-er/ is phonetically [-êr].

19 In fact, the generalization stated by this rule is not an accurate characterization of the language, as will be pointed out below.

20 In transcriptions up to this point, we have in general written final obstruents as fortis or lenis depending on their underlying value,

ignoring the merger of these two series in (typically voiceless) lenis segments finally. Except where our attention is specifically devoted to the properties of final segments, we will continue this practice below. In order to obtain accurate phonetic representations, of course, it would be necessary to apply the rule of "final lenition", which will be stated in due course below.

21 It is interesting to note that there is an exception to the neutralization of final obstruents as lenis. A verb followed directly by a complement may optionally behave as if its final consonant were word-medial, rather than final. From the verb [kasãn] "I send", for instance, we would expect a third singular form [ka:z̥]. We do indeed find this form, but when "he sends" is immediately followed by an object, for instance, we may get a phrase like [kas ãn amé:zog̥] "send the neighbor" - which shows a medial voiceless fortis, and thus forms a minimal pair with [ka:z ãn amé:zog̥] "the neighbor's cat." It is apparently the case that this optional treatment of verbs is similar in character to liaison in French. As Selkirk (1972) has shown, this can be described in large part by a rule demoting certain phrase-internal boundaries. In fact, Selkirk posits for French a rule which is quite similar in content to the one we would need here: a rule which reduces ## to #_ (or perhaps +) between a V and a following complement internal to the $\bar{V}$ constituent (in the terms of an $\bar{X}$-theory of phrase structure).

22 Jackson (1967:50) implies that Falc'hun has shown that the segments in final /s/ plus stop clusters are fortis, not lenis. In fact, Jackson's only evidence for this appears to be the fact that they remain voiceless even before a sonorant initial word. Since voicelessness is not equivalent to fortisness in Breton, however, this argument does not hold: as we see, the rule for assigning voicing in this environment may well have the following form:

i. [+obst] ---> [+voice] / [-obst] ___ # [-obst]

In general, Breton obstruents are only voiced when surrounded on both sides by sonorants; this condition also comes into play in describing the internal structure of compounds in the language. In this case, obviously, Jackson's conclusion would not hold; and indeed, Falc'hun (1951) treats the fact that final clusters still induce length in preceding vowels as an argument in favor of treating them as lenis, not fortis.

23 I am indebted to Laurie Tuller of UCLA for presenting the following facts in an unpublished manuscript, and for drawing my attention to the example and its significance.

References:

Anderson, Stephen R. (1969). "An Outline of the Phonology of Modern Icelandic Vowels." Foundations of Language 5:53-72.

Anderson, Stephen R. (1974). The Organization of Phonology. New York: Academic Press, Inc.

Anderson, Stephen R. (1976a). "On the Description of Multiply-Articulated Consonants." Journal of Phonetics 4:17-27.

Anderson, Stephen R. (1976b). "On the Description of Consonant Gradation in Fula." Studies in African Linguistics 7:93-136.

Anderson, Stephen R. (1978a). "Tone Features." in V. Fromkin (ed.), Tone: A Linguistic Survey. New York: Academic Press.

Anderson, Stephen R. (1978b). "Syllables, Segments and the Northwest Caucasian Languages." in Bell & Hooper (eds.), Syllables and Segments, pp. 47-58. Amsterdam: North-Holland.

Anderson, Stephen R. (1979). "On the Subsequent development of the `Standard Theory' in Phonology." in Dinnsen (ed.), Current Approaches to Phonological Theory, pp. 2-30. Bloomington: Indiana University Press.

Anderson, Stephen R. (1980). "Notes on the Development of Phonological Theory." Language & Speech, 23:115-122.

Bach, Emmon, & Robert T. Harms (1972). "How do Languages Get Crazy Rules?" in Stockwell & Macauley (eds.), Linguistic Change and Generative Theory. Bloomington: Indiana University Press.

Baudouin de Courtenay, Jan (1895). "An Attempt at a Theory of Phonetic Alternations." (translation of Versuch einer Theorie phonetischer Alternationen; ein Kapitel aus der Psychophonetik) in Stankiewicz, 1972, pp. 144-212.

Bjerrome, Gunnar (1957). Le patois de Bagnes (Valais). Stockholm: Almqvist & Wiksell.

Chen, Matthew (1970). "Vowel Length Variation as a Function of the Voicing of the Consonant Environment." Phonetica 22:129-159.

Chomsky, Noam (1980). Rules and Representations. New York: Columbia University Press.

Chomsky, Noam, & Morris Halle (1965). "Some Controversial Questions in Phonological Theory." Journal of Linguistics 1:97-138.

Chomsky, Noam, & Morris Halle (1968). The Sound Pattern of English. New York: Harper & Row.

deChene, Brent, & Stephen R. Anderson (1979). "Compensatory Lengthening." Language 55:505-535.

Disner, Sandra F. (1978). Vowels in Germanic Languages. UCLA Working Papers in Phonetics, vol. 40.

Donegan, Patricia J., & David Stampe (1979). "The Study of Natural Phonology." in Dinnsen (ed.), Current Approaches to Phonological Theory, pp. 126-173. Bloomington: Indiana University Press.

Drachman, Gaberell (1969). "Adaptation in the Speech Tract." in Papers from the Fifth Regional Meeting, Chicago Linguistic Society (pp. 314-329).

Einarsson, Stéfan (1945). Icelandic. Baltimore: Johns Hopkins University Press.

Falc'hun, F. (1951). Le système consonantique du breton. Rennes: Plihon.

Fant, Gunnar (1970). "Sounds, Features, and Perception." in Proceedings of the Sixth International Congress of Phonetic Sciences. Prague: Academia.

Flege, James Emil, & Robert F. Port (1980). "Cross-Language Phonetic Interference from Arabic to English." Research in Phonetics, Report # 1 (pp. 99-136). Bloomington: Indiana University Dept. of Linguistics.

Fowler, Carol A. (1977). Timing Control in Speech Production. Bloomington, Ind.: Indiana University Linguistics Club.

Fromkin, Victoria (1979). "Persistent Questions Concerning Distinctive Features." in Lindblom & Öman (eds.), Frontiers of Speech Communication Research. pp. 323-334. London: Academic Press, Inc.

Gandour, Jackson T. (1978). "The Perception of Tone." in Fromkin (ed.), Tone: A Linguistic Survey. New York: Academic Press.

Gandour, Jackson T, Berndt Weinberg, & Diane Rutkowsky (1980). "Influence of Postvocalic Consonants on Vowel Duration in Esophageal Speech." Language and Speech 23:149-158.

Hallberg, Staffan (1978). "Unnatural Phonology." Journal of Linguistics 14:157-177.

Hallberg, Staffan (1980). "Apparent Naturalness in Faroese Phonology." Nordic Journal of Linguistics 3:1-24.

Halle, Morris (1979) "Formal and Functional Explanations in Phonology." in Brogyanyi (ed.), Festschrift for Oswald Szemerényi (Current Issues in Linguistic Theory, vol. 11) Part I, pp. 325-342. Amsterdam: Benjamins.

Halle, Morris, & Kenneth Stevens (1979). "Some Reflections on the Theoretical Basis of Phonetics." in Lindblom & Öman (eds.), Frontiers of Speech Communication Research. London: Academic Press.

Hombert, Jean-Marie (1978). "Consonant Types, Vowel Quality, & Tone." in V. Fromkin (ed.), Tone: A Linguistic Survey. New York: Academic Press.

Hyman, Larry (1976). "Phonologization." in Juilland (ed.), Linguistic Studies Offered to Joseph Greenberg. (vol. 2) pp. 407-418. Saratoga, Cal.: Anma Libri & Co.

Ingria, Robert (1980). "Compensatory Lengthening as a Metrical Phenomenon." Linguistic Inquiry 11:465-495.

Jackson, Kenneth (1967). A Historical Phonology of Breton. Dublin: Dublin Institute for Advanced Studies.

Kiparsky, Paul (1973) "Phonological Representations." in O. Fujimura (ed.), Three Dimensions of Linguistic Theory. Tokyo: the TEC corporation.

Klima, Edward, & Ursula Bellugi (1979). The Signs of Language. Cambridge: Harvard University Press.

Kozhevnikov, V. A., & L. A. Chistovich (1965). Speech: Articulation and Perception. Translated by J.P.R.S., Washington, D.C., No. JPRS30,543. Moscow-Leningrad.

Ladeoged, Peter (1968). A Phonetic Study of West African Languages (revised edition). Cambridge: Cambridge University Press.

Ladefoged, Peter (1980). "What are Linguistic Sounds Made of?" Language 56:XXX-XXX.

Ladefoged, Peter, & Tony Traill (1980). "The Phonetic Inadequacy of Phonological Specifications of Clicks." UCLA Working Papers in Phonetics 49:1-27.

Lehiste, Ilse (1970). Suprasegmentals. Cambridge: MIT Press.

Lehiste, Ilse (1971). "Temporal Organization of Spoken Language." in Hammerich, Jakobson & Zwirner (eds.), Form & Substance, pp. 159-170. [Copenhagen:] Akademisk Forlag.

Lehiste, Ilse (1977). "Isochrony Reconsidered." Journal of Phonetics 5:253-263.

Lieberman, Philip (1970). "Towards a Unified Phonetic Theory." Linguistic Inquiry 1:307-322.

Lindblom, Bjorn, J. Lubker & T. Gay (1979). "Formant Frequencies of Some Fixed-Mandible Vowels and a Model of Speech Motor Programming by Predictive Simulation." Journal of Phonetics 7:147-162.

Lopez, Barbara (1979). The Sound Pattern of Brazilian Portugese (Cariocan Dialect). unpublished Doctoral dissertation, UCLA.

Lovins, Julie B. (1979). "A Study of "Nasal Reduction" in English Syllable Codas." (unpublished revision of paper from CLS 14, circulated in duplicated form, Bell Laboratories, Murray Hil, N. J.)

McCarthy, John (1979). "On Stress and Syllabification." Lingustic Inquiry 10:443-466.

Monahan, K.P. (1979). "On Syllabicity." MIT Working Papers in Linguistics 1:182-190.

Ohala, John (1979a). "Universals of Labial Velars and deSaussure's Chess Analogy." in Proceedings of the IXth International Congress of Phonetic Sciences, Copenhagen, Denmark. vol. II, pp. 41-47.

Ohala, John (1979b). "The Contribution of Acoustic Phonetics to Phonology." in Lindblom & Öman (eds.), Frontiers of Speech Communication Research. London: Academic Press.

Ohala, John, & J. Lorentz (1977). "The Story of [w]: An Exercise in the Phonetic Explanation for Sound Patterns." Proceedings of the Berkeley Linguistic Society 3:577-599.

Pétursson, Magnus (1974). Les articulations de l'islandais à la lumière de la radiocinématographie. Paris: C. Klincksieck.

Safir, Ken (1979). "Metrical Structure in Capanahua." MIT Working Papers in Linguistics 1:95-114.

Selkirk, Elizabeth (1972). The Phrase Phonology of English and French. unpublished MIT Doctoral dissertation.

Sommerfelt, Alf (1921 [1978]). Le breton parlé à Saint-Pol-de-Léon [new edition with notes by F. Falc'hun & Magne Oftedal]. Oslo: Universitetsforlaget.

Stankiewicz, Edward (1972). A Baudouin de Courtenay Anthology. Bloomington: Indiana University Press.

Wipf, Elisa (1910). Die Mundart von Visperterminen im Wallis (Beiträge zur scweizerdeutsche Grammatik, vol. II). Frauenfeld: Huber.

Zee, Eric (1979). "Effects of Vowel Quality on Perception of Nasals in Noise." UCLA Working Papers in Phonetics 45:130-149.

Cross linguistic studies of speech production*

Peter Ladefoged
Phonetics Laboratory, Linguistic Department, UCLA

[Paper presented at a conference on Mechanisms of Speech Production]

We are all tempted to think that the normal is what we do. Cross-linguistic studies of speech are a healthy corrective to this attitude in that they demonstrate that what is commonplace for us is exotic for others. We may think that there is nothing very unusual about American English, but in fact the vowels in words such as beard, bard, and bird are most peculiar, and occur in few, if any, of the other languages in the world. Conversely, words with clicks in them sound curious to us, but they are nothing special to speakers of Bushman languages, in which the majority of words begin with clicks. Studying the speech production mechanisms that occur in other languages enables us to realize the full, extraordinary, capabilities of the human vocal apparatus.

We are, in fact, at a time in history which is probably the most advantageous for cross-linguistic studies. The range of languages spoken is almost as great as it ever has been. About ten thousand years ago, when the world's population was very much smaller, there may have been no more than a dozen languages spoken in Europe; now there are over one hundred. Similar multiplication has no doubt occurred all over the world, as groups of people have become more widespread and diversified. But now, with wider communications making it possible for one group to extend its influence and dominate larger regions, speakers of the smaller, less politically powerful, languages are disappearing. The process of language loss has been accelerating in recent times. We are now soon after a peak in the number of languages in the world, and thus have an unrivalled opportunity to observe the greatest possible range of human linguistic capabilities. In a few hundred years, when the accidents of history may have resulted in Russian, Chinese, and English being the only living languages, linguists will not be so fortunate.

When linguists describe the languages of the world, they often limit themselves to specifying just the linguistic contrasts -- the sounds that can be used to change the meanings of words. The phonemic contrasts that occur in most languages can be specified by reference to a fairly limited set of categories. The contrasts among the consonants of English (or Swedish, or German, or Japanese, for that matter) can be adequately specified by saying whether they are voiced or voiceless, and where and how they are articulated. Thus each consonantal segment in these languages can be characterized by taking one term out of each of the three columns in the lefthand side of Table 1.

---

*This paper is typed using our new text processing system, which, at the moment, does not incorporate phonetic symbols. As a result some unorthodox symbols are used. It should not be inferred that we are departing from our normal devotion to IPA symbols. These are, after all, only working papers.

Table 1. Three term descriptions for the specification of English speech sounds

consonants			vowels		
voiceless	bilabial	stop	high	front	rounded
voiced	labiodental	nasal	mid-high	central	unrounded
	dental	fricative	mid-low	back	
	alveolar	approximant	low		
	palatoalveolar	trill			
	retroflex	flap			
	palatal	lateral			
	velar				

The vowels of most languages can also be described by means of two or three terms, in these cases specifying what are traditionally called the degrees of height and backness of the tongue, and the position of the lips, as shown on the righthand side of Table 1. Linguists differ slightly in the labels that they use, but the notion of a three term specification is generally valid for all descriptions of contrasts, except for those using binary features; and even these can usually be reduced to three term descriptions by taking into account the constraints on possible combinations of values of features that can occur. This paper examines the extent to which the contrasts in some less well known languages can be categorized by means of sets of terms such as those in Table 1. It will also discuss whether these ways of classifying contrasts provide a valid picture of the capabilities of the speech production mechanism.

Speech sounds that do not occur in English are of two types: those that use different combinations of the terms in Table 1; and those that necessitate the use of additional descriptive terms. The latter may be further split into those that require simply an additional possibility in one of the three columns, and those that require an additional column with terms that can be combined with one from each of the existing three columns.

There are numerous examples of sounds in other languages that can be described as combinations of the consonant terms in Table 1 that do not occur in English. Thus Burmese has not only voiced nasals and laterals as in English, but also the corresponding voiceless sounds (symbolized by a subscript [̥]) as illustrated in (1).

(1)	mà	(healthy)	nà	(pain)	la	(moon)	ŋâ	(fish)
	m̥à	(order)	n̥à	(nostril)	l̥á	(beautiful)	ŋ̊â	(rent)

Melpa, a language spoken in Papua New Guinea, has, in addition to the alveolar lateral familiar to us, both dental laterals (symbolized by adding [̪] to form [l̪]) and velar laterals (symbolized by [ʟ]), as in (2). These sounds may be allophonically voiced or voiceless. Illustrative spectrograms are given in Ladefoged, Cochran, and Disner (1977).

(2)	kial̪t̪im	(fingernail)	lola	(speak improperly)	paʟa	(fence)
	wal̪̥	(knitted bag)	bal̪̥	(apron)	raʟ̥	(two)

Table 2. Symbols for combinations of terms in Table 1. Alternate rows: (a) voiceless (b) voiced; Pairs of columns: 8 places of articulation as in Table 1; Pairs of rows: 7 manners of articulation as in Table 1. Asterisks denote combinatorial possibilities that do not and/or cannot occur, as discussed in the text. Parenthesized symbols denote possibilities that have not been observed as contrasting with at least one of the unparenthesized adjacent sounds in the same row.

	1	2	3	4	5	6	7	8
1(a)	p	*	t̪	t	t̠	ṭ	c	k
(b)	b	*	d̪	d	(d̠)	ḍ	ɟ	g
2(a)	m̥	*	(n̪̥)	n̥	(n̠̥)	(ṇ̥)	(ɲ̥)	ŋ̊
(b)	m	*	n̪	n	n̠	ṇ	ɲ	ŋ
3(a)	ɸ	f	θ	s	ʃ	ṣ	ç	x
(b)	β	v	ð	z	ʒ	ẓ	j^	ɣ
4(a)	ʍ	ʋ̥	(θv)	ɹ̥	(ʃv)	(ɹ̣̥)	j̊	(xv)
(b)	w	ʋ	(ðv)	ɹ	(ʒv)	ɹ̣	j	ɣv
5(a)	B	(ṛ̥)	(r̪̥)	(r̥)	(r̠̥)	*	*	*
(b)	B	(ṛ)	r̪	r	(r̠)	*	*	*
6(a)	*	(Ɐ̥)	(ɾ̪̥)	(ɾ̥)	*	(ɽ̥)	*	*
(b)	*	Ɐ	(ɾ̪)	ɾ		ɽ	*	*
7(a)	*	*	(l̪̥)	l̥	(l̠̥)	(ḷ̥)	(ʎ̥)	(ʟ̥)
(b)	*	*	l̪	l	l̠	ḷ	ʎ	(ʟ)

Kele, another Papua New Guinea language, has voiced bilabial trills (symbolized by [B]), as well as the more familiar alveolar trills, as in (3). Ladefoged, Cochran, and Disner report that the rate of vibration of the lips is typically about 30 Hz, and is not significantly different from that for lingual or uvular trills.

(3)	ᵐBin	(vagina)	ᵐBulim	(face)	ᵐBuwen	(testicle)
	ⁿril	(song)	ⁿruwin	(bone)	ⁿrikei	(leg)

Margi, a Chadic language spoken in Northern Nigeria, has a voiced labiodental flap (symbolized by [ⱱ]) as in (4). Photographs of the lip action in this sound have been given in Ladefoged (1968).

(4) bəⱱú (ideophone descriptive of sudden appearance and flight)

Considering all these (to our ears) unusual sounds, it becomes a challenging exercise to find physiologically possible combinations that do <u>not</u> occur. It is not, however, always easy to say which combinations of categories can be ruled out as physiologically impossible. Is it, for example, possible to make labiodental stops and nasals? I would say that these sounds can be made only by people with exceptional dentition, such that there are no gaps between the teeth. Others, however, call the allophone of /m/ that occurs in 'symphony' a voiced labiodental nasal. (I prefer to think of it as a voiced nasalized labiodental fricative; but in any case discussion of this sound need not detain us further as it has not been found to be contrastive in any language.)

Some of the other combinations are more obviously impossible. Thus, although it is possible to make a trill using both lips as in Kele, or a flap using only one lip as in Margi, a labiodental trill using only the lower lip is beyond my capabilities, and a bilabial flap is not at all easy. Physiological constraints on trills and flaps also apply in the retroflex and palatoalveolar regions. I can curl the tip of my tongue up and back and then let it strike the back of the alveolar ridge as it goes forward, thus forming a retroflex flap; and I think I can make a trill in the retroflex position, but I do not know of one in any language. If we take palatoalveolar to imply raising of the tongue blade in the post-alveolar region, then the fricative r-sound that occurs in Czech may be categorized as a voiced palatoalveolar trill. But from a physiological point of view it is difficult to get the blade of the tongue to form a flap in this area. Tongue tip trills and flaps in the palatal and velar regions also seem to be impossible. But are bilabial laterals (the two lips meeting at the center, but not at the side) viable linguistic possibilities? Similarly, how about labiodental laterals? They can be made (with a bit of a grimace), but they have not been observed in any known language.

Apart from the sounds discussed in the previous paragraph, all the other combinations of consonantal terms in Table 1 are symbolized in Table 2.[1] It is instructive to try to find pairs of items that do not contrast. As it is usually fairly easy to find examples of contrasts between sounds with different manners of articulation, or with very different places of articulation, we will limit this discussion to an assessment of the contrasts between sounds that are very similar to one another. In Table 2, parentheses enclose all those items that do not contrasts with one of the adjacent sounds in the same row. In some cases an arbitrary choice has to be made as to which of two symbols should be parenthesized. Thus, considering the lateral symbols in the bottom row, either the palatal lateral [ʎ] or the velar lateral [ʟ̥]could have been put in parentheses. There are languages such as Melpa, illustrated in (2) above, which have velar laterals [ʟ̥] contrasting with alveolar laterals [l] and dental laterals [l̪]. There are also Arandic languages spoken in Australia, such as Kaititj, that have palatal laterals contrasting with dental, alveolar, and retroflex laterals. Examples of contrasts in final position are given in (5).

(5) ilbal̪ (smoke) irmal (fire saw) aldimaḷ kuraʎ (star)

But, as far as I know, no language contrasts velar laterals and palatal laterals, and accordingly either [ʎ] or [ʟ̥] has to be parenthesized. There are also problems in determining which items can be said to contrast in that there are no well defined boundaries between each of the eight places of articulation that have been listed. Thus Yanuwa, another Australian language, contrasts seven out of the eight places of articulation, missing only a contrast between bilabial and labiodental sounds. Examples of oppositions between nasals in intervocalic position are given in (6). The dental nasals are symbolized by [n̪].

(6)	umuwaḍala (in the canoe)	wun̪un̪u (cooked)	wunala (kangaroo)	waṇura (white crane)	n̠an̠alu (tea)	luwaɲu (strips of fat)	waŋulu (big boy)

It must also be remembered that the categories that we have been using do not distinguish between the different parts of the tongue that may be involved in an articulation. A more detailed analysis as in Kirton and Charlie (1978) would show that the differences include tip of the tongue (apical) and blade of the tongue (laminal) distinctions, as well as what might more properly be called a pre-velar rather than a palatal articulation.

In the light of this more detailed analysis it might be thought somewhat procrustean to use the data in (6) to demonstrate that the items in Table 2 row (a) can be said to contrast. But it is inevitable, if we are simply specifying contrasts, that we either have to postulate a very large number of features, or we have to admit that the terms for different features have to be interpreted in different ways in different languages. Linguists invariably try to describe contrasts in terms of a small number of features. If we do this, we cannot consider these features as designating simple properties. Despite the earlier claims of Chomsky and Halle (1968), it now seems completely clear that such a small set of features cannot be regarded as designating phonetic properties definable in terms of physical scales. Instead, each value of a feature (and it does not really matter whether we are talking about traditional features such as Place of articulation, or Chomsky-Halle innovations such as Coronal) designates a cluster of related properties. In some languages some of these properties are necessary (eg the fact that certain sounds are laminal as opposed to being apical in Yanuwa); in other languages other properties are more important (eg in making Malayalam contrasts what matters is the precise part of the upper surface of the vocal tract contacted by the tip or the blade of the tongue). Sounds with the same values of the features have a family resemblance to one another in the Wittgensteinian sense, rather than being alike in their values on some simple physical scale.[1]

I will not try to illustrate all the possible contrasts between adjacent items in the rows in Table 2. Examples of contrasts involving some of the more unfamiliar sounds have been given above. Examples of contrasts between adjacent items for nearly all the other unparenthesized sounds have been given elsewhere (Ladefoged 1968, 1971).

There are several points that are worthy of note concerning the distribution of the parenthesized items in Table 2. Firstly the fricative symbols form the only sets for which a full range of contrasts has been observed. In some ways this is surprising in that fricatives are fairly complex sounds to make. They necessitate more precise positioning of the tongue than, say, stops, in which the lower articulator can be raised against the upper articulator with varying degrees of force as long as a closure is formed. In the case of fricatives the degree of narrowing of the vocal tract has to be very precisely controlled. Secondly, and conversely, there are considerably fewer contrasts among approximants. Indeed, many of these sounds have no special symbol, and I have had to use the corresponding fricative symbol followed by a lowering mark [˅] indicating that the articulation involves a lesser degree of stricture. Thirdly, 21 of the 28 parenthesized items that do not contrast with adjacent sounds are voiceless. All these (and several similar) facts are hard to explain simply from the point of view of avoiding difficulties in speech production. Fricatives are certainly no easier to produce than approximants; and there is certainly not much more physiological difficulty in making any of these sounds voiceless rather than voiced (although it may be significant that more air is used up when making voiceless sounds, so that they require more respiratory effort).

Full explanations of the distribution of the parenthesized and non-parenthesized items lie outside the realm of speech production. Voiceless sounds are less often contrastive perhaps in part because they are in some way less distinct. Conversely, there may be more oppositions among fricatives than there are among stops or nasals because fricatives may be more distinct. Differences among stops or nasals are often marked by only very small differences in the formant transitions. At this point in time, when languages are so numerous and diverse, the observed linguistic phenomena probably reflect a natural balance between the demands of speech production, perception, learnability, memory, and other cognitive factors, all these conflicting demands being tempered by the stresses put on the system by foreign borrowings and other influences resulting from contacts between languages.

We must now consider sounds that cannot be described in terms of the categories in Table 1. Even if we limit ourselves to simply adding terms to each column, the number of possibilities becomes too great to try to represent the observed contrasts as in Table 2. Instead of alternate rows indicating just voiced and voiceless sounds, we would have to take into account at least two other states of the glottis: murmur (breathy voice) as observed in Marathi and other Indo-Aryan languages; and laryngealization (creaky voice) as in Hausa and other Chadic languages. I used to think that no language contrasted all four of the possibilities: voiced, voiceless, murmured, and laryngealized. But now I have observed all these states of the glottis and an additional fifth state involving a tightening of the upper

---

1. This view of features as designating Wittgensteinian family resemblance types is considered further by Lindau (1980), and is also mentioned by Anderson (1980).

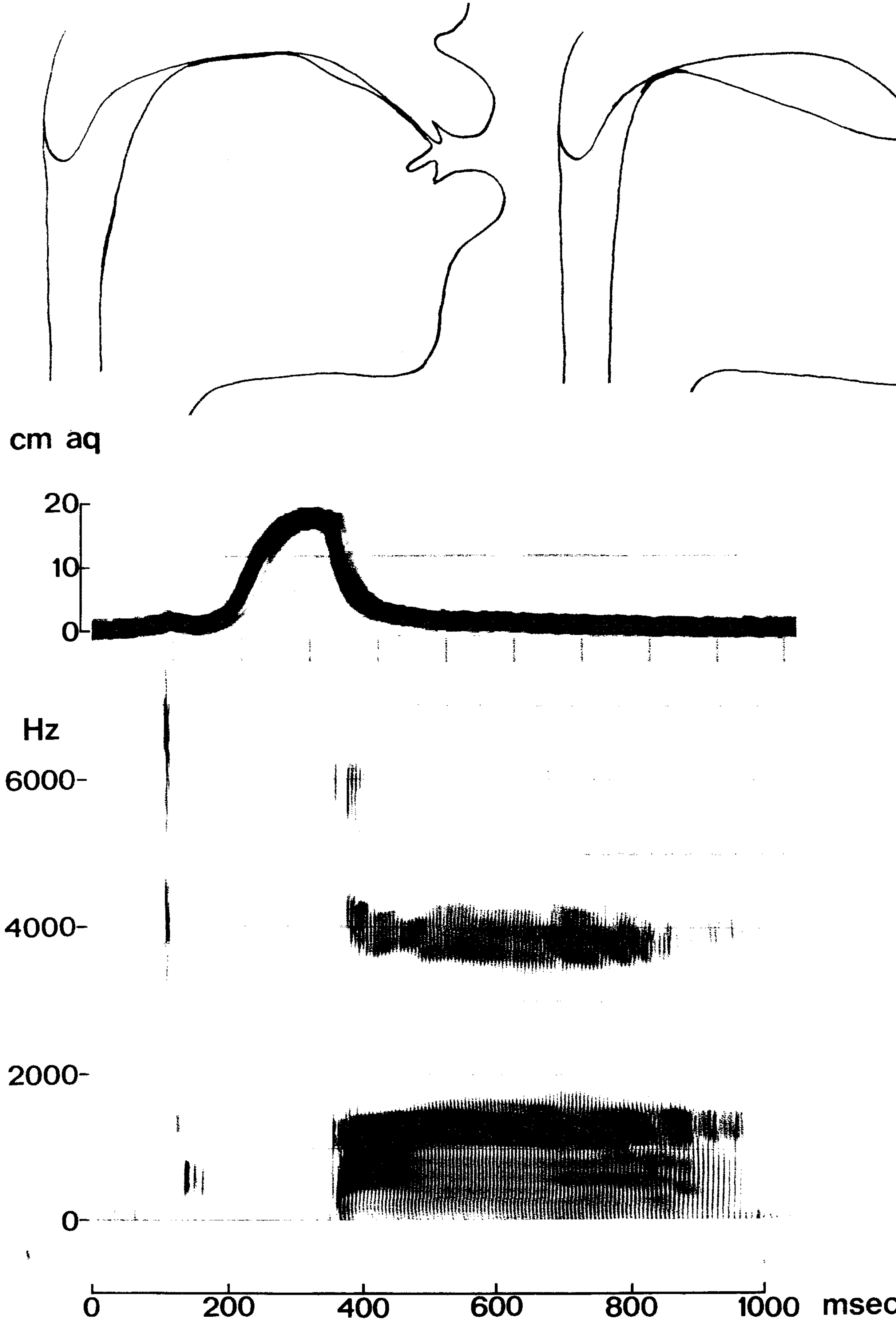


Figure 1. The pronunciation of a voiced uvular ejective dental click in the !Xóõ word g/q'àã (chase). The upper part of the figure shows the estimated positions of the vocal organs at approximately times 0 and 100 msec. The curve in the middle is the pressure of the air in the pharynx. The lower part of the figure is a wideband spectogram.

part of the larynx to produce what may be called ventricular voice in !Xóõ, a Bushman language fully described by Traill (forthcoming). Ventricular phonation is probably more associated with vowels than with consonants, and therefore need not be considered as an extra category in the terms for consonants in Table 1. But as we will not be considering vowels further in this paper, examples of different phonation types in !Xóõ are listed in (7), in order to demonstrate another aspect of the range of the speech production mechanism. The examples also illustrate a phonation type involving a combination of murmur followed by laryngealization in the form of a brief glottal catch.

(7)	Voice	ǁáa	camelthorn tree
	Laryngealized	gǀà’je	bend
	Murmur	!a̤a̤	slope
	Ventricular	!ào	base
	Murmured laryngealized	ǀã̤’je	wait for him

There is some evidence from Traill`s work that speakers of particular languages may overdevelop certain muscles so that they take on an appearance that is not normal in speakers of other languages. The !Xóõ sounds with ventricular phonation are produced by constricting the upper part of the larynx so that the posterior portion of the aryepiglottic muscle nearly contacts the root of the epiglottis. Traill has been able to examine two speakers of this language using both cineradiology and fiberoptic techniques. Both these speakers have abnormally (from an ethnocentric point of view) enlarged muscular pads just above the larynx. Traill, who is himself a fairly fluent speaker of this language, had also been examined in the same way, prior to the recording of the !Xóõ speakers. The xray photographs revealed such curious enlargements of the posterior part of his larynx that it was feared that he might have a tumor, a worry that was finally laid to rest only when the xrays of the !Xóõ speakers showed similar results.

Additional phonation types are far from the only only extra terms needed in Table 1. There must be further terms in the list of places of articulation. The contrastive use of glottal stops (as in Javanese) requires recognition of the glottis as a place of articulation. We must also add terms to account for uvular stops and nasals (as in Eskimo), uvular fricatives (as in French), and pharyngeal sounds (as in Arabic). It has recently been shown (Laufer and Condax, 1980) that pharyngeal fricatives involve the active use of the epiglottis as an articulator, thus demonstrating one more degree of complexity in the use of the speech production mechanism in the formation of linguistic contrasts. Other sounds made at unusual places of articulation include apico-labials in languages such as Big Nambas. According to Fox (1979) this language has stops, nasals, and fricatives in which "the apex of the tongue comes into contact with the upper lip." Additional manners of articulation must also be provided (or a fourth column added) to account for central vs lateral distinctions among fricatives (as in Zulu) and taps (as in Chaga). All these contrasts are illustrated in Ladefoged (1971).

Extra columns are needed to describe sounds in which the primary source of energy is not just lung air moving outwards. I will not attempt to discuss all the ways in whch acoustic energy can be generated without the use of the respiratory system forcing air out of the lungs. Even linguists who are concerned only with classifying contrasts have to complicate the set of categories quite considerably at this point. In this paper I will simply exemplify the range of the speech production mechanism by giving a more detailed account of what goes on in the formation of a sound that may be classified as a voiced uvular ejective dental click. This sound has been chosen for illustrative purposes as it is perhaps the most complicated of the 80 contrasting clicks that occur in !Xóõ (Traill 1978).

Figure 1 is a composite of data related to the !Xóõ word [g|q̓àã] (chase). The upper part of the figure shows the estimated positions of the vocal organs at two different moments in time. These diagrams are based on cine-radiology data reported in Traill (forthcoming). The original tracings are actually of a different word, and do not show the position of the larynx. Cineradiology data illustrating the articulatory movements that occur in a voiced uvular ejective dental click are not available to me. There may, therefore be some inaccuracies in the positions I have reconstructed; but the general picture is fully validated by Traill's cineradiology data on a large number of similar sounds. The lower part of the figure shows actual data on the pressure of the air in the pharynx, and on the acoustic components of this sound, recorded by means of techniques described in Ladefoged and Traill (1980).

This sound begins with the raising of the back of the tongue to the soft palate, forming a velar closure, and the closing of the naso-pharyngeal port so that the pressure of the air in the pharynx begins to rise. As the spectrogram shows, the vocal cords continue to vibrate during this period. At about the same time as the velar closure is formed, the tip, blade, and sides of the tongue contact the hard palate, so that a very small volume of air is enclosed in the mouth, as shown in the lefthand diagram of the vocal organs. The tongue then moves down and back to the position in the righthand diagram, with the air in the mouth being rarefied within a considerably larger cavity. The spectrogram shows the small burst of noise that occurs when the tip of the tongue is lowered and air rushes into the mouth cavity. Shortly after this the glottis is constricted so that there are three slow irregular vibrations of the kind that occur in creaky voice. As the pressure in the pharynx goes down slightly at this moment we may infer that the pharyngeal cavity is being enlarged, perhaps by a slight downward movement of the constricted glottis. Shortly after the irregular vibrations cease we can deduce that the glottis must be completely closed and moving upwards so that the air in the pharynx behind the uvular closure is compressed. The pharyngeal pressure record shows that the pressure rises to nearly 20 cm aq, which is about double the pressure of the air in the lungs in this speaker's other utterances in the same set of recordings. This pressure increase can only be due to the piston-like action of the closed glottis. There is a small burst of noise which occurs when the uvular closure is released, which is followed fairly rapidly by the release of the glottal stop and the start of regular vocal cord vibrations for the vowel.

The complex timing of the events described above cannot be inferred from the classificatory label voiced uvular ejective dental click (or the equivalent set of distinctive feature values), which the linguist ascribes to the sound. Nor, for that matter, are the actions of the speech production mechanism adequately specified by the much simpler set of terms we have been discussing earlier in this paper. Even the different combinations of the terms in Table 1 can be considered only as labels, specifying linguistic contrasts, and grouping them into classes that are appropriate for phonological rules. They are not descriptions that enable speech scientists to know, without further information, the required movements of the vocal organs. Thus describing Burmese as having voiceless alveolar nasals as exemplified in (1) does not make it clear that in a typical production of a word such as [n̥à] (nostril) the tip of the tongue is raised to contact the alveolar ridge, and air escapes through the nose for the first part of this sound; then, about 20 msec before the tip of the tongue comes down, the glottis narrows and voicing begins. Finally, about 40 msec after the vocal cords have begun to vibrate, and 20 msec after the lowering of the tip of the tongue, the soft palate is raised to form a velic closure, so that air can no longer go out through the nose.

As a further example of the inadequacy of linguistic specifications of the speech production mechanism consider the description voiced velar lateral as applied to the sounds in Melpa exemplified in (2) above. Given only this specification, there is no way in which one could set up a computer model of the tongue along the lines suggested by Fujimura (1980) so as to form this sound. Both the term velar and the term lateral have to be given special definitions that apply only when they occur together if we want to use them to provide a precise specification of what part of the tongue is narrowed, and where (and for how long) there is contact with the roof of the mouth.

Furthermore, the inadequacies of linguistic specifications are not just a matter of filling in detail in a mechanical way. There cannot be an algorithm that, given the co-occurrence of three terms such as voiceless dental fricative, enables a full description of the speech production mechanism to be given. These three terms specify different sounds when, for example, they refer to different dialects of English. A typical speaker of Californian English makes the voiceless dental fricative in words such as "thin" with the tip of the tongue protruding below the upper front teeth (Ladefoged 1980). But most speakers of British English make this sound with the tongue behind the upper teeth (Jones 1956). Linguistic descriptions must be accompanied by a detailed, language specific, set of algorithms before they can be interpreted in terms of actual sounds.

As we noted at the beginning of this paper, cross-linguistic studies are useful in drawing our attention to the wide range of phenomena that occur in the languages of the world. But we must always remember that linguists are usually more concerned with classifying contrasts than with describing different actions of the vocal organs. Unless cross-linguistic studies are supported by detailed phonetic investigations, they do not give a valid indication of the capabilities of the speech production mechanism.

References

Anderson, S. (1980) "Why phonology isn't 'natural'" UCLA Working Papers in Phonetics 51, 36-93.

Chomsky, N. and Halle, M. (1968) The sound pattern of English New York: Harper and Row.

Fox, G.J. (1979) Big Nambas Grammar (Pacific Linguistics B-60.) Canberra: Australian National University.

Fujimura, O. (1980) [Paper in this same conference proceedings]

Kirton, J. and Charlie, B. (1978) "Seven articulatory positions in Yanuwa consonants." Pacific Linguistics A-51, 179-197.

Jones, D. (1956) An outline of English phonetics Cambridge: Heffer.

Ladefoged, P. (1968) A phonetic study of West African languages Cambridge: Cambridge University Press.

Ladefoged, P. (1971) Preliminaties to linguistic phonetics Chicago: University of Chicago Press.

Ladefoged, P. Cochran, A. and Disner. S. (1977) "Laterals and trills" J. International Phonetic Association 7.2, 46-54.

Laufer, A. and Condax, I. (1980) "The function of the epiglottis in speech" Language and Speech (in press).

Lindau, M. (1980) "Phonetic differences in Nigerian languages" UCLA Working Papaers in Phonetics 51, 105-13.

Traill, A. "Another click accompaniment in !Xhóõ" Khoisan Linguistic Studies 5, 22-29.

Traill, A. (forthcoming) Phonetic and phonological studies of !Xhóõ Bushman.

Phonetic Differences in Nigerian Languages

Mona Lindau

It is generally presumed that features on the phonological and phonetic levels are identical. In the standard theory of generative phonology contrast on the phonological level is described in terms of binary values of a feature. Phonetic differences between languages are accounted for at the phonetic level as differences on a multivalued scale of the same feature. In addition, each feature is related to a single articulatory or acoustic correlate. However, Ladefoged (1979) demonstrates that relating a feature to a single physical scale often constitutes an oversimplified view of feature correlates. Instead, a phonological feature is more like a cover term for several articulatory and acoustic parameters. A particular sound may differ between languages along any one of these various parameters.

I will discuss some phonetic differences between several Nigerian (and some other African) languages and examine one particular kind of relationship that may occur between features and their phonetic correlates. The sound classes of rhotics, implosives, and vowels will be considered. In these studies care has been taken to represent each language by as many speakers as was possible for the recording situations. The results should thus reflect differences between languages rather than between individuals. The recordings were made in a sound studio, where available, otherwise in the field with a good taperecorder. The utterances were words illustrating the relevant sounds, said in a frame. There was an attempt to find words with the rhotics and implosives in the same vocalic environment between open vowels, but this was not always possible.

The rhotics form a sound class with a great deal of phonetic variation. Phonologically, rhotics tend to behave in similar ways. They occupy the same place in consonant systems, usually in minimal opposition to a lateral. They also tend to occupy the same place in syllable structures.

		Language	Number of speakers	Number of tokens per speaker
Kwa	Ịjọ́	Kalaḅári	7	3
		Ḅụ́mọ́	4	3
		Ịzọ̣n	6	3
	Ẹ́do	Degema	4	3
		Ẹ́do (Bini)	4	3
		Ǵhotuọ	6	3
Chadic		Hausa	12	3

Table I

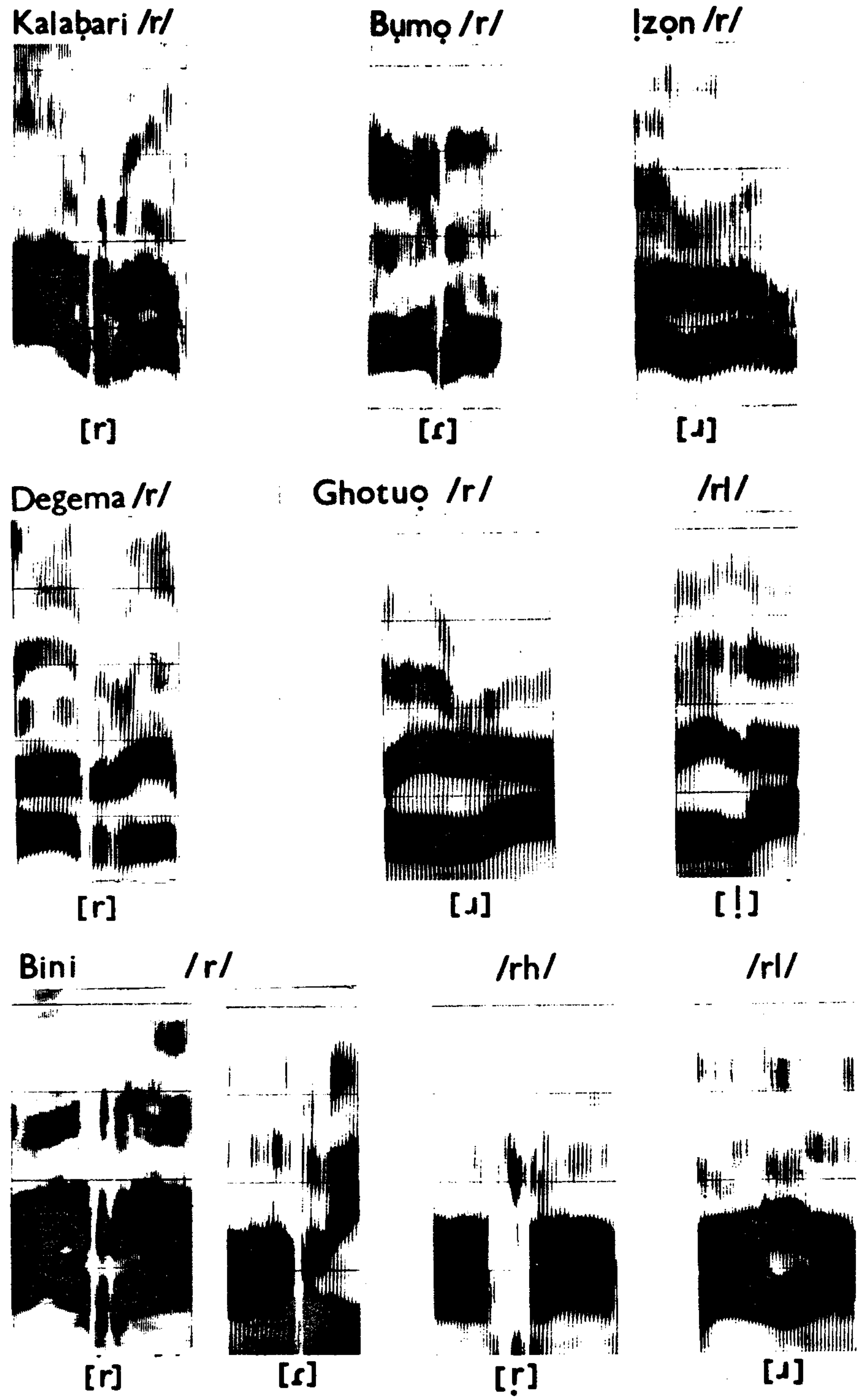


Figure 1. Spectrograms of r-sounds in some Nigerian languages.
Kaḷaḅari /ar(i) amaa/ 'her hometown'
Bụmọ /ara/ 'she/
Ịzọn /araụ/ 'she'
Ḋegema /ọḍare me/ 'let him accompany me'
Ghotuọ /rao/ 'beg', /rla/ 'rot'
Ẹdo (Bini) /raa/ 'will come', /lare/ 'come/, /rhaa/ 'will steal/, /rlaa/ 'will eat'

The r-sounds in the language samples were analyzed from wide band sound spectrograms. The languages, number of speakers and tokens are tabulated in Table 1. Figure 1 illustrates r-sounds in some Nigerian languages. On the spectrograms trills show up as a pulse train of alternating closures and openings, as in Kalaḅari, Degema and Bini. Taps appear as a single pulse in a trill, as in the Bụmọ and Bini examples. Approximant r′s have clear formants with somewhat less intensity than surrounding vowels, as in Ịzọn and Ghotuọ.

The formant frequencies of the r-sound are related to places of constriction in the vocal tract (Fant 1968). For r-sounds the first two formants seem to mostly reflect the quality of surrounding vowels; the third and fourth formants are more important indicators of the place of constriction in this type of sound. According to acoustic theory, a lowered third formant, close to the second formant, indicates a constriction fairly far back in the postalveolar-midpalatal region with strong retroflexion. This happens in Ịzọn. As the constriction moves forward in the mouth the third formant increases. A relatively high third formant, close to the fourth formant indicates a dental place, as happens in the Kalaḅari example.

Looking at rhotics in the individual languages it is found that in Kalaḅari four of the seven speakers use trilled r′s in free variation with taps. The other speakers use approximant r′s. The third formant is high and close to the fourth formant, indicating a dental/alveolar place of articulation.

None of the four Bụmọ speakers employs a trill. The tap is the most common: three speakers use taps (one of them will occasionally vary his taps with approximants), and the fourth speaker uses approximants throughout. The third formant is generally high (around 2,500 Hz), and not really close to the second formant, showing a retroflex postalveolar place. All six speakers of Ịzọn realize the /r/ phoneme as an approximant, alternating with a tap in only a few cases. The third formant is low and close to the second formant. showing a retroflex postalveolar place of constriction.

The three Ịjọ languages thus exhibit trills, taps and approximants at different places. The Ẹdo languages display a somewhat more complex situation.

Degema has one /r/ phoneme. This is realized as a trill for three of the four speakers. This trill may have voiceless allophones in initial and final position, and it may have a tap allophone in intervocalic position. The fourth speaker uses an approximant. The third and fourth formants are close for all types of r′s, showing a retroflex dental/alveolar place of constriction.

Ghotuọ and Ẹdo are described in a monumental work by Elugbe (1973). The following phonemic analyses rely on his work. Ghotuọ has three liquid phonemes: /l/, /r/, and /rl/. The regular lateral does not concern us here. Elugbe describes /r/ as a voiced alveolar trill. The six speakers in my recordings all used an approximant [ɹ] here. Only one speaker varied this with a trill in two instances. The third formant is lowered, reflecting a retroflex place. The /rl/ phoneme is described by Elugbe (1973, 1978) as a tapped lateral, which fits all of my speakers. One could add that /rl/ may also be produced as an approximant with lateral release, [ɹˡ], when words are produced in isolation.

Ẹdo contrasts four liquids: /l/ ≠ /rl/ ≠ /rh/ ≠ /r/. Again, the lateral will not be discussed. Elugbe describes /rl/ as a voiced alveolar approximant, [ɹ]. Three of the four speakers here are in accordance with this, and the fourth uses an approximant with lateral release, [ɹˡ]. The /rh/ is a voiceless alveolar trill for all speakers, and /r/ may be a trill or a tap, in all instances with a high F3, indicating a voiced alveolar trill.

The two r´s in Hausa have recently been described by Newman (1979). Figure 2 shows the Hausa trill/tap and the flap. My data are restricted to intervocalic position, and the speakers exhibited considerable variation between taps and approximants in this position. However, even with all this variation as to manner of articulation, the two r´s are still kept apart as to place, manifested in the fourth formant, which is consistently and significantly lower for the flap /ɽ/ than for the trill/tap /r/ according to a grouped t-test ($p<0.001$).

Figure 2. Spectrograms of r-sounds in Hausa.
/káɽáa/ 'cornstalk', /kàràatúu/ 'reading'

In the languages described above trills exhibit more variation than approximants. A person who uses a trill will often have taps in free variation. Occasionally a speaker with an approximant r-sound will have a tap allophone, but basically the approximants do not vary as much as other r-sounds. Approximants may, however, weaken to complete deletion as a sound change. Elugbe (1973) describes one type of weakening in the Edo languages, where proto-Edo *dh > ḷ >ɹ> Ø. Ghotuo has a tapped [ḷ] and Ø, Ẹdo has [ɹ̃]. I do not know of any cases of the opposite direction of sound change within the rhotic group, i.e. where approximants or taps have developed into trills. Trills do develop from sound changes, but to my knowledge only from alveolar non-rhotics. For example in Hausa, alveolar obstruents in final position developed into trilled [r] (Klingenheben's Law, Schuh 1972). Proto-Ẹdo *dh > Ẹdo [r]/ i---i (Elugbe 1973). Thus weakening seems to be the general type of sound change affecting rhotics. Trills are original or develop from weakened non-rhotics, then trills ma become taps, then taps may become approximants, and these may be deleted altogether. If the opposite pattern, strengthening, does exist, I would venture to claim that it is much less common.

It is clear that the class of [+rhotic] sounds do not have a simple phonetic correlate. In figure 1, the Kalaḅari trill just does not have much in common with the Ghotuọ approximant. But spectrograms of all the trills look similar to each other, irrespective of their place (position of F3) And the trills are similar to the taps in their pulse like quality. Taps and approximants exhibit a different relationship, in that they are related with respect to production. The step from one complete closure to no closure (and some lengthening) in an approximant constitutes a change in position along a single articulatory parameter of stricture.

Rhotics may also vary as to place of constriction in the vocal tract. The third formant is lowered and close to the second formant in the Ịzọn and Ghotuọ approximants, and it is somewhat less lowered for all the r's in Ẹdo. The third formant is fairly high in Bụmọ, and close to the fourth formant in Degema and Kalaḅari. The Hausa /r̃/ has the highest fourth formant. Thus r-sounds are also related to each other along a continuum going from a very low third formant to a very high fourth formant. Although the Kalaḅari trill and the Ghotuọ approximant are so different, they may be related through intermediate members of the family of rhotics.

In spite of the lack of a pervading, essential property for all rhotics, there are resemblances throughout the class. The nature of the relatonship between members of the class of rhotics is thus one of family resemblance in Wittgenstein's sense of the term (1958). Family member r1 resembles r2 which resembles r3 which resembles r4, and although r1 and r 4 may not be much alike, it is entirely possible to express their relatedness as a set of steps across other members. In doing so, we also express the fact that some members, like trills and taps, are closer related than other members, say trills and approximants. This helps in explaining why some variations in the class of r-sounds are more common than others.

Implosives constitute another example of a sound class whose members have a family resemblance kind of relationship. Ladefoged (1979) makes the point that the phonetic differences between implosives in Hausa and Kalaḅari cannot be described using the features proposed by Chomsky and Halle (1968). Figure 3 shows oscillomink tracings of the intensity (top line) and the sound wave (bottom line) of word with voiced bilabial plosives and implosives in Bụmọ Ịjọ, Okrika Ịjọ, and Degema. The illustrations are representative of a number of speakers for each language (table II).

Language	Number of speakers	Number of tokens per speaker
Hausa	12	3
Bu̦mo̦ I̦jo̦	4	3
Okrika I̦jo̦	4	3
Degema	4	3

Table II

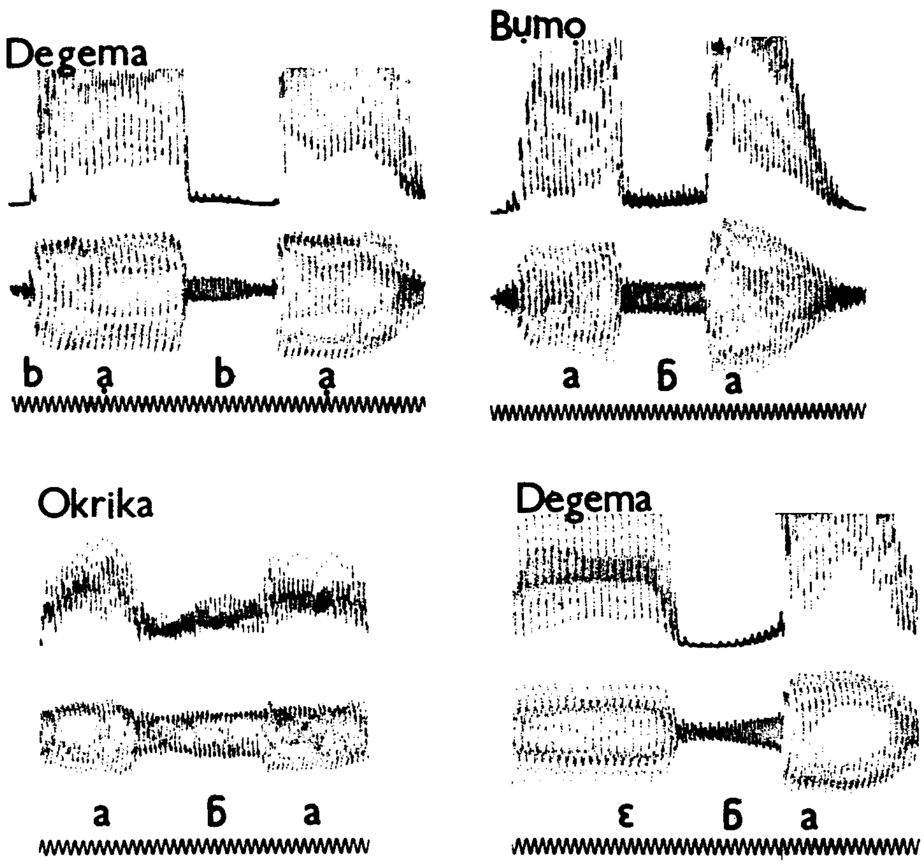


Figure 3. Oscillomink tracings of intensity (top line) and soundwave (bottom line) of words with bilabial plosive and implosives.

Degema /b̦a̦ba/ 'carry (child on back)', /e̦b̦a/ 'mad'
Bu̦mo̦ I̦jo̦ /aɓa/ 'to kill her'
Okrika /aɓa/ 'to kill her'

The implosives in the I̦jo̦ languages and in Degema exhibit regular voicing throughout the closure and a relatively high intensity level. In comparison, consider the "regular" /b/, as in the Degema egressive voiced [b]. With voicing going on throughout the closure, and with no adjustment of oral tract size, the oral pressure will increase during the voiced closure up to a point where the supraglottal and subglottal pressures are equalized, and voicing fizzles out towards the end of the closure. In a regular [b] the intensity is relatively low during the closure. It is possible to prolong and intensify voicing by lowering the larynx and thereby increasing the size and lowering the pressure in the oral tract. This is apparently what happens in the Bụmọ, Okrika, and Degema implosives. In Bụmọ the larynx lowers enough to keep up steady, fairly intense voicing throughout the closure period. In Okrika and Degema the voicing intensity even increases during the closure, indicating larynx lowering to a greater degree than in Bụmọ. The implosives in Degema also sound very intense.

The Hausa implosives show more variation between speakers than one finds in the previous languages. Figure 4 shows the Hausa /ɓ/ produced by four of the twelve speakers. Speaker 1, and one other speaker, use a fully voiced implosive of the same kind as the Ịjọ and Degema speakers. The most common production (eight speakers) of implosives in Hausa is shown by speakers 2 and 3 in the figure. These speakers combine regular and irregular voicing. The intensity tends to increase during the closure, which indicates that this type is still implosive. Two speakers, represented by speaker 4 in the figure, had an irregularly voiced onset of the closure, followed by no voicing at all during the rest of the closure.

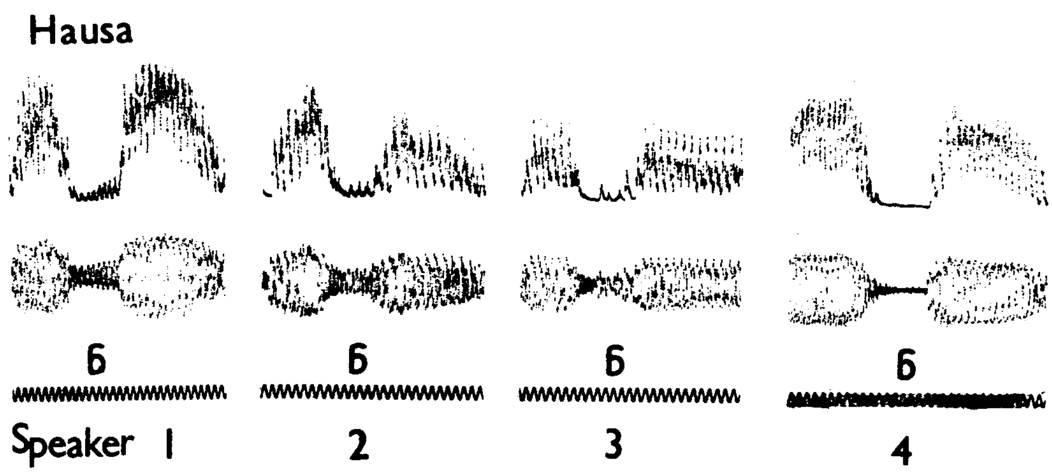


Figure 4. Oscillomink tracings of intensity (top line) and soundwave (bottom line of Hausa /gáɓàa/ 'joint' of four speakers.

The implosives in Bụmọ and Degema are related on a larynx lowering parameter. The larynx lowers enough in Bụmọ to keep up steady intensity voicing. In Okrika and Degema it lowers to a greater extent, so that the voicing intensity increases. The implosives in these languages are related to the implosives in Hausa on a voicing type parameter. The Hausa implosives may be regularly voiced, as in the Ịjọ languages and Degema, or they may combine regular and irregular voicing, or they maybe voiceless. The Bumo /ɓ/ and the

Hausa /ɓ/ can only be related through intermediate types, so again this is an example of family resemblance type of relationship.

A final example of this type of relationship can be found in the vowel harmony of African languages. Vowel harmony of structurally similar kinds is found in both Niger-Congo and Nilo-Saharan languages (Hall et al. 1974). In many of the Kwa languages of Niger-Congo the vowel harmony is controlled by variation in the size of the pharynx. The one set of vowels is produced with an expanded pharynx (by advancing the tongue root and lowering the larynx), and the other set of vowels is produced with a constricted pharynx (by retracting the tongue root and raising the larynx) (Lindau 1975, 1979). In Ateso (Eastern Nilotic/Nilo-Saharan) the tongue root does vary but not to any greater extent than can be accounted for by a concomitant variation in tongue height. Jacobson (1978) found that in Dho Luo (Western Nilotic/Nilo-Saharan), either tongue height or the tongue root with or without the laryngeal displacements, may serve to distinguish the harmonizing sets. Jacobson also points out that in other Western Nilotic languages like Shilluk and Dinka, however, vowel harmony is best related to distinct phonation types.

Vowel harmony in African languages is thus correlated with several articulatory parameters. These parameters are not entirely independent of each other. Tongue root movement is necessarily part of varying tongue height. Laryngeal displacements may be connected to phonation types. When a sound is produced with a strongly constricted pharynx that includes a raised larynx, as happens for example in Arabic pharyngealized sounds, a creaky phonation type is often an accompaniment. The description of vowel harmony in African languages requires one phonological feature so that the structural unity can be captured. This feature serves as a cover term for a family of articulatory correlates, any member of which resembles some other member.

In conclusion, searching for the essence of a particular class of sounds in terms of a single phonetic correlate is sometimes a futile task. Some features are better described as cover terms for a family of sounds, any member of which resembles some other member. The relationship between phonology and phonetics is thus considerably more complicated than what so far has been considered.

References

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

Condax, I.D. and G.S. Nathan. 1979. "It sounds like some kind of 'r'. Why?" Paper given at the 54th meeting of the Linguistic Society of America, Los Angeles, Dec. 1979.

Delattre, P. and D. C. Freeman. 1968. "A dialect study of American r's by x-ray motion picture." Linguistics 44:29-68.

Fant, G. 1968. "Analysis and synthesis of speech processes" In B. Malmberg (ed.) Manual of Phonetics:171-277.Amsterdam: North-Holland Publishing Co.

Hall, B., R.M.R. Hall, M.D. Pam, A.Myers, S.A. Antell, and G.K. Cherono. 1974. "African vowel harmony systems from the vantage point of Kalenjin." Afrika und Übersee, Band LVII, Heft 4:241-267.

Jacobson, L. 1978. "Dho Luo vowel harmony." UCLA Working Papers in Phonetics 43.

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

Ladefoged, P. 1975. A Course in Phonetics. New York: Harcourt, Brace & Jovanovich, Inc.

Ladefoged, P. 1979. "What are linguistic sounds made of?" UCLA Working Papers in Phonetics 45:1-24. Forthcoming in Language.

Lindau, M. 1978. "Vowel features." Language 54.3:517-540.

Lindau, M. 1979. "The feature Expanded" Journal of Phonetics 7:163 -176.

Newman, P. 1980. "The two R's in Hausa" African Language Studies 17.

Schuh, R. 1972. "Rule inversion in Chadic." Studies in African Linguistics 3.3:379-397.

Stanford Phonology Archive. Interim Report.

Wittgenstein, L. 1958. The Blue and Brown Books. Preliminary studies for the Philosophical Investigations. New York: Harper & Row.

My thanks to Peter Ladefoged for many discussions on this paper and to Richard Harshman for referring me to Wittgenstein. I am also indebted to Ben Elugbe, Charles Jenewari, and Kay Williamson for all their help with data gathering.

The story of /r/

Mona Lindau

Features on the phonological and phonetic levels are generally presumed to be identical and related to a single articulatory or acoustic parameter. This makes it possible to describe contrasts in binary terms, as in the standard theory of generative phonology. In addition, phonetic differences between languages can be described in terms of different values along a particular phonetic parameter. This paper will argue that at least some phonological features are associated with the phonetic level in a more complex and considerably different way than is generally assumed. The class of r-sounds provide a good illustration.

The rhotics form a class of phonetically quite different members. Phonologically, rhotics tend to behave in the same way: they occupy the same place in consonant systems and in syllable structures. In consonant clusters the rhotics tend to be vowel adherent. Rhotics participate in the same kind of rules. For example, rhotics often alternate with other rhotics. In Persian, a trilled r has a tap allophone in intervocalic position and a voiceless trill variant in wordfinal position. In Fulfulde, a trill is realized as an approximant [ɹ] before a consonant. Postvocalic r′s tend to become vowels or disappear altogether, as happens in British English, German, Danish and the Swedish spoken in Skåne. Rhotics have similar effects on environments: vowels before r′s tend to lengthen, as in English and Swedish; vowels before or after r′s tend to lower, as in French, Danish, and Swedish.

From a phonetic point of view the rhotics exhibit a wide variety of manners and places of articulation. It is difficult to imagine a single phonetic correlate among all this surface variation, although it has been proposed that there is a common acoustic factor, namely a lowered third formant, that is also close to the second formant (Ladefoged 1975, Lindau 1978, Condax and Nathan 1979).

Language		Number of speakers	Number of tokens per speaker
California English		9	2
Stockholm Swedish		3	3
Skåne Swedish		5	8
Chicano Spanish		6	2
Hausa		12	3
	Degema	4	3
Ẹdo	Ẹdo (Bini)	4	3
	Ghotuọ	6	3
	Kalaḅari	7	3
Ịjọ	Bụmọ	4	3
	Ịzọn	4	3

Table I

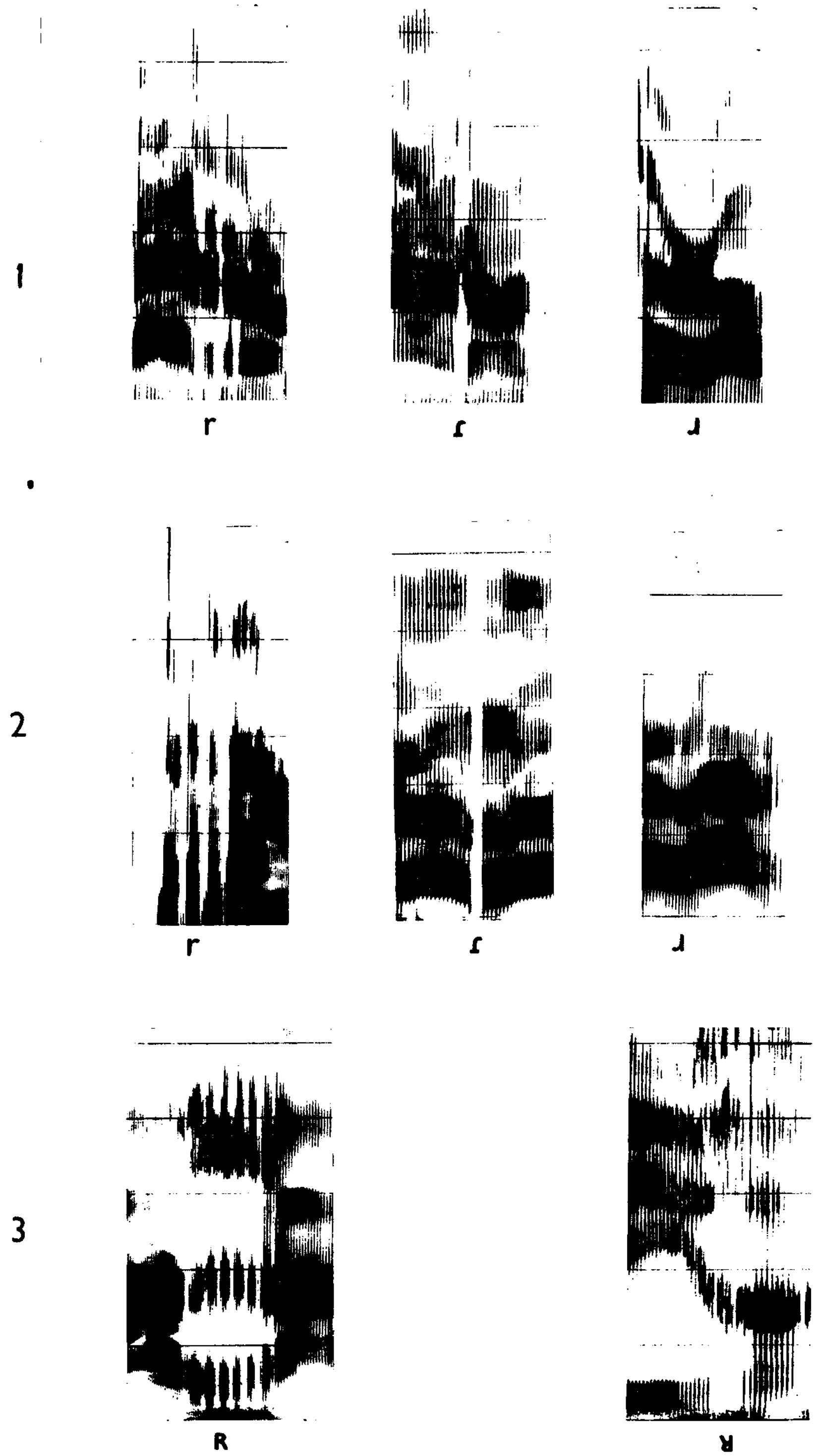


Figure 1. Spectrograms of trills, taps, and approximants. Each spectrogram is individually calibrated in steps of 1000 Hz. The r's come from the following languages and words:

1: Spanish /karo/ 'car'; Spanish /kaɾo/ 'expensive'; English 'thorough'

2: Stockholm Swedish /ramla/ 'fall'; Degema /ararar/ 'things', Degema /ararar/ 'things' (another speaker)

3: Skåne Swedish /bar:a/ 'to shed (of pinetrees)'; Skåne Swedish /fi:ra/ 'celebrate' (another speaker)

I investigated r-sounds in three Indo-European languages and seven languages spoken in West Africa. Each language was represented with several speakers so that the results reflect differences between languages rather than between individuals. Table 1 contains a list of languages and speakers. The data were recorded in a sound studio where available, and otherwise in the field with a good taperecorder. The utterances consisted of words illustrating the r-sounds said in a frame. There was an attempt to find words with the r's in the same vocalic environment, usually between two open vowels, but this was not always possible. The data were analyzed from wide band sound spectrograms.

While it is true that the rhotics in some languages can be characterized in terms of a lowered F3, this is not the case in other languages. In figure 1 the top row illustrates languages with trills, taps and approximants with a lowered F3. At the one extreme of this type of r is English, with a very low F3, coming close to F2. Spanish has two phonemic r's, realized as a trill and a tap, both of which are produced with a lowered F3. Thus, in spite of their different production mechanisms r's in English and Spanish resemble each other acoustically in the lowered F3. However, a lowered F3 is not a characteristic of all rhotics. In the second row the trill from Stockholm Swedish, the tap from a Degema speaker, and the approximant from another Degema speaker do not have a lowered F3. In fact, F3 is somewhat raised and in Degema it is close to the fourth formant.

The fact that a lowered third formant is not a pervading property of all r's should not be a surprise. Formant frequencies reflect constrictions in the vocal tract. From studying the nomograms in Fant (1968) it is clear that the third formant in rhotics indicates the places of constriction fairly well. The American English /r/ has a lowered F3. A lowered third formant may be obtained by retroflexion and a postalveolar-palatal constriction; alternatively, it may be produced with the tongue tip down, some midpalatal constriction, and a constriction in the lower pharynx. Adding lip rounding also lowers the third formant. All of these mechanism are used by speakers in producing the English /r/ (Delattre and Freeman 1961, Lindau 1978, Ladefoged 1979). The articulation of the American English /r/ in <u>herd</u> is shown by the radiographic tracings in Figure 2. All six speakers produce this /r/ with constriction in the pharynx just above the epiglottis, and a bunching of the body of the tongue towards the palate. In addition P3 and P5, and perhaps P4 use some lip rounding as well.

Rhotics in other languages may be produced with constriction in other places in the vocal tract. From Fant's nomograms one may conclude that the close F3 and F4 in the Degema trills and taps indicate a constriction further forward, in the postdental region, than in the r's with a lowered F3.

The third formant will increase even further when the place of constriction is in the uvular area. The uvular r's from Skåne Swedish in the third row of figure 1 display a very high F3, close to F4 for both the uvular trill and the uvular fricative.This was typical for four of the five speakers. The fifth speaker used a somewhat lower F3.

If the phonetic correlate of r's is not the third formant, then what property makes us recognize all these r's as the "same" sound?

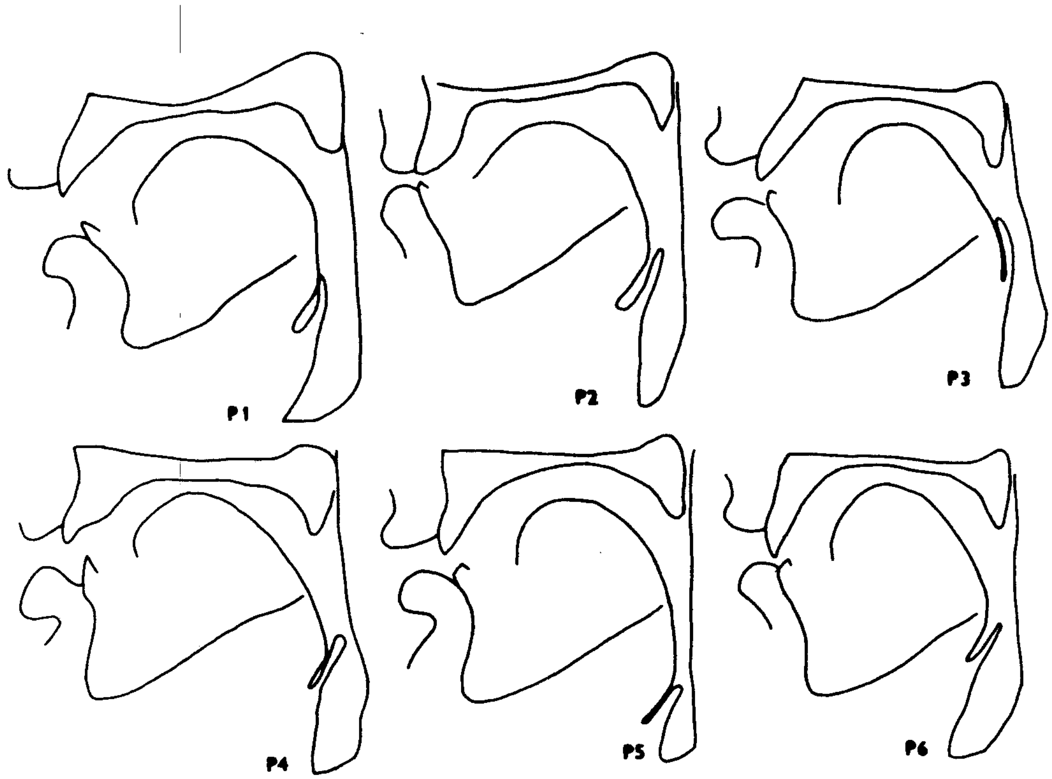


Figure 2. Cineradiographic tracings of the /r/ in herd from six American English speakers (P1-P6).

The first and second formants do not seem to correlate very well with r-ness. More than anything else they seem to reflect coarticulation with surrounding vowels. There is usually a decreased formant intensity associated with the rhotics, but this tends to be true for all sonorants, not just for the rhotics.

The uvular trill used in Skåne and the English approximant /r/ do not seem to have much in common with each other. The uvular trill is however strikingly similar to the other trills in the pattern of pulses of closures and openings, where the pulses are of about the same duration. The trills sound similar. All the uvular rhotics are alike in their formant patterns. The similar pulsing patterns in all the trills explain the change from tongue tip trills to uvular trill in languages like French, German, and Swedish. Once the r-sound is manifested as a uvular, it often weakens and there is free variation between trills, fricatives and approximants. The acoustic similarity among these rhotics is in fact so great that in some areas in Sweden on the border between tongue tip r's and uvular r's members of the same family may use either front or back r's and never notice the difference for other members in that family (Ohlsson et al. 1977).

Figure 1 also shows that trills and taps are alike. The taps appear as a single pulse in a trill. The duration of taps and single pulses are about the same. A tap is a frequent allophone of a trill, particularly in intervocalic position. Trills and taps are often in free variation. Out of 85 languages with trills in the Stanford Phonology Archive, about 45 have tap variants.

Taps and approximants exhibit a different relationship than that between taps and trills. Although the third formants in taps and approximants may be similar, the general acoustic patterns of these two rhotics are considerably less related than those of trills and taps. However, from an articulatory point of view the step from the complete closure of a tap to no closure (and some lengthening) for the approximant could be regarded as a change in position along a single articulatory parameter of stricture. Thus taps and approximants are similar with respect to production.

The approximant r-sound is a less common manifestation of an /r/-phoneme than are trills and taps. In the Stanford Phonology Archive, trills and taps occur about ten times as often as approximants. Approximant allophones of either trills of taps do occur, but only about a third of the time (13 languages) one finds allophonic alternations between trills and taps (45 languages).

Thus there is no physical property that constitutes the "essence" of all rhotics. Instead, each member of the class of r-sounds resembles some other member with respect to some property, but not with respect to the same property across all r-sounds. Trills and taps resemble each other in the duration of the pulses involved, tongue tip trills and uvular trills resemble each other in the patterns of consecutive rapid pulses, uvular trills and uvular fricatives are alike in their formant pattern, tongue tip taps and approximants are close to each other on an articulatory stricture scale and perhaps as to the position of the third formant. The nature of the relationships between members of the class of rhotics can profitably be characterized in terms of Wittgenstein's (1958) notion of family resemblance. Family member r_1 resembles r_2 which resembles r_3 which resembles r_4. Although member r_1 and r_4 may not be much alike, it is entirely possible to express their relationship as a set of steps across other members, as illustrated above. In doing so the fact is also expressed that some members, like trills and taps, are closer related than other members, say alveolar trills and approximants. This helps in explaining why some variations in the class of r-sounds are more common than others.

Clearly, searching for a single phonetic correlate underlying a whole class of sounds may not always be a profitable task in phonology. Some phonological classes, like the nasal, can be properly characterized in that way, but the relationship between phonological classes and phonetic classes may be considerably more complex than the one to one relationship that is generally assumed. The class of rhotics provides one example of a class where the relationship with the phonetic level is best dealt with in terms of family resemblance.

References

Condax, I. D. and G.S. Nathan. 1979. "It sounds like some kind of 'r'. Why?" Paper given at the 54th meeting of the Linguistic Society of America, Los Angeles, Dec. 1979.

Delattre, P. and D.C. Freeman. 1968. "A dialect study of American r's by x-ray motion picture." *Linguistics* 44:29-68.

Fant, G. 1968. "Analysis and synthesis of speech processes." In B. Malmberg (ed.) *Manual of Phonetics*: 171-277. Amsterdam: North-Holland Publishing Co.

Ladefoged, P. 1975. *A Course in Phonetics*. New York: Harcourt, Brace & Jovanovich, Inc.

Ladefoged, p. 1979. "What are linguistic sounds made of?" *UCLA Working Papers in Phonetics* 45:1-24. Forthcoming in *Language*.

Lindau, M. 1978. "Vowel features" *Language* 54.3:517-540.

Ohlsson, S. Ö., J.P. Nielsen, and K. Schaltz. 1977. "Om r-gränsen pa Öland: pa Östfronten intet nytt?" *Arkiv för Nordisk Filologi* 92:177-199.

Stanford Phonology Archive. Interim Report.

Wittgenstein, L. 1958. *The Blue and Brown Books.* Preliminary studies for the Philosophical Investigations. New York: Harper & Row.

Thanks to B.Elugbe, R.Harshman, J.Fordyce, M.Garba, M.Godinez, C.Jenewari, P.Ladefoged, S.Okpara, and K.Williamson for all their help with data collection and discussions. This work was supported by the National Science Foundation.

Palato-alveolar Affricates in Several Languages

Ian Maddieson

Introduction

This paper reports on the first stage of an investigation which is using measurements of certain properties of palato-alveolar affricates in several languages to investigate whether the phonological inventory of a language determines, at least in certain respects, the phonetic realizations of its phonemes. The investigation began with two particular hypotheses in mind, namely, that when a language has more contrasts in a particular section of its phoneme inventory, more extreme values of appropriate phonetic parameters are found to occur in the language, and less variation is permitted in the production of those segments.

Voiceless palato-alveolar affricates were examined in three languages which have differently structured phonological inventories resulting in differing degrees of contrast along a stop-affricate-fricative continuum in this area. Standard Spanish has /tʃ/ but there is no palato-alveolar fricative /ʃ/ in the inventory. The nearest stop is the dental plosive /t̪/. Standard British English has alveolar /t/ and palato-alveolar /ʃ/ in contrast with /tʃ/. Standard Italian has both a single affricate /tʃ/ and a geminate /tʃ:/ and also has a fricative /ʃ/ and a single /t/ and geminate /t:/. Italian thus has the largest number of distinctions to maintain and Spanish the least.

The predictions made by the hypotheses outlined above are therefore a) that Italian /tʃ/ and /tʃ:/ will be produced with more extreme values on some appropriate phonetic parameter(s) than Spanish or English /tʃ/, and, b) that Italian will show the least variability in the production of the segments examined, Spanish the most, and English an intermediate amount.

Method

For each language a list of sentences was prepared containing target words embedded in a carrier phrase. The carrier phrase consisted of two syllables preceding the target word and two syllables following it. The phrases used are given in (1).

(1)	Spanish	Diga ____ a mi.	/diɣa ____ a mi/
	English	Please say _____ for me.	/pliz seɪ ___ fə/fɔ mi/
	Italian	Dica ____ per me.	/dika ____ pɛr me/

The target words all have the structure CVCV, where the initial C is /k/, the vowels are the closest approach to a low central vowel which is phonotactically permitted in the environment in the respective languages, and the medial C is one of the target phonemes. The list of words is given in (2), together with a gloss where relevant.

(2)	Spanish	/tʃ/	cacha	'segment of the peel of a fruit'
	English	/tʃ/	cutcher	(nonsense word)
	Italian	/tʃ/	cacia	(nonsense word)
		/tʃ:/	caccia	'hunting'

In general the words used were either nonsense words, or were words that were unfamiliar; they were however orthographically unambiguous and caused no difficulty for any of the subjects.

Eight subjects were used for each language (Spanish 5F 3M; English 3F 5M; Italian 4F 4M). They read a list of sentences in which the target words occurred 10 times, along with words that had minimal differences and words of quite different phonological structure. Subjects read the list twice, once at a slow speed and once at a fast speed. In order to control the variation in speed both readings were paced by an electronic metronome. During the slow reading the metronome was set to indicate intervals of 3 seconds, and the subjects were instructed to stretch out the sentences to fill the time between the metronome signals. For the fast reading the metronome indicated 1.75 second intervals, and the subjects were instructed to squeeze the sentences into the time between the signals. Before each reading subjects were rehearsed until the correct reading speed had been satisfactorily achieved. Subjects were seated in a sound-attenuated booth while a high quality tape-recording of their readings was made.

The audio recordings were digitally sampled at 10KHz onto the UCLA Phonetics Laboratory PDP-12 computer and durational properties of the affricates were measured. Durations were determined using an automatic segmentation algorithm implemented by James Wright for the initial location of boundaries. This program allocates each 5 msec of speech to one of the categories 'voice', 'noise' and 'silence'. The boundaries obtained in this way can then be fine-tuned interactively by inspection of the waveform in the boundary area and adjustment of a cursor.

Results

The following measures were obtained: a) the duration of the silent interval (closure phase) of the affricates, b) the duration of the fricative portion of the affricates, and c) the duration of the vowel preceding the affricates. From (a) and (b) the ratio of silence/friction has been calculated. Mean durations for silent and friction portions in fast and slow speech for each affricate are given in Table 1, which also gives the percentage of reduction in duration of each portion in the fast speech compared with the slow speech.

SILENCE

		slow	fast	% reduction
Spanish		78	54	31%
English		100	68	32%
Italian	/tʃ/	60	36	40%
	/tʃ:/	147	86	41%

FRICTION

		slow	fast	% reduction
Spanish		74	65	12%
English		94	77	18%
Italian	/tʃ/	83	68	18%
	/tʃ:/	101	73	28%

TABLE 1. Mean durations in msec of silent and fricative portion of palato-alveolar affricates in three languages at slow and fast speech rates.

The overall mean durations of both the silent and fricative portions of these affricates are significantly different between the four groups, with the exception of the friction duration of English and Italian (single) /tʃ/. In other words, the Spanish affricate is significantly shorter in each portion than the English, which is in turn shorter than the Italian geminate. The Italian single affricate has a significantly shorter silent portion than any of the others.

Because speaking rate and utterance length were controlled, these differences cannot be explained by variation in rate or syllable count. However, this does not mean that they are necessarily entirely due to language-specific factors. In particular, durations of adjacent segments are often correlated. A tendency for a vowel to be inversely related to a following consonant in duration is especially widespread. The relationship of the preceding vowel to the affricate durations was therefore investigated. The durations of the low central vowels preceding the affricates in each language are given in Table 2.

VOWEL DURATION

		slow	fast	% reduction
Spanish		148	113	27%
English		85	78	8%
Italian	/tʃ/	208	153	26%
	/tʃ:/	132	112	15%

TABLE 2. Mean durations of vowels preceding affricates at fast and slow speech rates.

As expected, the geminate affricate in Italian bears a rather different relationship to the preceding vowel duration than the three single affricates. Nonetheless, it is clear that the duration difference between vowels before Italian /tʃ/ and /tʃ:/ plays a major role in signalling the single/geminate distinction.

As for the single affricates, there is a good inverse correlation between vowel duration and affricate duration. On closer examination this can be shown to be due to the correlation between vowel duration and silent duration only; the duration of friction is not well correlated with the duration of the preceding vowel. The vowel/silence relationship is rate-dependent (cf. Isenberg 1978, Repp et al. 1978) but within each rate is very precise. This is shown in Figure 1.

Although the origin of this correlation is not obvious, there are several reasons why it seems more reasonable to predict the silent duration from the vowel duration rather than vice-versa. These include the following. The occurrence of a short vowel in the English test word is expected because the low central vowel /ɐ/ is a member of a set of distinctively short vowels including /ɩ,ɷ,ɑ/ (as in putt, pit, put, pot). The occurrence of a long vowel in Italian 'cacia' is expected, as the long/short distinction is consistently involved in signalling the single/geminate contrast in following consonants including those that have no silent portion. Spanish, lacking any similar phonological involvement of vowel duration, might well be expected to have a vowel of intermediate duration. Variation in vowel duration between these languages may thus be anticipated independently of the differences observed in the affricates. Therefore it may well be that the silent duration in single affricates is not a language-specific property, but is a consequence of vowel duration differences and overall speech rate.

Friction duration is not correlated with preceding vowel duration. Furthermore, it is markedly less affected by change of speech rate than silent duration is (see Table 1). It shows only a 19% reduction in fast speech in the pooled data compared with 36% for silence. The differences between the languages in friction duration may be viewed as more representative of inherent characteristics of affricates in these languages. Spanish, with the briefest friction, has the fewest phonological distinctions to maintain on the stop/affricate/fricative continuum. English has the longest friction of the single affricates. The Italian single affricate is intermediate between these two. Such a ranking does not conform to the prediction of the original hypothesis of this paper.

However, if the friction duration is examined in terms of a proportional relationship to the silent duration, a measure is obtained which does separate the Italian affricates widely from each other and groups the Spanish and English affricates together with intermediate values. This measure is expressed as the silence/friction ratio. Figure 2 displays the means and standard deviations of this measure for the four groups.

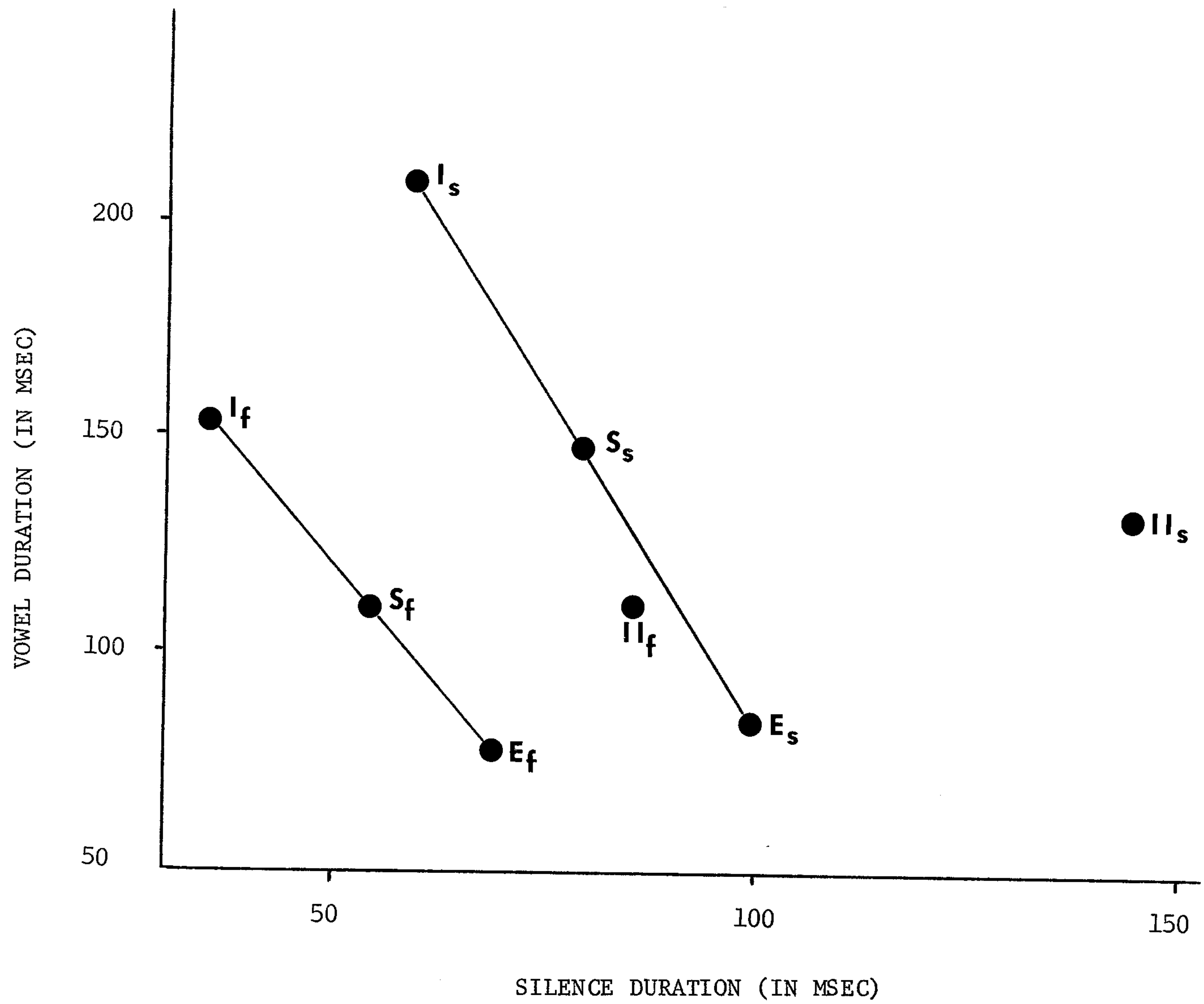


Figure 1. Relationship of preceding vowel duration and silent portion duration in affricates in Spanish (S), English (E) and Italian (I) at slow ($_s$) and fast ($_f$) speech rates. Points are also plotted for the Italian geminate (II).

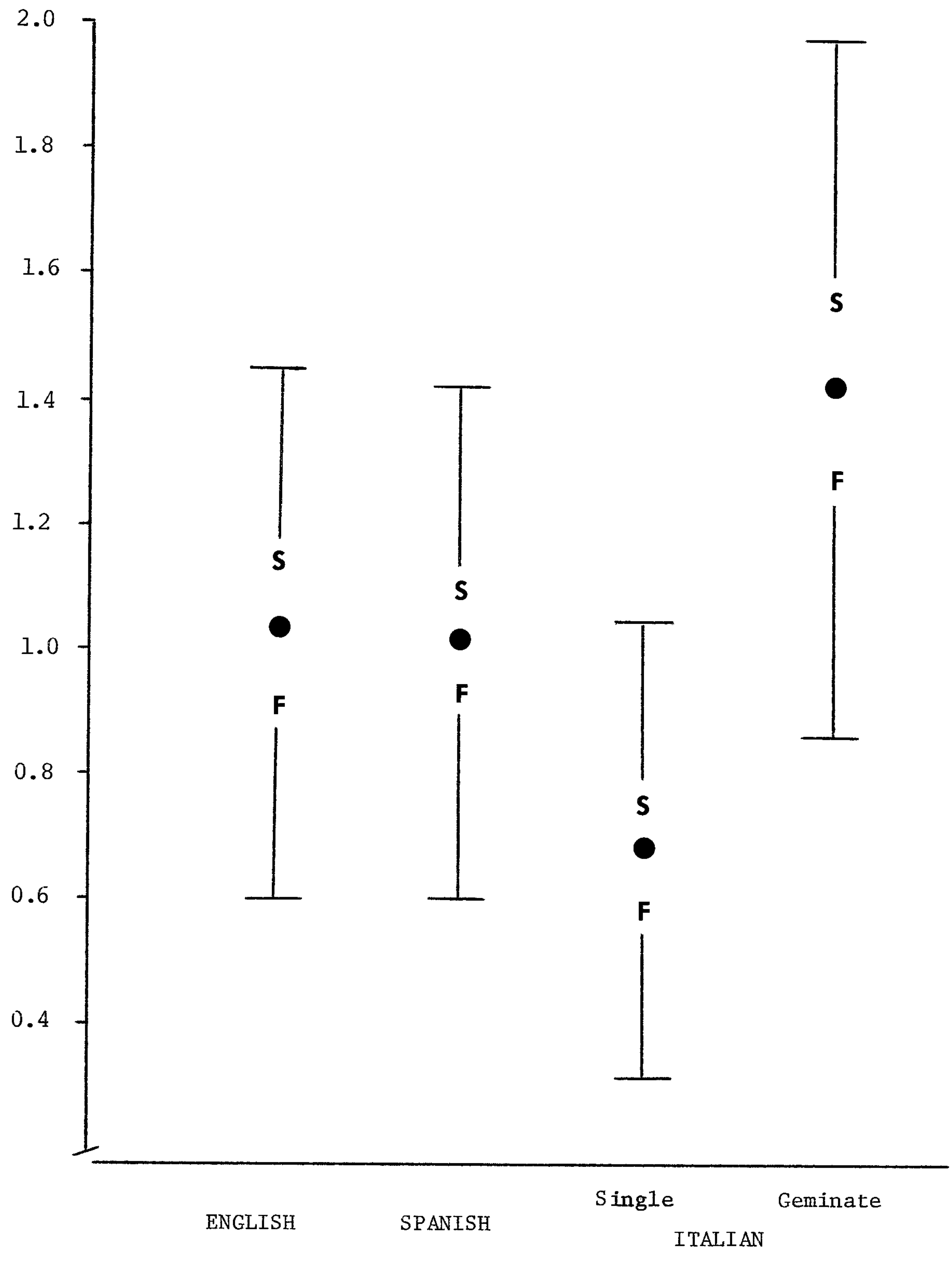


Figure 2. Overall mean (dot), and fast (F) and slow speech (S) means for the silence/friction ratio for each affricate. The bars indicate one standard deviation of the overall mean above and below the mean.

That this ratio is an appropriate measure is suggested by the finding of Repp et al. (1978: 626). In their examination of perceptual cues to the distinction between 'shop' and 'chop' in American English, they report that "within each speaking-rate condition, the fricative-affricate boundary is associated with a constant ratio between silence and noise durations". The values they obtained for this ratio at this perceptual boundary are consistent with the values found in this paper for typical naturally-produced affricates, i.e. they are at the lowest end of the range. However, the finding of a lower ratio for the slower speech rate boundary differs from the finding of consistently lower ratios for the faster speech rate in the production data in the present study. Whereas our finding is that the silent duration is more affected by the different speech rates than the friction duration, Repp et al. make the supposition that "variations in rate of articulation cause the duration of the fricative noise to change more than the duration of the silence" in order to explain their results. This matter obviously deserves further investigation.

Figure 2 also provides data to address the hypothesis that differences in variability should be correlated with differences in phonological inventory. Comparing the standard deviations of the silence/friction ratio indicates that the variability is greatest in the Italian geminate affricate, and is no less in the Italian single affricate than in the English or Spanish. Thus the prediction that variation is reduced in the presence of a larger number of competing segments is false for this measur of timing in affricates. It is also false for the separate measures of the component durations, which were compared by calculating the coefficient of variation for each of the four groups in this data.

Conclusion

Differences have been found between the affricates studied with respect to durational characteristics. We have argued that the duration of friction may be the best measure of inherent differences between the languages, as silent duration may be predicted from preceding vowel duration. However, if the silence/friction ratio is (roughly) constant, then friction duration may also be predicted from vowel duration. This may be the case in languages with only one voiceless palato-alveolar affricate, such as Spanish or English, but different ratios obtain for the affricates in the language examined which maintains a contrast of single and geminate affricates, Italian. Friction duration, or silence/friction ratio, must therefore be a parameter along which languages may differ, although such differences may be predictable in part from the structure of the phoneme inventories involved.

References

Isenberg, D. (1978) "Effect of speaking rate on the relative duration of stop closure and fricative noise" Status Report on Speech Research (Haskins Laboratories) 55/56: 63-79.

Repp, B.H., A.M.Liberman, T.Eccardt & D.Pesetsky (1978) "Perceptual integration of temporal cues for stop, fricative and affricate manner" Journal of Experimental Psychology: Human Perception and Performance 4: 621-637.

Dimensions of tone systems

Ian Maddieson

This paper describes patterns in the structure of tone systems in terms of an understanding of the relative importance of the dimensions along which tones may contrast, and explains a marked typological dissimilarity between tone and vowel systems as resulting from the different kinds of dimensionalities that underlie them.

In order to obtain reliable estimates of the variety and structure of tone systems in the languages of the world, a survey has been made of a carefully constructed sample of tone languages. The sample has been selected to represent the maximum of genetic diversity of tone languages, while reflecting the numbers of tone languages found in the various genetic groupings. Where adequate descriptive material is available, one tonal language is been included from each identified cluster of closely related languages. Figure 1 illustrates how such a selection might be made from an imagined family of languages including both tonal and non-tonal languages related to each other in the way indicated by the tree diagram. The sample includes 300 languages.

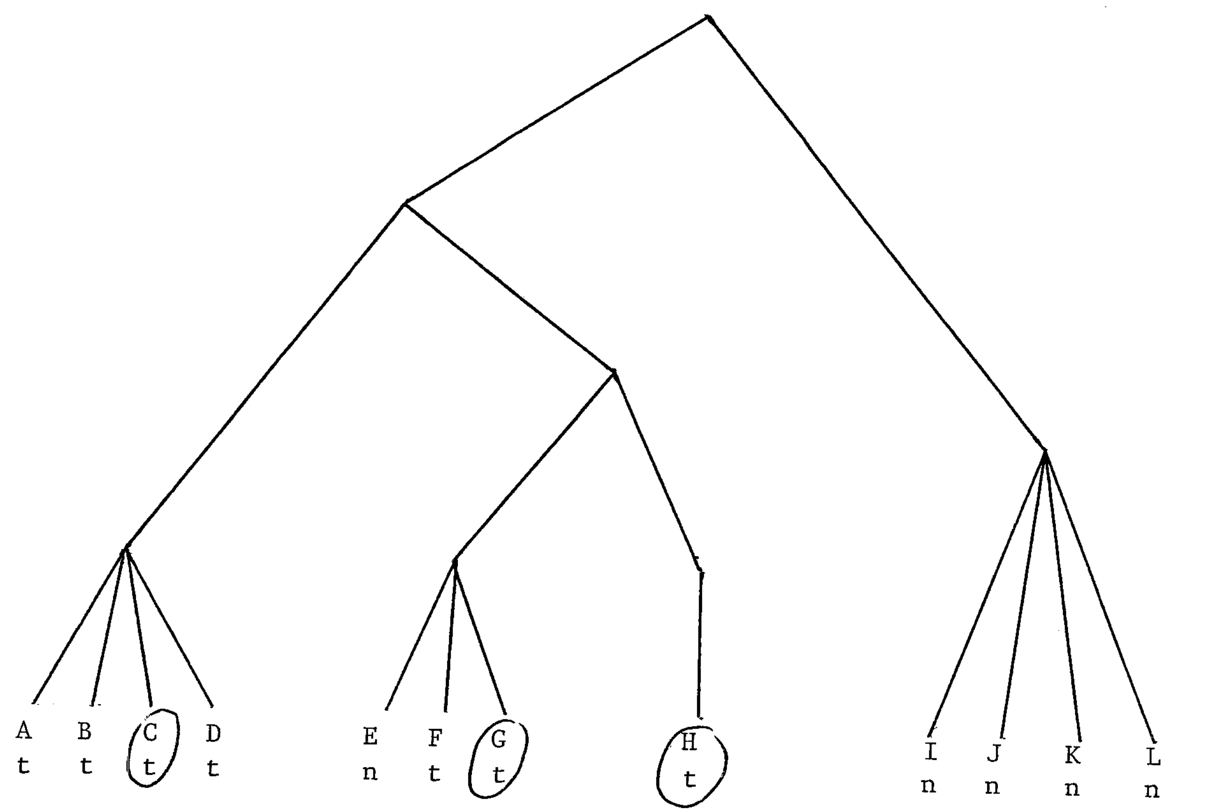


Figure 1. Imaginary family tree of languages A–L, some being tonal (t) while others are not (n). If the circled languages were selected for inclusion in the survey, the selection criteria would be satisfied with respect to both completeness and genetic diversity.

A careful attempt has been made to ensure the standardization of the descriptive level for each language. The objective has been to determine the number and nature of underlying contrastive tones that would be posited if only plausibly 'natural' rules are used, but if full reference to relevant phonological, morphological, lexical and syntactic facts is permitted. Such rules include processes such as downdrift, and normal coarticulation effects between adjacent tones or between tones and segmental properties. Tones with a marginal status in the phonology of a language, such as those with a very restricted distribution, have been excluded. Most cases of downstep and special sandhi tones are among those excluded. Only the basic pitch shape of a tone is considered in this survey. Non-pitch phenomena which co-occur with tones are regarded as independant features. Where different 'tones' are restricted to particular syllable types they have been considered contrastive only if the pitch shapes differ in ways that are not attributable to the effect of the difference in syllable type.

For the purposes of this survey it is accepted that both level tones, requiring only one pitch height target for their specification, and contour tones, requiring two or three pitch heights targets to be specified in sequence, occur as phonological units. However, where phonetic contours can be shown to undergo rules like those that affect an equivalent sequence of phonological level tones, or to derive from the combination of elements (e.g. separate morphemes) bearing underlying level tones, the phonetic contours are regarded as constituted from level tone sequences, and not as unit contour tones. On the other hand, distibutional arguments alone, such as restriction of contours to long vowels, are insufficient to justify decomposition of contours into constituent levels.

The data in the sample thus reflect a level of phonological analysis which has been made as uniform as possible. This level is similar to that of a conventional phonemic analysis. This form of analysis was chosen partly for pragmatic reasons because more phonological descriptions are available using a phonemic model than any other theoretical framework. The phonemic model is however modified to allow for 'mixing of levels' and for considerations of naturalness in the interpretation of the data.

The sample allows for estimates to be made of the frequency with which tone systems with particular numbers of tones occur. These results are given in Table 1. This shows that roughly half the languages have only

Number of tones in the inventory	percentage of languages in the survey	n
2	51%	154
3	28%	86
4	9%	27
5	4%	13
6	6%	17
7 (+)	1%	3

Table 1.

two basic tones. About a quarter have systems contrasting three tones, while the remainder contrast four or more tones, with four being markedly more frequent than five, six or seven tones. The larger the size of the tone inventory, the less frequently it is found to occur in the languages of the world. We will contrast this with the structure of vowel inventories later in the paper.

There are also structural properties of tone systems that are related to the size of the system. These are summarized in Table 2. Systems contain-

Number of tones	structure of the system	percent of cases with given structure
2	level tones only others	100% 0%
3	level tones only others	81% 19%
4	level tones only 3 levels + 1 fall 3 levels + 1 rise others	30% 22% 22% 26%
5	3 levels, 1 fall, 1 rise others	62% 38%
6	3 levels, 1 fall, 2 rises 3 levels, 2 falls, 1 rise others	24% 18% 62%

Table 2.

ing only two tones consist of level tones only. Systems containing three tones also for the most part contain only level tones. The most frequent kind of system containing four tones is one which has only level tones, but those with three level tones and a rise or a fall are both almost as frequent. Systems with five or more tones typically include both falling and rising contours. The most frequent system of five tones contains three level tones, one fall and one rise. The most frequent six tone systems contain an additional rising or falling tone, but there are several different patterns and none is clearly predominant. However, the larger inventories, with rare exceptions, contain two contour tones which move in the same direction. Two rises or two falls may be distinguished from each other by the amount of pitch change they exhibit, or they may each have the same amount of pitch change but be located in different parts of the pitch range used in the language: the systems containing two contour tones that move in the same direction are about twice as likely to distinguish them by the amount of pitch change than to distinguish them by location in different fractions of the pitch range.

Many of these frequently found patterns connecting the size and the structure of tone inventories can be summarized in simple generalizations. The smaller and most commonly found inventories typically contrast tones a along only one dimension - that of pitch level. As the size of the inventory increases, pitch movement is exploited for contrastive purposes, adding a dimension of direction - level vs falling and/or rising. The larger and

and less frequently found inventories may also exploit a dimension of amount of pitch change, for example by having a rising tone which covers the whole of the pitch range, and another which covers only a portion. These three dimensions are clearly ranked both in terms of the number of languages which use them and in terms of their exploitation within individual inventories. Pitch level contrasts have precedence over pitch direction contrasts which have precedence over contrasts of amount of pitch change.

This ranking of these dimensions of contrast in tone systems corresponds with the ranking of the cognate dimensions labeled *average pitch*, *direction*, and *slope* found by Gandour and Harshman (1978) in a perceptual study of tone. Briefly stated, Gandour and Harshman asked subjects to indicate perceived dissimilarities between pairs of tones in an inventory of nine tone shapes consisting of three level tones, three falling tones and three rising tones. (A long/short difference is ignored here). These judgments were subjected to a multidimensional scaling analysis using the PARAFAC procedure (Harshman 1970). From this analysis the three linguistically interpretable dimensions mentioned above were obtained. The perceptual dimensions can be ranked in importance by comparing the mean subject weights for each dimension. The ranking obtained is shown in Figure 2.

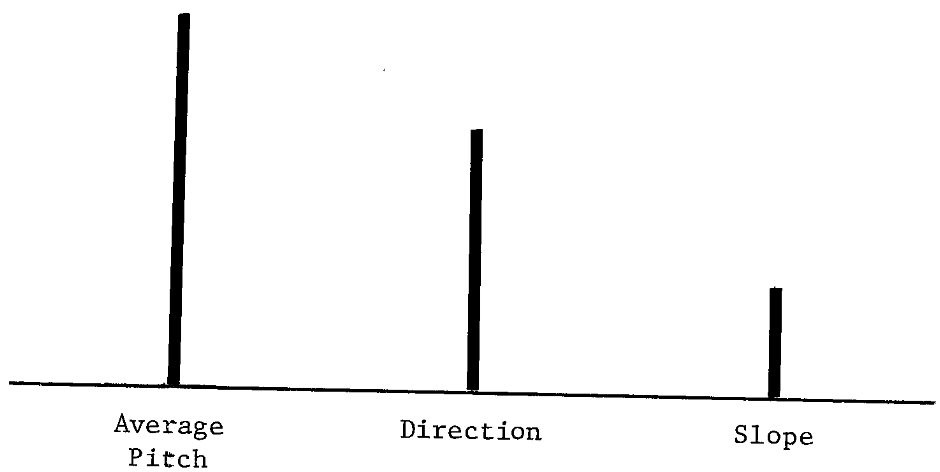


Figure 2. Relative weight of three perceptual dimensions of tones (on an arbitrary scale). Weightings calculated from the mean of group means in Gandour and Harshman (1978).

This implies, roughly speaking, that subjects typically relied most on the average pitch dimension to discriminate between tones, then next they relied on the direction dimension, and then on the slope dimension the least. Figure 2 represents the relative salience of the dimensions calculated by taking the mean of the group means of three groups of subjects, but the same rank order is repeated in the separate solutions for each group. The subjects for this study were speakers of Thai, Yoruba and English. A fourth dimension, interpreted as reflecting *extreme endpoint*, was salient in the results from the English group. It is unclear what relationship this dimension may have to the structure of tone inventories. However, the ranking of the other three dimensions remains the same for each set of subjects.

The correspondence found between the ranking of the dimensions in tone inventories and in the perception experiment suggests that, in general, tone inventories are elaborated by recruiting progressively less salient dimensions for contrastive purposes as their size increases.

It does not seem appropriate to propose that other phonological subinventories, such as vowel qualities, are constructed in the same way. Vowel systems differ from tones in the number of contrasts usually found. Several independent surveys (e.g. Crothers 1978, Maddieson 1980) agree that the most frequent number of vowel qualities found in the world's languages is five, with both higher and lower numbers of vowels being rarer. Vowel systems are only very rarely reduced to as few as two contrasting qualities. Also, in perceptual experiments on vowels, Terbeek (1977) found no reliable ranking of the six vowel quality dimensions he extracted from the judgments of subjects speaking five different languages. Rather, several dimensions are all strongly salient. It thus seems possible to typologize phonological subinventories according to whether they have a hierarchical set of dimensions defining them, like tones, or an unranked (or, at least, partially unranked) set of dimensions defining them, like vowels. This distinction may well parallel that between subsystems that typically have small numbers of contrasts or may be absent from a language altogether, and subsystems that typically have multiple contrasts and are found in all languages. Tones are an example of this first category, and liquids and clicks might well be too. Vowel qualities are an example of the second category, and plosives might be too. The availability of multiple salient dimensions for contrasts helps to explain why certain types of systems are typically more elaborate than others.

References

Crothers, J. (1978) "Typology and universals of vowel systems" in Universals of Human Language, 2: Phonology ed. J.H.Greenberg et al., Stanford University Press, Stanford: 91-152.

Gandour, J.T. & R.A.Harshman (1978) "Cross-language differences in tone perception: a multi-dimensional scaling investigation" Language and Speech 21: 1-33.

Harshman, R.A. (1970) Foundations of the PARAFAC procedure; models and conditions for an 'explanatory' multi-modal factor analysis (UCLA Working Papers in Phonetics 16).

Maddieson, I. (1980) "Phonological generalizations from the UCLA Phonological Segment Inventory Database" UCLA Working Papers in Phonetics 50: 57-68.

Terbeek, D. (1977) A cross-language multidimensional scaling study of vowel perception (UCLA Working Papers in Phonetics 37).

Pitch in Sherpa

Ian Maddieson, Sharon Hargus, Jonas N.A. Nartey

Sherpa is a Sino-Tibetan language spoken by 50,000 people in parts of Eastern Nepal. It is included with various Tibetan dialects in the Central Bodic branch of Tibeto-Burman (Shafer 1974). Our analysis of pitch in Sherpa refers to the speech of only one individual, Konchhok Zangbu Lama of Trakshingdu monastery, Solu Khumbu, and we recognize that it may not apply entirely to all varieties of Sherpa.

During the course of field-work on Sherpa, a wide range of phonetic pitch levels was observed in the speech of our consultant. When such differences are found in similar environments in a language that is presumed to be tonal, it usually indicates that a large number of contrastive tones exist. Research on phonetic properties of tone systems has shown that the more tone levels a language has, the more phonetic space between the extremes of the tone system will be required (Maddieson 1977, 1978, 1979). In Sherpa, we found high and low tones differing by as much as 70 Hz in a normal male voice. Thus, we expected that there would also be contrastive tones at intermediate levels. However, minimal tonal contrasts between segmentally identical syllables were surprisingly hard to find. Despite our perception of various 'mid' tones, we could only feel confident of a binary opposition in tone. On the other hand, it was sometimes difficult to agree on and describe the crucial difference between two lexical items which our consultant would maintain to be different.

This paper presents the results of a phonetic investigation into pitch designed to resolve these matters. We show that Sherpa does in fact have a basically simple tone system, but that a number of rules and the interaction of tone with other aspects of the phonology combine to present the puzzling superficial complexity we first observed.

For our phonetic investigation a selected word-list of 274 items was prepared. These were written on prompt cards in the Tibetan script. Three complete readings of this list were recorded by our consultant, each item being read first in isolation and then in the sentence frame below.

ŋà	náklúŋ	________	sìrìn
I	again		say-be

"I (will) say ________ again."

Narrow-band spectrograms have been made of all items in the first reading. Spectrograms have also been made of selected items in the second and third readings. Thus, between two and six examples of each word have been analyzed. Measurements were made of the pitch at the beginning and the end

Consonants.

	bilabial	dental	alveolar	retroflex	palatal	velar	
stops	p p^h b	t t^h d		ṭ $ṭ^h$ ḍ	c c^h ɟ	k k^h g	
affricates			ts ts^h		tç $tç^h$		
fricatives			s z		ç ʑ		
nasals	m	n			ɲ	ŋ	
approximants					j		h
laterals			l̥ l				
trills		r̥ r					

Vowels.

short	long
i u	iː uː
e o	eː oː
	ɛː ɔː
a	aː

Table 1 Phoneme Inventory of Sherpa.

of each vowel and also at the highest pitch point in the vowel where this differed from the beginning or the end. These measurements and visual examination of the spectrograms formed the basis for a classification of the words into groups according to their pitch patterns and other relevant properties. In addition, a careful auditory analysis of all three readings was prepared by each of the authors. The greater percentage of the words on our word-list are monosyllabic in structure, and it is these monosyllables that will be the focus of attention in our report. Disyllabic and longer words appear to have less of the superficial complexity we are examining.

We have identified certain factors which contribute to the apparent multiplicity of tones at the surface level of analysis. And, at the same time, we have reached a better understanding of certain limits on the distribution of consonants, tones and voice qualities which reduce the contrastive possibilities of the system. It is the cumulative effect of these two kinds of processes that produces the confusing situation which was the original motivation for our study.

We believe that Sherpa has a basic phonological contrast only between high and low tones. This simple situation is complicated by the occurrence of an apparent set of mid tones and by a largely predictable distinction between falling and level tone shapes. The apparent distinction between high and mid tones can be shown to be predictable on the basis of a contrast in vowel length throughout the system that is unrecognized in the previous literature on Sherpa (see Bibliography). The phonemes of Sherpa are given in Table 1.

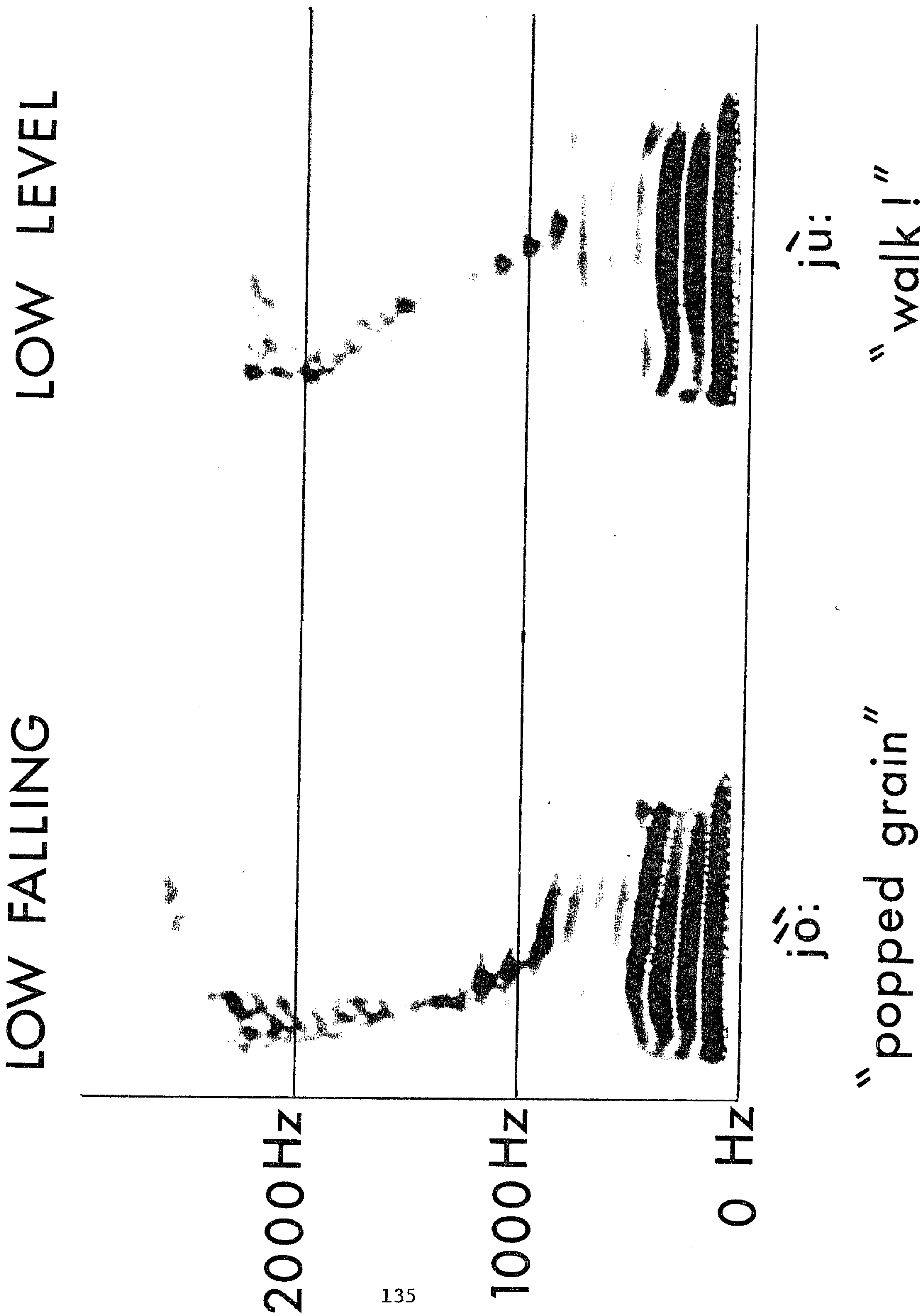
Pitch is considerably higher on a short vowel than it is on a long vowel in a word belonging to the high-tone group. This can be illustrated with the data in Table 2.

	n	peak pitch mean	peak pitch s.d.	vowel duration mean	vowel duration s.d.
a) CV̋ words	27	184 Hz	9.9	168 msec	22.3
b) CV̋: words	12	158 Hz	4.9	362 msec	33.1
significance of difference		p. > .0001		p. > .0001	

Table 2 Peak pitch and vowel duration of open monosyllabic high-falling tone words with (a) short and (b) long vowels.

This shows the mean peak pitch value in all of the high-falling words in the first reading with short and long vowels respectively in open syllables, and the duration of the vowels in these two groups of words. The peak pitch is the highest pitch measured in the item. For many items the highest value is the measurement at vowel onset, but in other cases a higher pitch value is obtained at some later point in the vowel. The peak pitch often seems a more stable measure (i.e., has less variance) of the

Figure 1.



LOW FALLING
LOW LEVEL
2000Hz
1000Hz
0 Hz
jô:
"popped grain"
jû:
"walk !"

inherent pitch of a word or class of words. The peak pitch mean includes both the isolation and sentence contexts, but vowel duration is represented only by measures made on the isolated words, since duration is significantly affected by context.

Although the difference in pitch is statistically significant, the contrast between high and mid cannot be phonemic because of the differences in vowel length. Vowel length is also contrastive with low-tone words (cf. /d̤à̀:/ "rice", /d̤à̀/ "enemy"), and in this case does not co-occur with significant pitch differences. (These differences are substantially less easy to discriminate than corresponding pairs of high-tone words.) The mean peak pitch for long low-falling words is 130 Hz and for short low-falling words it is 131 Hz. Therefore, it is not appropriate to predict vowel length in high-tone words from the pitch difference. It should also be noted that this pitch difference is not plausibly due to segmental differences between the words examined. Although the two sets of high-tone words involved are not exactly matched in segmental composition, they have broadly similar frequencies of segment types. For example, the long vowel set has 85% initial voiceless obstruents, and the short vowel set has 83%.

We therefore propose that Sherpa has a rule of high-tone pitch-raising affecting (at least) monosyllables with short vowels. If the rule is stated in this way, then the narrower pitch range used for the high/low contrast in long vowels is regarded as the basic pitch range of the language, and Sherpa is consequently less out of line with other languages with similar numbers of tonal contrasts than it at first appeared to be.

The second complication of surface tonal patterns concerns the distinction between falling and level pitch patterns. The majority of items in our sample have a falling pitch contour, as in the high-falling sets just discussed. However, a minority of our words have level pitch. These level patterns may occur with low-toned long and short vowels, or with high-toned long vowels. Although the distinction between level and falling is not salient, the difference is consistently observed. Narrow-band spectrograms of a near-minimal pair of words which differ in this way are shown in Figure 1. Both words are low-toned and have long vowels. (We have transcribed the falling pitch by doubling the tone-mark.) This difference in pitch pattern is largely predictable on the basis of lexical class. In items on our word-list, it redundantly reinforces the difference between nouns and non-finite verbs--imperatives and infinitives. The items in Figure 1 are an example of this distribution.

A small number of nouns, however, occur with level pitch. Deictic words and sentence particles may be found in either the level or falling class. For example, /dɛ̀̀:/ "here" has a falling pitch contour, whereas /tɛ̀:/ "there" is level. Despite such examples, the functional load of the contrast between level and falling patterns is extremely small. Nonetheless, this is a reliable and consistent distinction. In Table 3 we show the mean pitch values for the ends of low-toned, long vowels in open syllables of the first reading, for falling and level pitch patterns, respectively.

		n	final pitch mean	final pitch s.d.
a)	CV̀:̀	25	93 Hz	4.4
b)	CV̀:	16	110 Hz	6.7
significance of difference			p. > .0001	

Table 3 Pitch at the end of the vowel in open monosyllabic words with long vowels and (a) low-falling or (b) low-level pitch pattern.

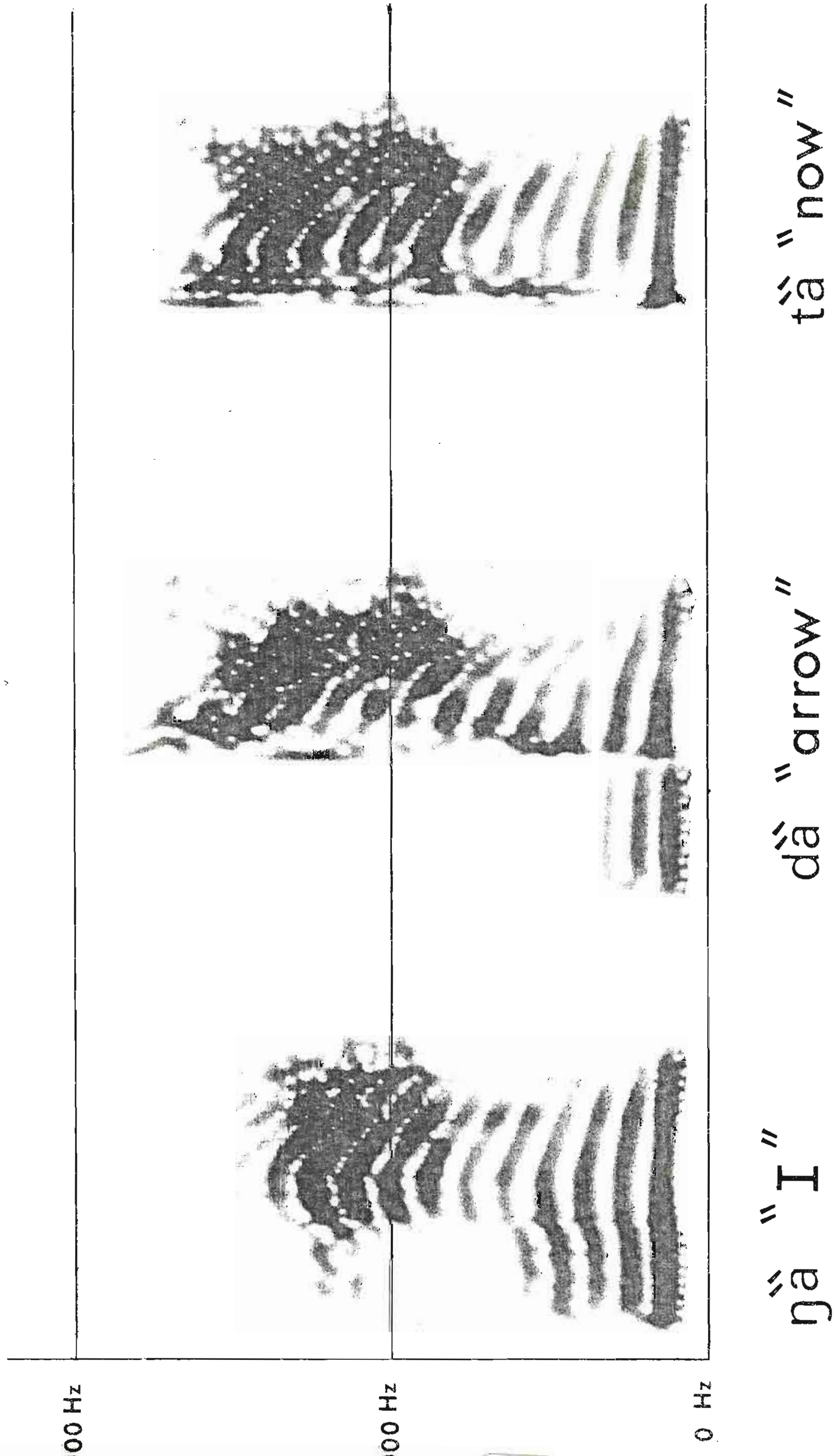
Although the mean peak pitch levels of these two groups of words are only 2 Hz apart at 130 Hz and 128 Hz respectively, the vowel-final pitch levels for the groups are 17 Hz apart and this difference is highly significant. For this reason, perhaps, low-tone level syllables were fairly consistently transcribed as 'mid' in our auditory analysis. The great majority of the observed level pitches can be accounted for by a rule stating that non-finite verbs, except for those with short vowels and high tone, have level pitch. Elsewhere, tones are falling in pitch.

There is also a handful of "sentence particles" with idiosyncratic tone patterns. This includes the words /ɬâ̌ː/ "yes", with a rising-falling pattern, and/rǎ̌ːŋ/ "also, at all", with a rising contour. Since such contours are restricted to such special items only, we feel justified in regarding them as outside the basic tonal system of Sherpa.

We have now described two processes, high-tone raising in short vowels, and lexically-governed derivation of level pitch, which require statement as extrinsic rules of the language. In addition to these, there are pitch differences which correlate with the segmental composition of the word and which may be regarded as intrinsic effects of the particular way such segments are produced in Sherpa.

Our investigation of these effects is far from complete, but we would like to mention some of our findings. First of all, voiced continuants, most especially nasals, seem to have a marked lowering effect on the beginning of the following vowel. This may be illustrated by the spectrograms of the three words /ŋȁ/, /dȁ/, and /tȁ/ in Figure 2. These words all have low-falling tone and short vowels. Apparently, Sherpa nasals are intrinsically low-pitched. The lowering effect of nasals occurs because they transmit their low pitch through to the following vowel. There is a gradual transition in pitch from the nasal to the vowel in /ŋȁ/, which begins in this token at 121 Hz. Compare this with the very abrupt transition in pitch between the voiced stop and the vowel in /dȁ/. In this case, notice that although pitch is low during the closure for /d/, there is no lowering of the following vowel, which begins at 139 Hz. The lowness of the /d/ is not transmitted to the vowel. In fact, the vowel in /dȁ/

Figure 2.



begins higher than the vowel in /tȁ/, which in this token also begins at 121 Hz.

Indeed, low tone vowels generally tend to be higher after voiced stops in Sherpa than after their voiceless unaspirated counterparts. Table 4 shows the pooled means of initial and peak pitch for all six tokens of three pairs of words differing only in initial consonant voicing. These are labelled the D-group and the T-group.

		initial pitch			peak pitch		
	n	mean	s.d.		mean	s.d.	
a) D-group dȁ ḍȁ ḍȕ "arrow" "enemy" "grain"	18	132 Hz	11.9		137 Hz	7.4	
				n.s.			p.> .01
b) T-group tȁ ṭȁ ṭȕ "now" "wheat" "boat"	18	128 Hz	6.1		130 Hz	4.7	
				p.> .05			n.s.
c) ŋȁ "I" (1 sg. pn.)	6	119 Hz	6.1		125 Hz	5.6	
				n.s.			p.> .05
d) lȁ "hill"	6	119 Hz	3.1		119 Hz	3.1	

Table 4 Comparison of initial and peak pitch measures following different consonants in selected open monosyllabic words with low falling tone and short vowels. Significance levels for the differences between adjacent pairs are shown by the bracketing to the right of the means for each measure. Non-adjacent pairs are significant at a level equal or superior to the level of intervening pairs.

There is no significant difference between these two groups in the initial pitch of the vowel, itself a surprising finding, and when the peak pitches are compared, the D-group is significantly higher than the T-group (The test used for significance was Duncan's multiple range test.) Both the D-group and the T-group are significantly higher in initial pitch than the six tokens of /ŋȁ/ and /lȁ/ ($p > .05$). The lowering effect of voiced continuants can also be seen in the data given in Table 5, which compares initial and peak pitch after various consonants in high-falling tone words with short vowels.

		n	initial pitch mean	initial pitch s.d.	peak pitch mean	peak pitch s.d.
a)	l̥a̋ "god"	6	179 Hz	7.3	182 Hz	7.0
b)	TH-group: tʰa̋ "end", tsʰa̋ "salt", pʰű "high country"	18	178 Hz	13.7	184 Hz	10.3
c)	T-group: ta̋ "horse", tsa̋ "grass", pű "skin hair"	18	176 Hz	10.0	181 Hz	8.9
d)	la̋ "wages"	6	170 Hz	5.2	170 Hz	5.2
e)	ŋa̋ "five"	6	154 Hz	7.8	160 Hz	8.2

Bracketing: initial pitch, a)–d) n.s.; d)–e) p.>.05. Peak pitch, a)–c) n.s.; c)–d) p.> .05; d)–e) n.s.

Table 5 Comparison of initial and peak pitch measures following different consonants for selected open monosyllabic words with high falling tone and short vowels. Significance is indicated by the bracketing in the same way as for Table 4.

Initial pitch in /la̋/ and /ŋa̋/ is lower than in /l̥a̋/ and in two groups of three words each, which represent voiceless aspirated and unaspirated stops and affricates (labeled TH-group and T-group, respectively). The difference here is only statistically significant for /ŋa̋/. However, if peak pitch is compared, there is a significant difference between both /la̋/ and /ŋa̋/ and the comparison items.

Note that there is no significant difference between the aspirated and unaspirated groups of these high-tone words. In the limited data available for studying the effect of aspiration with low tones, shown in Table 6, there is conflicting evidence.

			n	initial pitch mean	initial pitch s.d.		peak pitch mean	peak pitch s.d.	
1. a)	tçà:ŋ	"North"	6	126 Hz	6.6	n.s.	126 Hz	6.6	p.>.05
b)	tçʰà:ŋ	"beer"	6	116 Hz	6.0		118 Hz	6.4	
2. a)	gà:	"liking"	6	124 Hz	9.1	n.s. (a–b); p.>.05 (a–c)	125 Hz	12.1	n.s.
b)	kʰà:	"snow"	6	118 Hz	7.7	n.s. (b–c)	121 Hz	3.7	
c)	kà:	"mountain"	6	112 Hz	7.4		118 Hz	3.7	

Table 6 Comparison of initial and peak pitch measures for two sets of low tone words with a minimal contrast involving aspiration. Statistical analysis of these items was performed jointly with those in Table 4. Note that /gà:/ and /kà:/ differ significantly in initial pitch though neither is significantly different from /kʰà:/.

Whereas the peak pitch in /tcʰà:ŋ/ is significantly lower than in /tcà:ŋ/, in /kʰà:/ the pitch is higher than in /kà:/, although the difference does not reach significance. Sherpa has no voiced affricates, but does have voiced stops. There is a significant difference in initial pitch between /kà:/ and /gà:/; and /kà:/ is also significantly lower in both initial and peak pitch than the T-group and D-group words with low tones and short vowels in Table 4. Whether inconsistencies of this sort can be entirely accounted for by the effects of place and manner of articulation differences between the consonants involved is not clear. What is clear is that intrinsic segmental effects of this kind can be quite noticeable in our consultant's speech, and, especially as some are of an unexpected nature, they also contribute to creating the superficial impression of a complex tone system.

We now turn away from discussion of the factors leading to the multiplicity of surface pitch patterns, and move on to the opposing tendency in Sherpa which restricts the large number of theoretically possible combinations within the phonological system.

The consonant inventory of Sherpa has been given in Table 1. Sherpa has voiced, voiceless unaspirated and voiceless aspirated stops. In addition there are voiced and voiceless fricatives, and voiceless unaspirated and voiceless aspirated affricates. The voiced obstruents are found only in low-tone syllables; they may not initiate high-toned syllables. Therefore, voiced, voiceless unaspirated and voiceless aspirated stops contrast only in low-tone syllables.

However, another restriction acts to reduce this already limited voicing contrast among stops. The aspirated stops hardly ever occur in the low-tone syllables, and the same is true of the aspirated affricates and voiceless liquids. In our corpus, we have only one minimal triplet which contrasts voiced, voiceless unaspirated and voiceless aspirated stops is this way (see Table 6). As a result of these restrictions, syllables which begin with obstruents generally cannot contrast minimally in tone unless the obstruent they begin with is voiceless and unaspirated.

In addition to tonal contrasts and vowel length contrasts, Sherpa vowels may differ in phonation type. Besides plain voicing, both breathiness and creakiness have been observed. The occurrence of breathiness is restricted to low-tone syllables, and is an obligatory correlate of vowels following voiced stops. Therefore, for syllables initiated by stops, breathiness contrasts are again found only after voiceless stops. Laryngealization of vowels is likewise restricted to low tones and, moreover, to long vowels of certain qualities only, namely /ɛ̀ː/, /àː/, and /ɔ̀ː/. These distinct phonation types have not yet been measured or investigated by other than auditory means.

These various factors--breathiness, creakiness, voiced obstruent distribution, and distribution of aspirates--combine with the fundamental distinctions between high and low tones, short and long vowels, to permit only a limited set of the theoretically possible contrasts between Sherpa syllables to occur. In other words, these segmental features serve largely as redundant reinforcement of the more basic tonal and vowel length distinctions between syllables.

Our phonetic analysis of the phonology of Sherpa has enabled us to reach an understanding of the basically simple tonal system and its complex interaction with other aspects of the phonology. At the same time, we are now better able to account for the superficial multiplicity of pitch levels we first encountered during auditory analysis of the language.

Bibliography.

Glover, Warren. (1971) "Register in Tibeto-Burman Languages of Nepal: A Comparison with Mon-Khmer." _Papers in South East Asian Linguistics_, 2. (Pacific Linguistics, Series A, No. 29.) 1-22.

Gordon, Kent. (1970) "Sherpa tone and higher levels." In Hale & Pike, Part I. 186-206.

Gordon, Kent & Burkhard Schoettelndreyer. (1970) "Sherpa segmental synopsis." In Hale & Pike, Part 1. 345-67.

Grierson, George, ed. (1909) _Linguistic Survey of India, Volume III, Part_ I. Reprinted 1967; Delhi: Motilal Banarsidass.

Hale, Austin. (1968) "Some dimensions of contrast in Sherpa phonology." _Journal of the Tribhuvan University_, Special linguistic number. 89-104.

Hale, Austin. (1973) _Clause, sentence, and discourse patterns in selected languages of Nepal. Part IV: word lists_. (Summer Institute of Linguistics Publications, 40.) Norman, Oklahoma.

Hale, Austin & Kenneth Pike, eds. (1970) _Tone Systems of Tibeto-Burman Languages of Nepal_. In four parts. (Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, 3.) Urbana: University of Illinois.

Maddieson, Ian. (1977) "Tone loans: a question concerning tone spacing and a method of answering it." _UCLA Working Papers in Phonetics_ 36. 49-83.

Maddieson, Ian. (1978) "Universals of Tone." J.H. Greenberg, et al., eds. _Universals of Human Language, 2: Phonology_. Stanford: Stanford University Press. 333-365.

Maddieson, Ian. (1979) "Tone spacing: evidence from bilingual speakers." _UCLA Working Papers in Phonetics_ 45. 84-88.

Pittman, Richard. (1970) "Gurung, Tamang, Thakali, Sherpa, and Chepang Prosodies." In Hale & Pike, Part II. 1-8.

Schoettelndreyer, Burkhard. (1975) "Vowels and tone in the Sherpa verb." _Nepal Studies in Linguistics_, 2. 59-70.

Schoettelndreyer, Burkhard & Austin Hale. (1970) "A note on Sherpa vowels." In Hale & Pike, Part I. 368-80.

Shafer, Robert. (1974) _Introduction to Sino-Tibetan_. Wiesbaden: Otto Harrassowitz.

Sprigg, R.K. (1980) "'Vocalic Alternation' in the Balti, the Lhasa, and the Sherpa verb, as a guide to alternations in written Tibetan, and to proto-Tibetan reconstruction." _Bulletin of the School of Oriental and African Studies_ 43. 110-22.