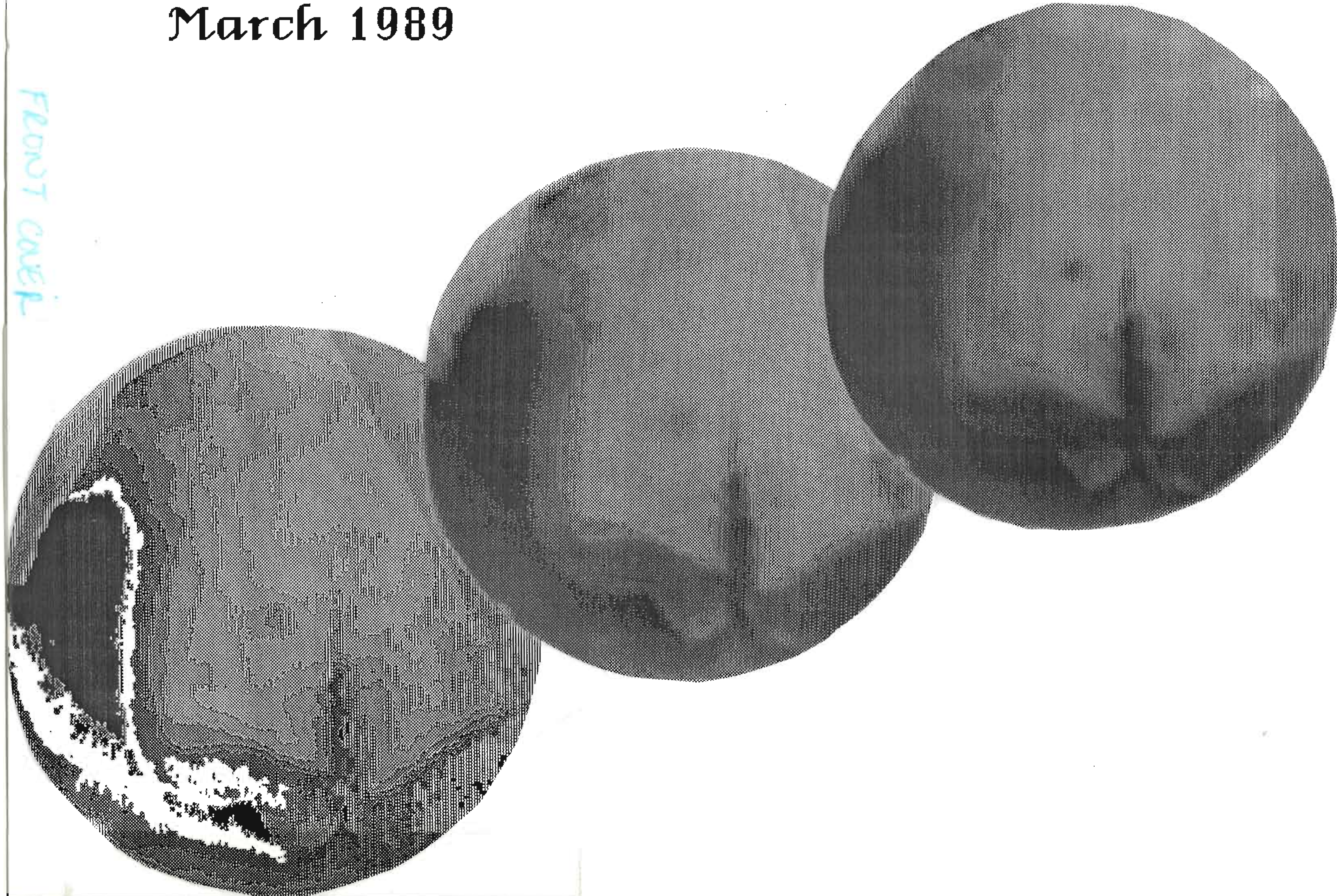
Title
WPP, No. 72

Publication Date
1989-03-15

UCLA Working Papers in Phonetics 72
March 1989

FRONT COVER



The UCLA Phonetics Laboratory Group

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

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

Funds for the UCLA Phonetics Laboratory are provided through:

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

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

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

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

UCLA Working Papers in Phonetics 72

March 1989

Contents

NOTES AND NEWS

Cross-linguistic articulatory models of vowels

(Presented at the 114th Meeting of the Acoustical Society of America, Miami, Florida 20 Nov 1987)

Michel T. T. Jackson

The goal of this paper is to present a quantitative cross-linguistic articulatory model of vowel production. The model is based on sagittal X-ray profiles of vocal tracts during vowels. X-rays of four speakers of Akan (two male, two female), three of Chinese (sex unknown), and four of French (two male, two female) from the literature were used to construct the model. The Akan data were courteously provided by Mona Lindau (1979, 1986). Two of the Chinese speakers are from Ohnesorg & Svarny (1955); the third is from Zhou & Wu (1963). The French data are from Bothorel, Simon, Wioland, & Zerling (1986).

Measurements of articulator position in each vocal tract profile were made on measurement grids similar to those used by Harshman et al. for tongue positions in English, with some additions and modifications suggested by other work in the field. A typical measurement grid, for one of the French speakers, is shown in Figure 1. The large dots on the figure mark the positions on the lips, jaw, tongue, velum, dorsal wall of the pharynx, epiglottis, and larynx that were measured. In this way, a total of 4559 measurements were collected (Table 1). Most of the missing points are due to the fact that tracings were not available for the vowel /a/ as produced by two of the Akan speakers. Another missing set is due to the fact that larynx positions were not traced in the X-rays of Chinese vowels.

Table 1: Measurements collected

Language	speakers	vowels	measures	missing	total
Akan	4	x 9	x 44	- 117	= 1467
Chinese	3	x 5	x 44	- 17	= 643
French	4	x 14	x 44	- 15	= 2449
					4559

Before plunging into detail, I would like to discuss the mathematical structure of the model, and to show how it is related to the model that has been proposed by Lindblom & Sundberg (1971). Lindblom & Sundberg model the space of possible tongue shapes

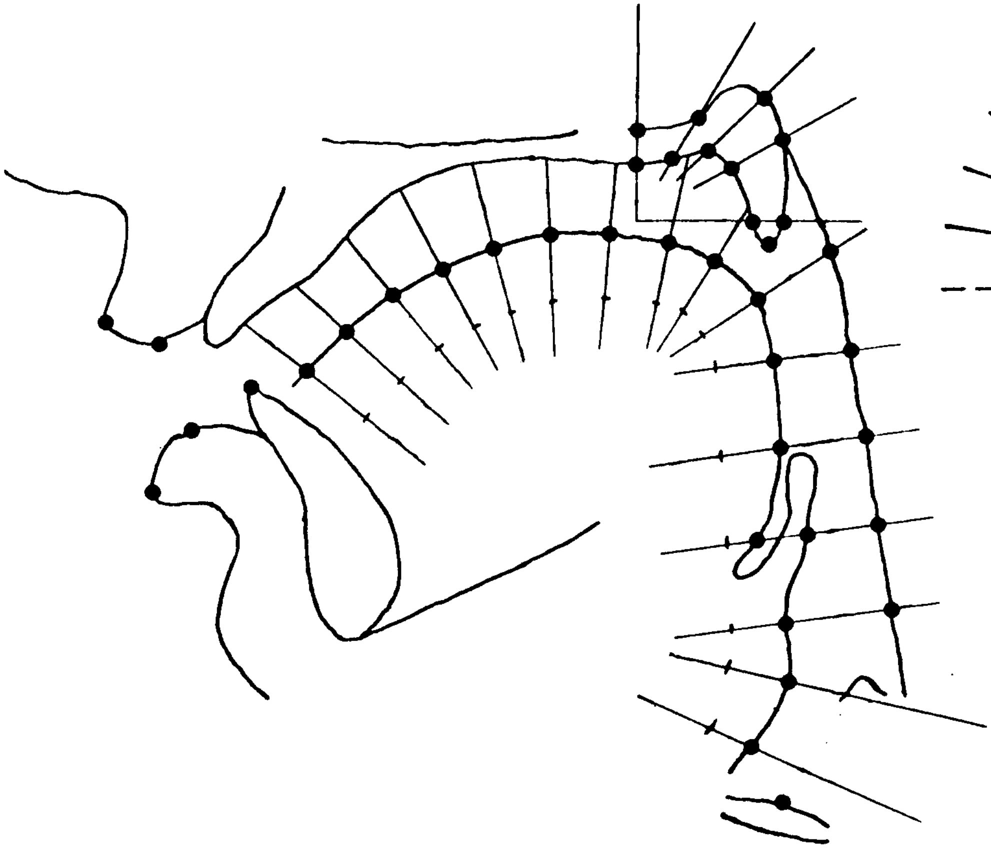


Figure 1: A measurement grid.

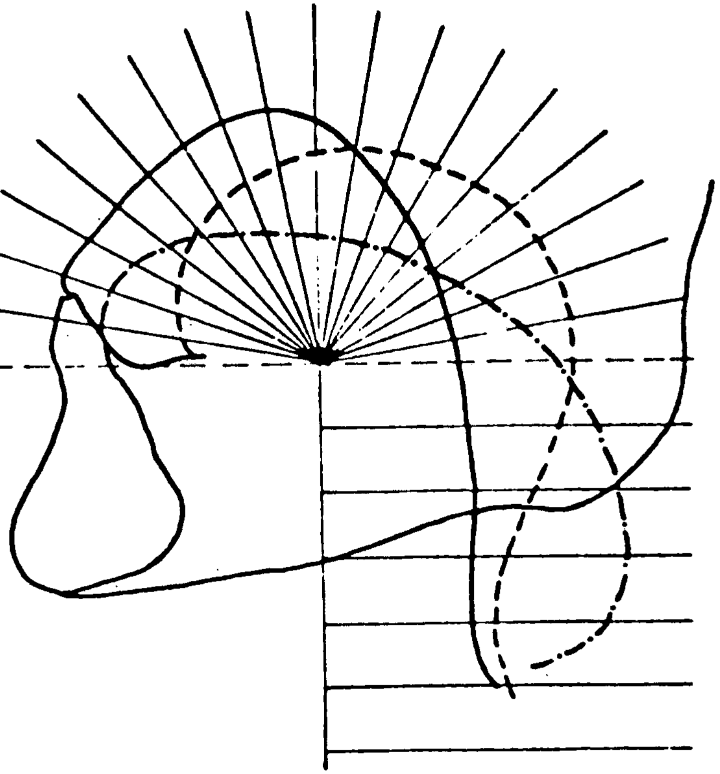


Figure 2: Lindblom and Sundberg's scheme for interpolating tongue shapes.

during vowels by interpolating between four given tongue shapes - a neutral shape, and the ones observed for the vowels [i], [u], and [a] as produced by a speaker of Swedish (Figure 2). This scheme is motivated by Lindblom & Sundberg's observation that tongue shapes during the articulatory extremes of vowels seem to fall into a few distinct families, exemplified by the given vowels. Other vowels are naturally represented by less extreme members of these families, as in Figure 3. This representation treats the vowel space as a set of planar patches, with the given vowel shapes at its apexes, and other vowels in between.

The model itself can be considered an operator that maps parameter sets into articulatory configurations. The parameters are interpolation coefficients giving the proportional contribution of each of the given shapes; the model itself can be represented as a linear operator or matrix. The elements of the matrix are the measurements of tongue position with reference to the neutral position. A parameter vector is converted into an articulatory position by multiplying it with the matrix to perform the interpolation, and then by adding the neutral position, which serves as the origin of the vowel space. Let us denote the position of the tongue during an arbitrary vowel along the *i*th gridline of Figure 2 by y_i, and similarly the position of the tongue along the *i*th gridline during the neutral vowel, [i], [u], and [a] by n_i, i_i, u_i, and a_i, respectively. The matrix **X** that embodies the model may be written as

$$\begin{array}{ccc} i_1-n_1 & u_1-n_1 & a_1-n_1 \\ i_2-n_2 & u_2-n_2 & a_2-n_2 \\ i_3-n_3 & u_3-n_3 & a_3-n_3 \\ \vdots & \vdots & \vdots \end{array}$$

Similarly, let us denote the proportional contribution of [i], [u], and [a] tongue shapes to the tongue shape in the arbitrary vowel by p_i, p_u, and p_a, respectively. Form column vectors **y** and **p** containing the (y_i) and (p_v). Then the interpolated vocal tract shape **y** can be represented in the model as in (1).

$$(1)\ \mathbf{y} = \mathbf{n} + \mathbf{Xp}$$

Lindblom & Sundberg regard the [i-u-a] dimension as a single continuum, and thus constrain only one of p_i and p_a to be non-zero for any particular vowel specification. This

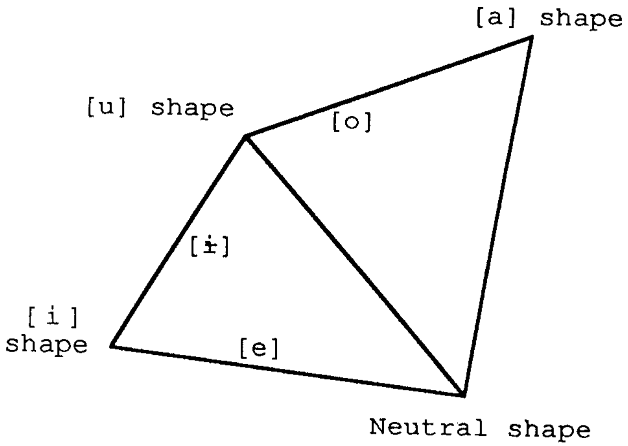


Figure 3: The vowel space as modeled by Lindblom and Sundberg.

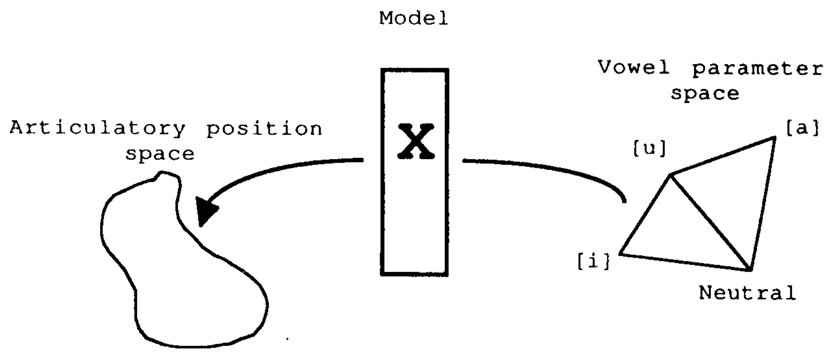


Figure 4: The Lindblom and Sundberg model as an operator from the vowel space to the space of possible tongue positions.

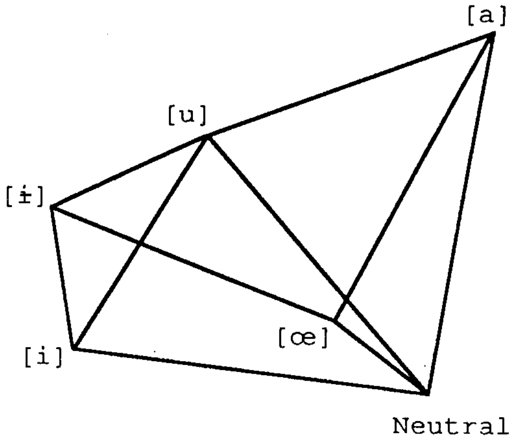


Figure 5: Possible extensions of Lindblom and Sundberg's model.

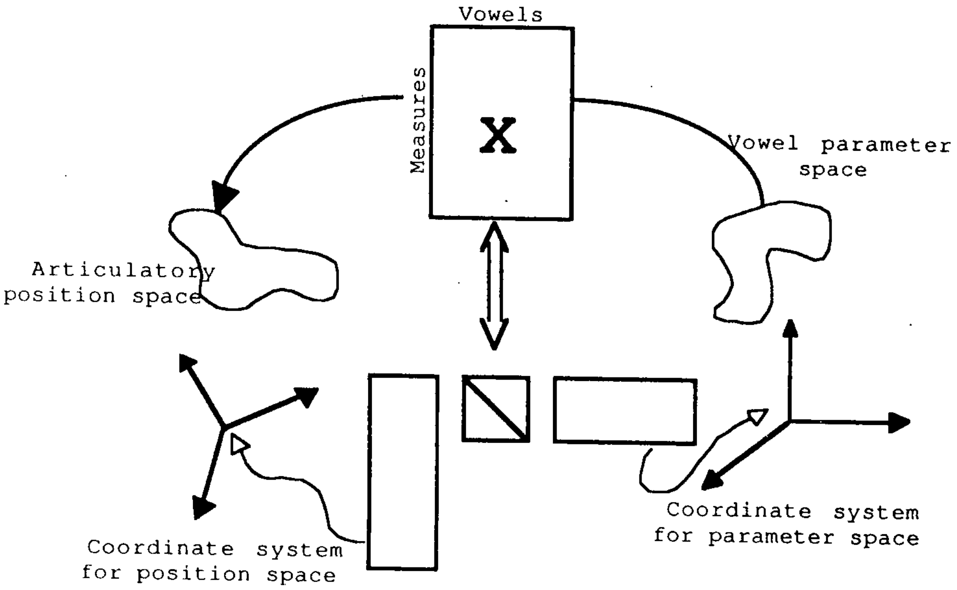


Figure 6: SVD of the operator.

model maps vectors corresponding to points of the vowel space to tongue positions, as schematized in Figure 4.

One can, however, legitimately ask whether the particular vowels used exhaust the families of tongue or vocal tract shapes. It would be simple enough to extend the model by adding more and more vowels to the sample, thus enclosing a vowel volume in some complicated polyhedron (Figure 5). But then we would want to know what constituted a sufficient set of landmarks in the vowel space, what the articulatory relationships between the various vowels were, and how well the model generalized across speakers.

The singular value decomposition (SVD) can be used to answer some of these questions. The SVD of a matrix that represents linear operator consists of three matrices: a matrix that provides a basis for the domain of the operator (the vowel parameter space); a matrix that provides a basis for range (the articulatory space); and a diagonal matrix that has weights proportional to the contribution of the corresponding element of the basis to the data sample. The decomposition of a linear operator such as the operator we have constructed above is schematized in Figure 6. (For further discussion of the SVD, see Press, Flannery, Teukolsky, & Vetterling 1986 and references contained therein).

The PARAFAC procedure (Harshman & Lundy 1984ab) is a multispeaker generalization of the SVD (see, e.g., Carroll & Chang 1970 or Kruskal 1984), and can be applied in much the same way as the SVD to analyze the characteristics of sets of operators. Figure 7 shows the design of a typical PARAFAC study. Given several speaker's mappings from a specific set of vowels to a set of articulatory postures, PARAFAC produces three kinds of matrices. Again, it produces a basis or coordinate system for the domain of these mappings, a vowel parameter space. Second, it produces a basis for the range space of articulatory postures used by the speakers. Third, PARAFAC produces a set of diagonal matrices, one per subject, that contain weights indicating the contribution of the corresponding element of the basis to the data set. These sets of matrices are known as factors, since their product estimates the mapping. The range basis and subject weight terms can be lumped to form an articulatory model. Just as in Lindblom & Sundberg's model, the product of this matrix and a vector of vowel parameters yields an articulatory posture.

There is no *mathematical* necessity that the subject-weight matrices be diagonal, and supposing that they are constitutes a strong hypothesis about the nature of the

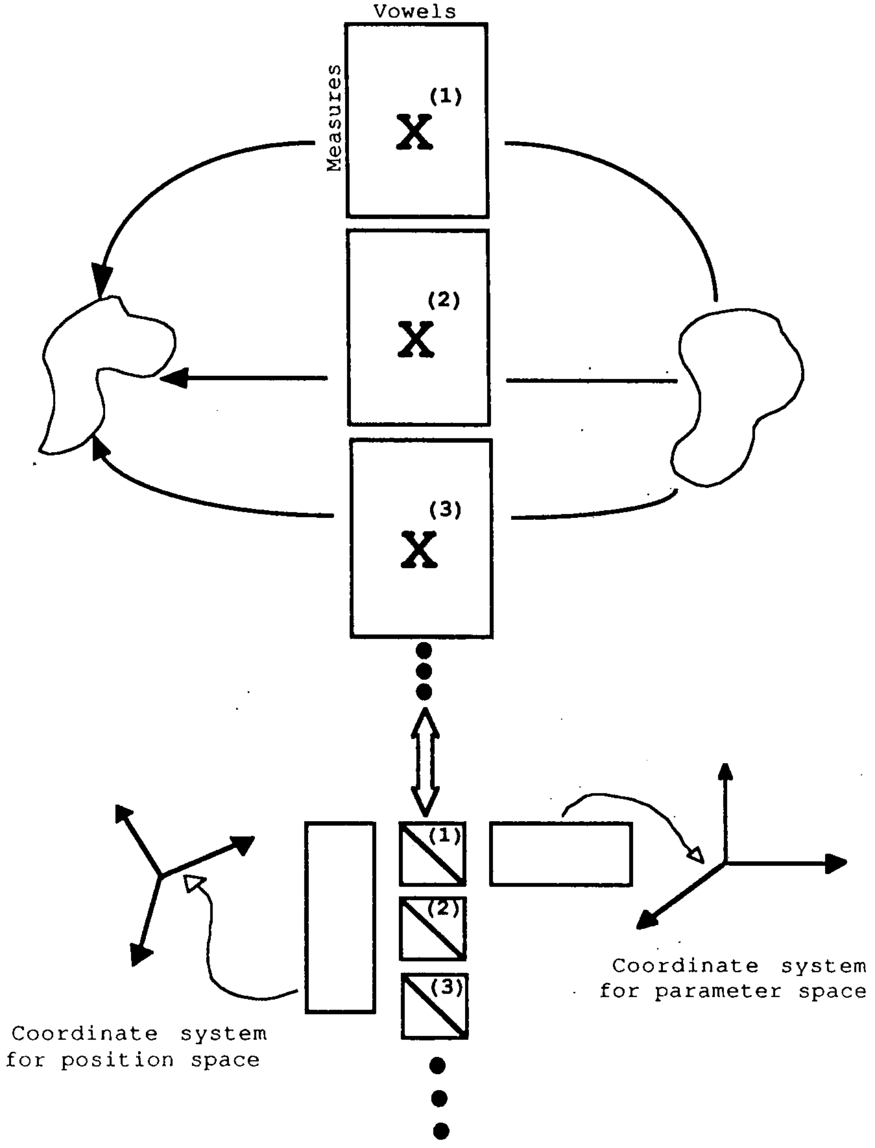
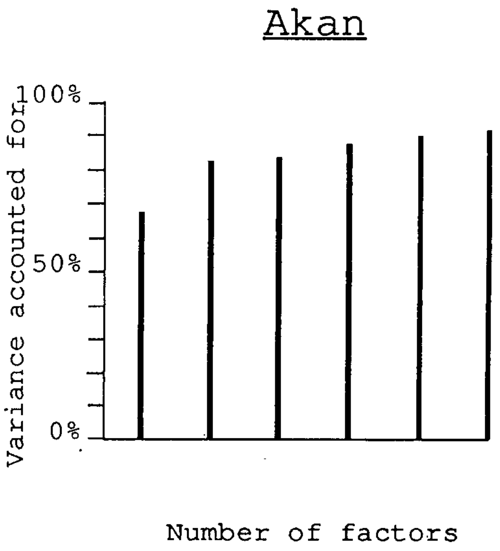


Figure 7: Design of a typical PARAFAC study.

Figure 8: VAF by PARAFAC models of the Akan data.



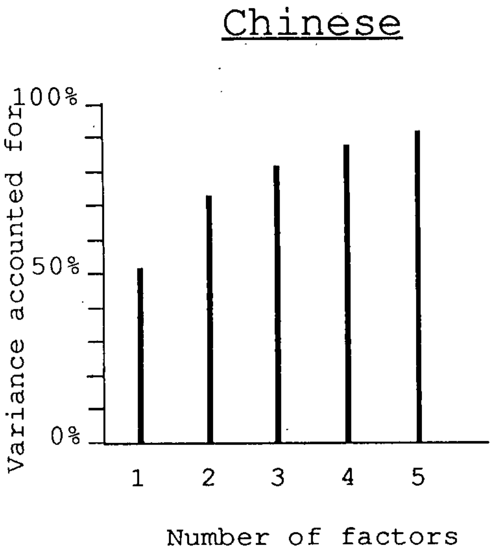


Figure 9: VAF by PARAFAC models of the Chinese data.

data. There are statistical models which allow more general mappings between the domain and range spaces of operators (e.g., the models discussed in Tucker 1966 and in Kroonenberg & de Leeuw 1980), but they do not appear to be necessary for this kind of data. Harshman, Ladefoged & Goldstein (1977) report that the two-factor PARAFAC solution accounted for over 96% of the variance in the raw data.

Like the SVD, PARAFAC in principle provides a unique representation, although data matrices that are not quite full rank can cause enormous numerical errors. Typically, one only retains factors which contribute non-vanishing variation to the modeled data. Another common practice is to retain only enough factors to account for some predetermined fraction of the variance, e.g. 75%. For instance, Figure 8 shows the percentage of the Akan data variance accounted for by PARAFAC models with various numbers of factors. Much as Lindau (1986) reported for Akan tongue shapes, two factors appear to account for most of the variance; later factors produce only minimal increases in the variance accounted for. The Chinese data, as shown in Figure 9, appears to support two or three factors; the French data (Figure 10) about 3.

Without dwelling on the specifics of the within-language models, let's consider how we can get a look at all the data simultaneously. The problem, as is schematized in Figure 11, is that there is no guarantee that we can identify the vowels of any particular language with the vowels of any other particular language. The vowels may come from different regions of the parameter space and may exploit different articulatory possibilities.

One way around this problem is to model the covariances between measurements of articulatory positions rather than the raw data. As Figure 12 suggests, each subject's variance-covariance matrix is related to the product of the raw data matrix with its transpose. Formally, this matrix can be characterized as an operator mapping the space of possible articulatory measurements into itself. We lose some information, in that the variance-covariance matrix no longer contains information about the parameter space, but on the other hand, we no longer have to cope with possible language-specific variation in that space.

Figure 12 also shows how we can get a simultaneous decomposition of these mappings from PARAFAC. In this case, PARAFAC yields a sort of simultaneous modal analysis of the data. We get a basis set of articulatory modes analogous to the eigenvectors of a single covariance matrix; we also get a set of weights analogous to

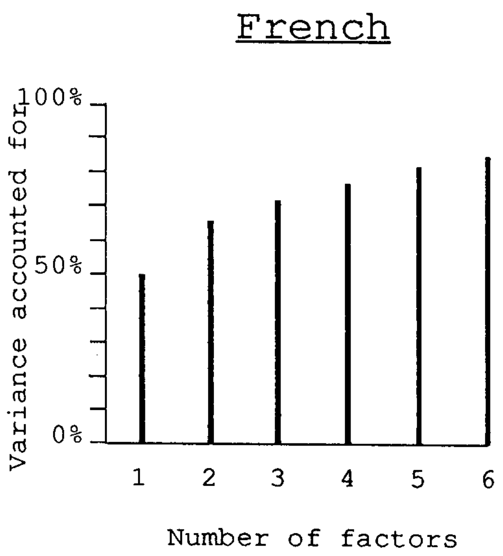


Figure 10: VAF by PARAFAC models of the French data.

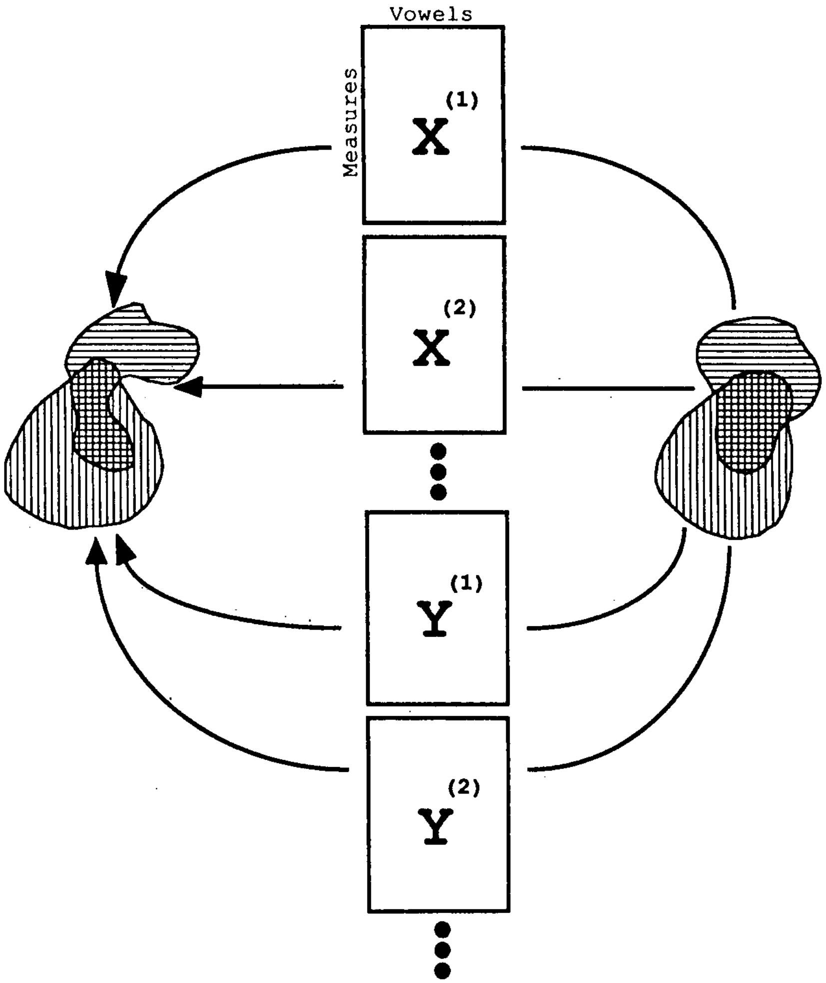


Figure 11: Vowel spaces of different languages do not necessarily overlap.

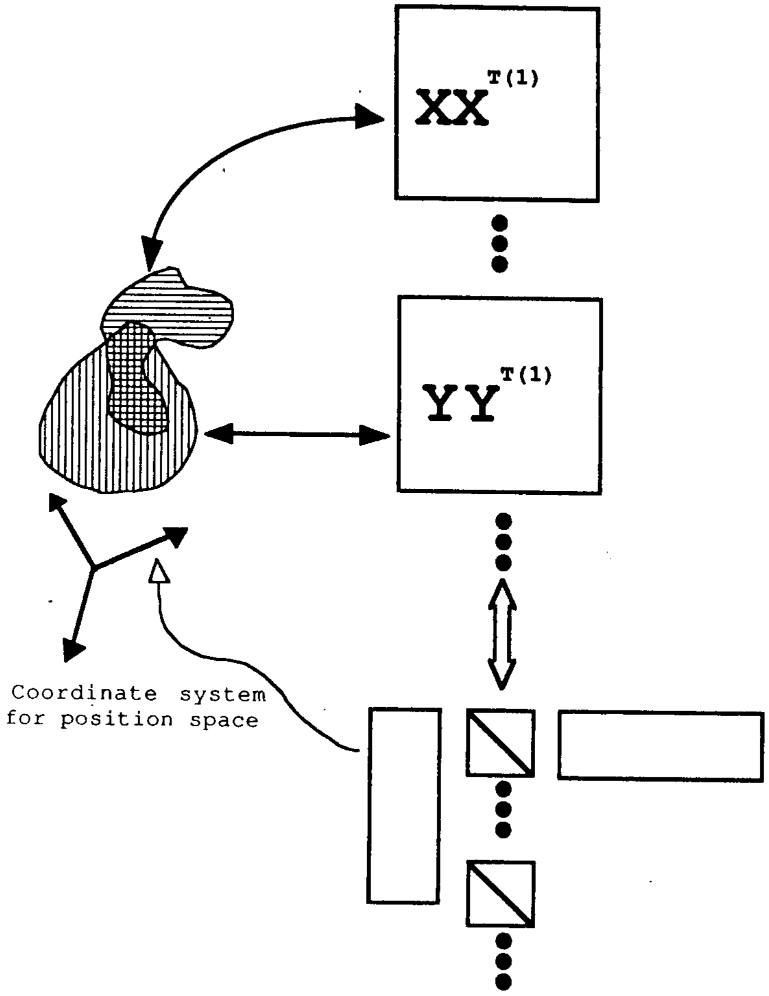


Figure 12: Modeling of covariances.

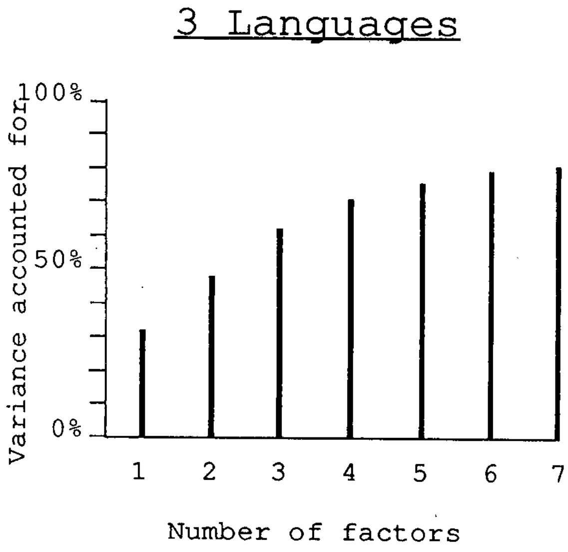


Figure 13: VAF by PARAFAC models of the covariances.

eigenvalues for each subject. Again, we normally only retain modes or factors that contribute non-vanishing variance to the modeled data.

I've applied this procedure to the data from Akan, Chinese, and French. Equal-average diagonal normalization (Harshman & Lundy 1984a) was used to remove scale differences due to differing x-ray reproduction sizes. Figure 13 shows the variance accounted for by models with one to seven factors. As we can see, it is not clear where successive factors become negligible; the data would appear to support four or five factors. I will present the five-factor solution here, since I think it has some interesting features.

The first two factors are easily described. Factor one, shown in Figure 14, generates large velum movements, together with some raising of the tongue, jaw and lips. This factor makes a relatively large contribution to the French data set, which includes the phonemically nasal vowels, and negligible contributions to the data from other languages, which don't include nasalized vowels.

Factor two, shown in Figure 15, generates large displacements at the root of the tongue and epiglottis, together with movements of the dorsal wall of the pharynx, and lowering of the jaw and lower lip as a unit. This factor contributes largely to the Akan data set; Akan is known to have a phonemic expanded/constricted pharynx contrast. Since I haven't had time to add it to the display program, you can't see that the pharynx constriction goes along with a considerable amount of larynx raising.

Factor three (Figure 16) generates a range of positions from an [o]-like posture with protruded lips to an [i]-like tongue posture with retracted and compressed lips. It appears to contribute the most to Akan, which has only front unrounded and back rounded vowels. It contributes to the articulatory space of the other languages as well, though since they also have front rounded vowels, this factor's contribution to them is proportionately not as large.

Factor four (Figure 17) also contributes to all three languages. It also generates a family of tongue postures that ranges from an [o]-like posture to an [i]-like posture. Unlike factor three, it also involves pharyngeal expansion and (not shown) larynx lowering. This appears to be a factor involved in the production of the front rounded vowels found in Chinese and French. The significance of this factor in the Akan data is surprising, though perhaps explained by the occurrence of labial consonant contexts, especially labial fricatives, in the data.

The similarities and differences between factors three and four merit some comment. The two factors appear to generate very similar ranges of tongue position. But they generate different kinds of jaw and pharynx motion. Factor three involves nearly vertical jaw motion; factor four involves much more jaw motion, including a large horizontal component that is probably part of the mechanism for producing the movements of the dorsal wall of the pharynx. These results emphasize that a front rounded vowel is not the same as a front unrounded vowel with labial activity superposed. Rather, front rounded vowels appear to involve coordinated gestures that are rather different from those required for front unrounded vowels.

Factor five (Figure 18) produces shapes ranging from an [ɛ]- or [æ]-like posture to an [u]-like posture with slightly protruded lips. There is some dilation of the pharynx and arching of the soft palate associated with the tongue backing and raising. We might expect such a factor to be involved in the production of back rounded vowels, and indeed it accounts for a largish proportion of the data in all three languages.

It is worth noting that some factors, like the first two that we saw, are relatively language-specific. It is also the case that more factors are discernable in the cross-language solution than is discernable in any of the languages individually. This suggests that none of the languages in this sample exploits all the dimensions available for the articulation of vowels. In other words, the vowels of each of these languages lies in a subspace of the full vowel parameter space. The other dimensions are not used, even in a redundant and predictable way. For instance, the difference between vowels like [e/ɛ] or [o/ɔ] is often described as tense versus lax. Some workers (e.g., Perkell 1971) have suggested that this contrast should be assimilated to the articulatory contrast between expanded and constricted pharynx vowels in e.g., Akan. But the factor which appears to contribute to the expanded/constricted pharynx distinction in Akan does not appear to contribute substantially to French vowel articulation, which has tense/lax contrasts between the vowels [e/ɛ], [o/ɔ], and [ø/œ].

In summary, the articulatory evidence is that the phonetic system as a whole provides more degrees of freedom than any individual language exploits. A kind of phonetic underspecification thus appears to be a pervasive feature of articulatory systems.

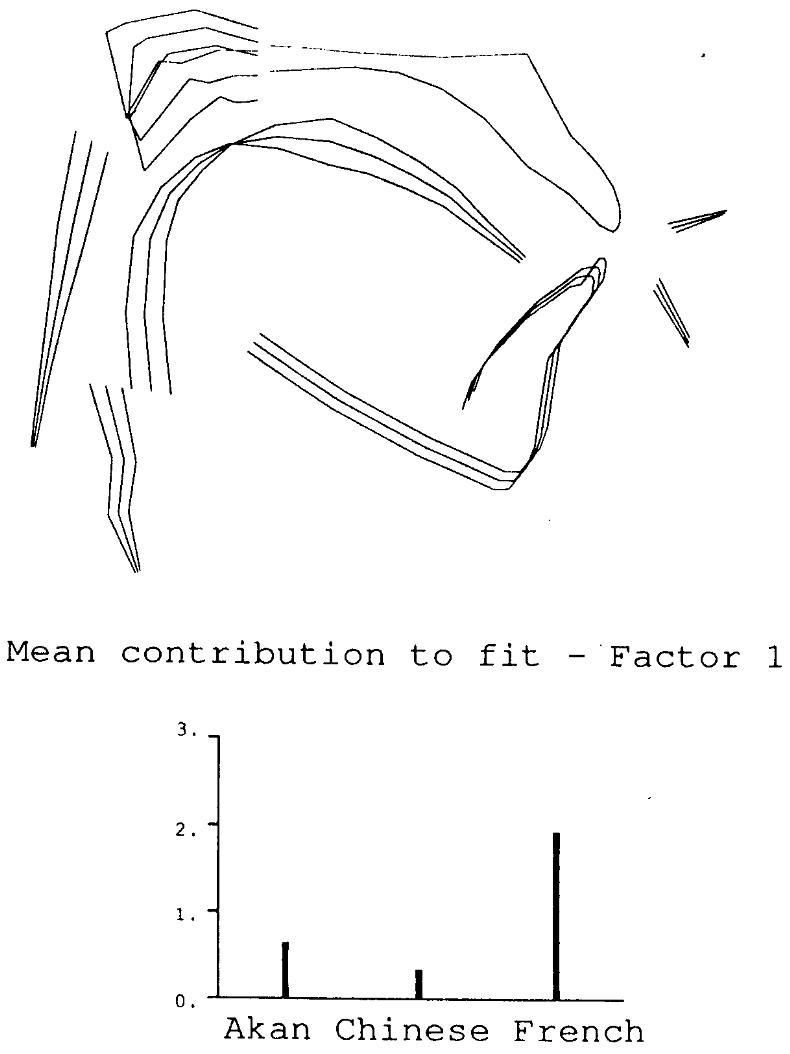


Figure 14: Factor 1.

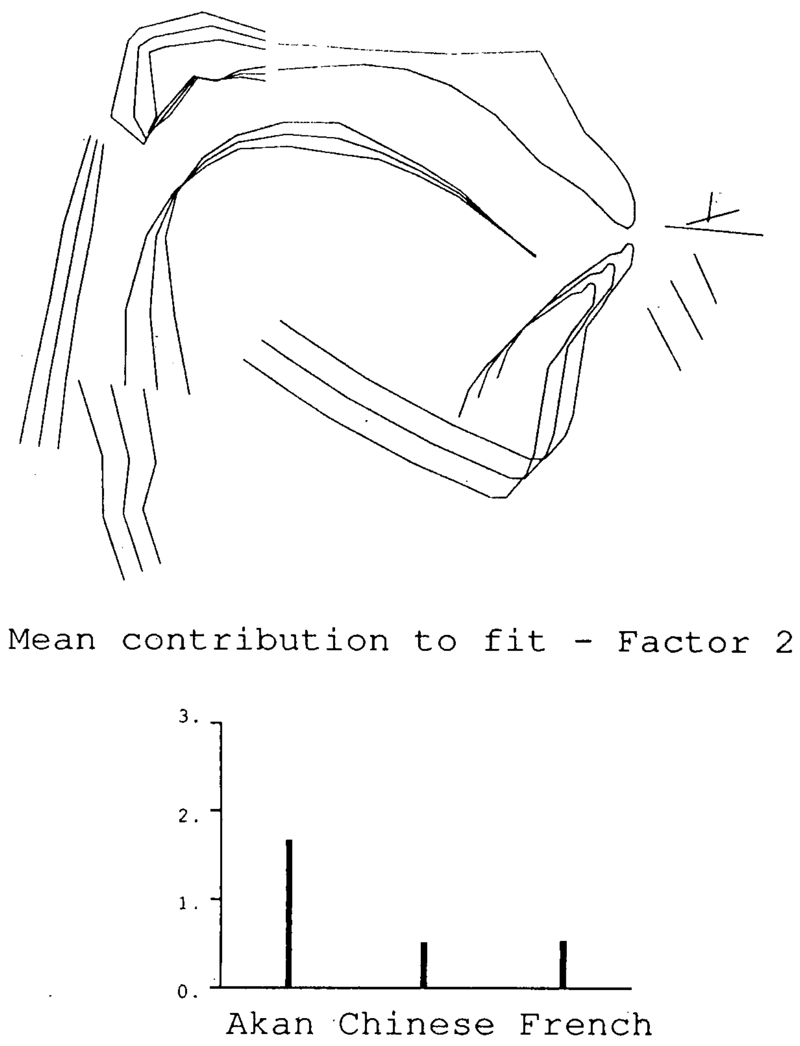
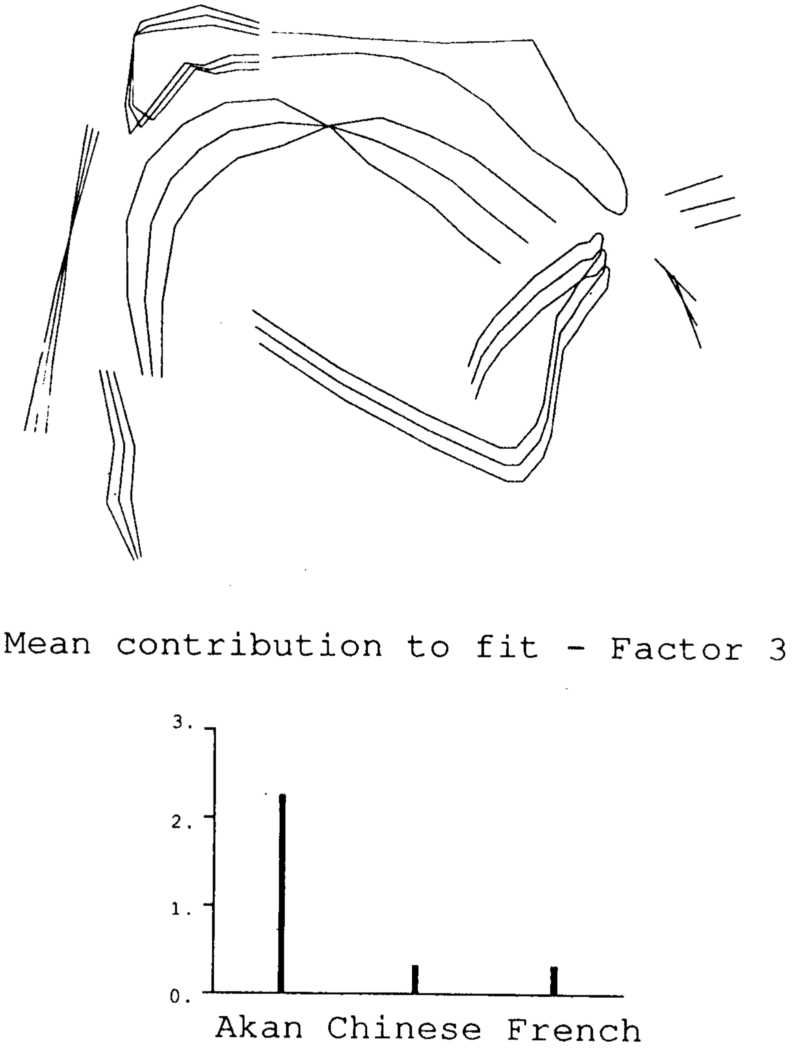
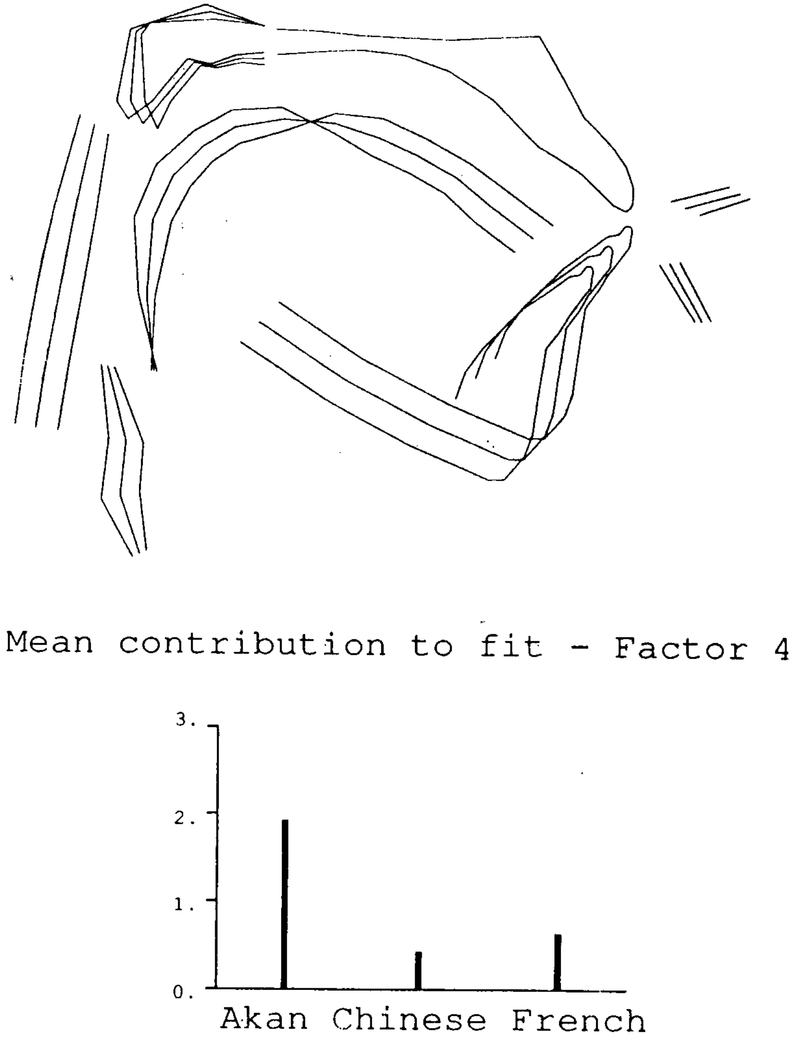


Figure 15: Factor 2.



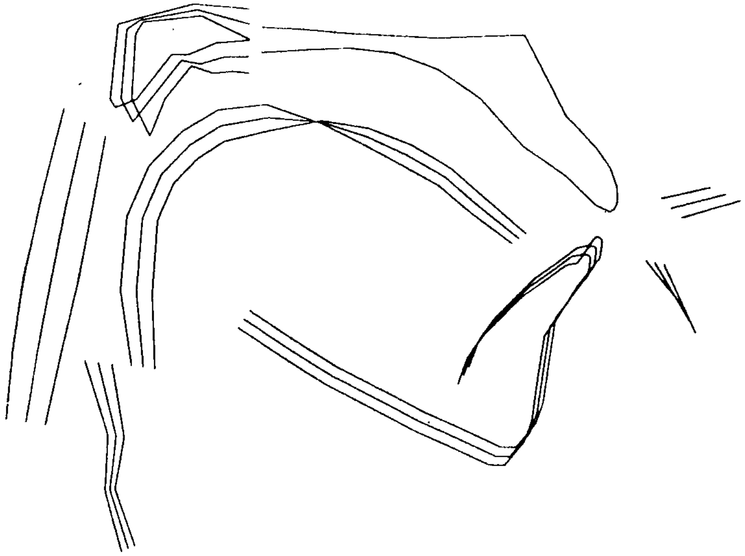
Mean contribution to fit - Factor 3

Figure 16: Factor 3.



Mean contribution to fit - Factor 4

Figure 17: Factor 4.



Mean contribution to fit - Factor 5

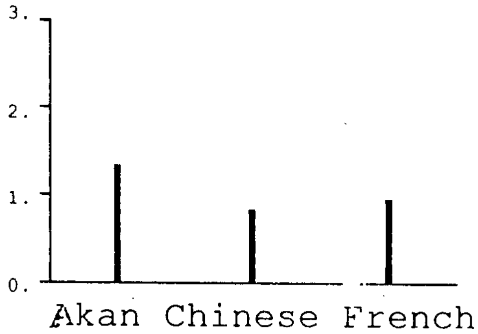


Figure 18: Factor 5.

References

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

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

Harshman, R. A., and M. E. Lundy. (1984a). "The PARAFAC model for three-way factor analysis and multidimensional scaling". In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, and R. P. McDonald (eds.), *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.

____. (1984b). "Data preprocessing and the extended PARAFAC model". In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, and R. P. McDonald (eds.), *Research methods for multimode data analysis* (pp. 216-284). New York: Praeger.

Harshman, R. A., P. N. Ladefoged, and L. Goldstein. (1977). "Factor analysis of tongue shapes". *Journal of the Acoustical Society of America* **62**, 693-707.

Kroonenberg, P. M. and J. de Leeuw. (1980). "Principal component analysis of three-mode data by means of alternating least squares algorithms". *Psychometrika* **45**, 69-97.

Kruskal, J. B. (1984). "Multilinear methods". In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, and R. P. McDonald (eds.), *Research methods for multimode data analysis* (pp. 36-62). New York: Praeger.

Law, H. G., C. W. Snyder Jr., J. A. Hattie, and R. P. McDonald. (1984). *Research methods for multimode data analysis.* New York: Praeger.

Lindau, M. (1979). "The feature expanded". *Journal of Phonetics* **7**, 163-176.

____. (1986). "Vowel features in Akan and English". Presentation EE6, 112th Meeting of the Acoustical Society of America, Anaheim CA, 10 Dec 1986. Abstracted in *Journal of the Acoustical Society of America* **80** Supp. 1, S62.

Lindblom, B., and J. Sundberg. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movement". *Journal of the Acoustical Society of America* **50**, 1166-1179.

Ohnesorg, K., and O. Svarny. (1955). "Etudes expérimentales des articulations chinoises". *Ceskosloveské Akademie Ved: Rozpravy* **65**: 5. Prague: Czech Academy of Sciences.

Perkell, J. S. (1971). "Physiology of speech production: a preliminary report on two suggested revisions of the features specifying vowels". *Quarterly Progress Report, Research Laboratory of Electronics* **102**, 123-139. Cambridge, MA: Massachusetts Institute of Techonology.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. (1986). *Numerical recipes: The art of scientific computing.* New York: Cambridge University Press.

Tucker, L. R. (1966). "Some mathematical notes on three-mode factor analysis". *Psychometrika* **31**, 279-311.

Zhou, D. and Z. Wu. (1963). *Articulation album of Putonghua.* Beijing: Shangwu Yingshuguan.

Are there cross-linguistic differences in front vowel production?*

Michel T. T. Jackson

Introduction

It is often assumed that the characteristics of a language's phonological inventory play some role in determining the phonetic substance of that language's segments. For instance, it is often believed that languages with more vowels tend to use more of the phonetic vowel space in order to maintain phonological contrasts. A high front vowel in a language with many vowels is expected, on this view, to be higher and fronted that a high front vowel in a language with only three vowels, in order to maintain the perceptual distinctness of the vowel.

The purpose of this article is to examine a similar claim that has been made about front rounded vowels in particular. In an often-cited study, Sidney Wood (1979) has claimed that languages with front rounded vowels use a prepalatal place of constriction, in contrast with languages without front rounded vowels, which generally use a midpalatal place of constriction. The prepalatal place of articulation is claimed to give a sharper acoustic quality to the unrounded front vowel, i.e. a higher F_3, thus increasing its perceptual separation from the flat unrounded front vowel. The midpalatal place of articulation apparently does not give a sharp enough quality to the vowel for the contrast between rounded and unrounded front vowels to be reliably perceived.

There is, however, a certain amount of evidence in the literature which suggests that this claim is ill-founded. As a brief example, Fant reports (1973, Chp. 4) formant frequencies for Swedish vowels, and compares these values to the values from the most similar American English (AE) vowels as reported by Peterson & Barney (1952). In most cases (Table I), the English front vowels have a higher F_3 than the comparable Swedish vowels. The Swedish vowels are thus flatter, contradicting Wood's claim.

---

*Expanded version of a paper presented at the Linguistic Society of America Annual Meeting, San Francisco CA, December 1987.

Arabic tongue shapes

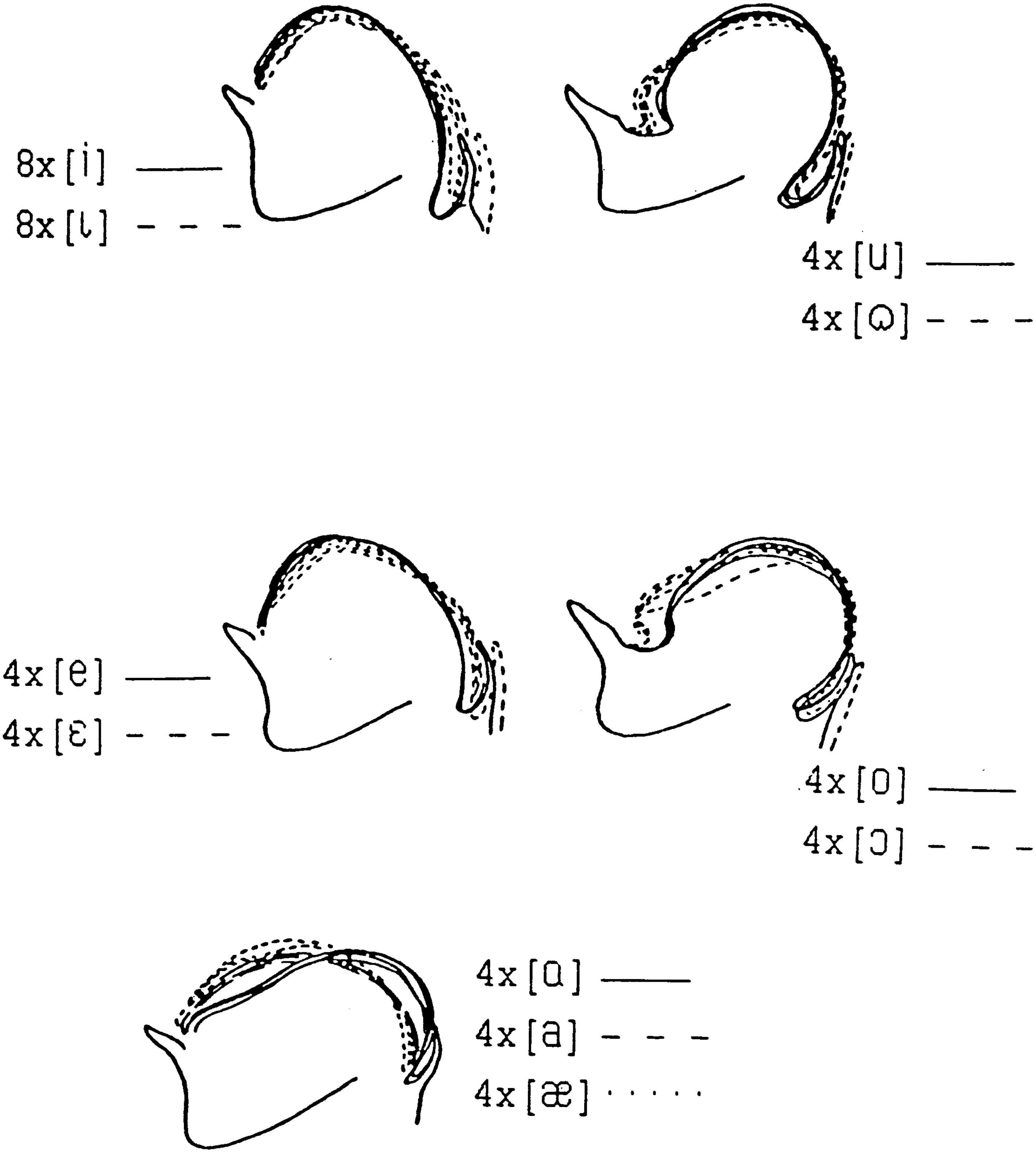


Figure 1

English tongue shapes

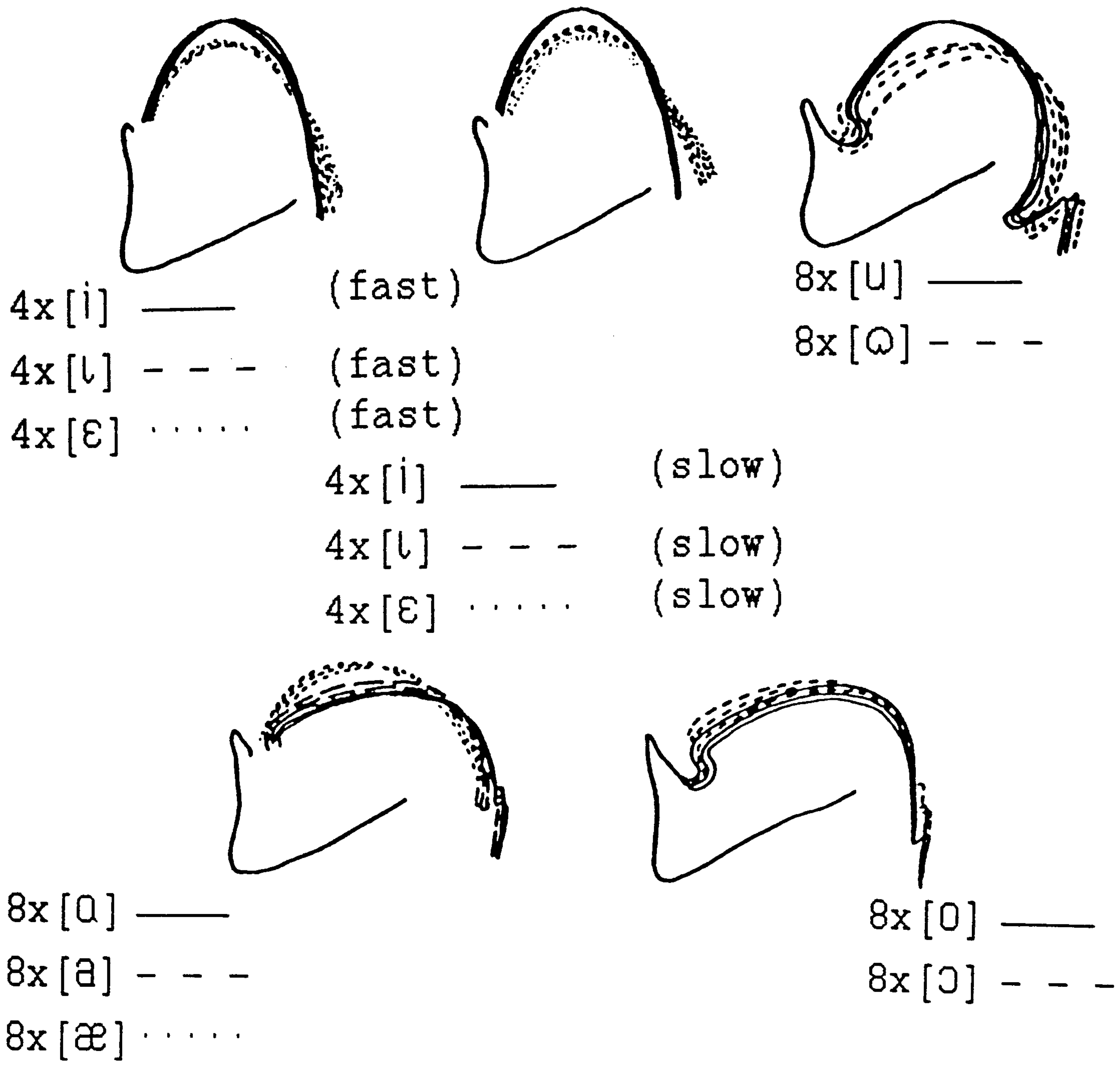


Figure 2

Table I: Formant frequencies of selected American English and Swedish vowels (after Fant 1973, Peterson & Barney 1952)

English				Swedish			
V	F_1	F_2	F_3	V	F_1	F_2	F_3
i	270	2290	3010	i_1	256	2066	2960
ι	390	1990	2550	e_1	334	2050	2510
ε	530	1840	2480	$\ddot{a}_1$	438	1795	2385
æ	660	1720	2410	$\ddot{a}_3$	606	1550	2450

However, this is only an indirect test of the hypothesis, which is an *articulatory* one, not an acoustic one. It is always possible that there is some difference between the speaker samples, e.g. age or speaking rate and style, that could account for the acoustic differences. We will therefore test Wood's claim on articulatory data.

Wood himself presents articulatory data in favour of his hypothesis. Some of the more convincing is the comparision between one speaker of Cairo Arabic (CA) and one speaker of Southern British English (SBE; Wood 1979). Fig. 1 shows families of tongue shapes from the vowels of CA as produced by the speaker. The tongues are superposed with reference to the lower jaw. Fig. 2 shows families of tongue shapes from the vowels of SBE as produced by the one speaker. Comparing the families of tongue shapes of the high front vowels from the two languages, we can see that there are some differences between the two sets of tongue shapes. The greatest variation in the family of tongue shapes used in high front vowels by the SBE speaker is more or less at the center of the tongue, as though the tongue were arching upwards. The CA speaker, on the other hand, appears to both arch the tongue and raise the blade of the tongue. Wood interpretes these observations as showing that CA has a prepalatal place of constriction for its front vowels, but that SBE has a midpalatal one.

However, there are several problems with these data. An important one is that Wood gives no indication of the palate position during these vowels, and so it is impossible to tell where the maximal constriction really is in these vowels. Also, since CA does not have front rounded vowels, Wood weakens his claim to say that languages with front rounded vowels strongly prefer the prepalatal place of articulation, whereas languages without front rounded vowels have a choice between the prepalatal and midpalatal places.

In light of these problems, it is worthwhile to give Wood's hypothesis a thorough evaluation based on articulatory data.

Using the UCLA X-Ray Database (Dart 1987) and other sources, I have collected tracings of midsagittal x-rays of vowels from six languages, with several speakers of each language. X-ray tracings of four speakers of Akan (Lindau 1979, personal communication); three speakers of Chinese (Ohnesorg & Svarny 1955); four speakers of French (Bothorel, Simon, Wioland & Zerling 1986); Icelandic (Pétursson 1974a, 1974b); five speakers of Spanish (Holbrook & Carmody 1937, Navarro-Tomás 1916, Parmenter & Treviño 1932, Quilis 1981, Russell 1929-30); and four speakers of Swedish (Fant 1969, Sundberg 1969, personal communication) were obtained. For all of these speakers, most of the vowels of the language were x-rayed and traced. The measurements of tongue position in five speakers' AE vowels reported in Harshman, Ladefoged, and Goldstein (1977) were also used.

Table II summarizes the data collected. The languages at the top of the table, Akan, English, and Spanish, do not have front rounded vowels, and the languages at the bottom, Chinese, French, Icelandic, and Swedish, do. The data includes 256 vowels produced by 27 speakers.

Table II: Data summary

Language	Speakers	Vowels	Missing	Total
Akan	4	9	2	34
American English	5	10	0	50
Spanish	5	7	4	31
Chinese	3	7	4	17
French	4	14	0	56
Icelandic	2	16	0	32
Swedish	4	10	4	36
				256

The tongue positions in all of these vowels were measured using grids similar to those used in Harshman *et al.*'s (1977) study of AE vowels. Fig. 3 shows a couple of sample grids; the bottom one is from a female speaker of Icelandic, the top from a male speaker of French. Here, we will only be concerned with the gridlines labelled 'Root', 'Blade', and 'Tip', as none of the measurements along the other gridlines yields significant information.

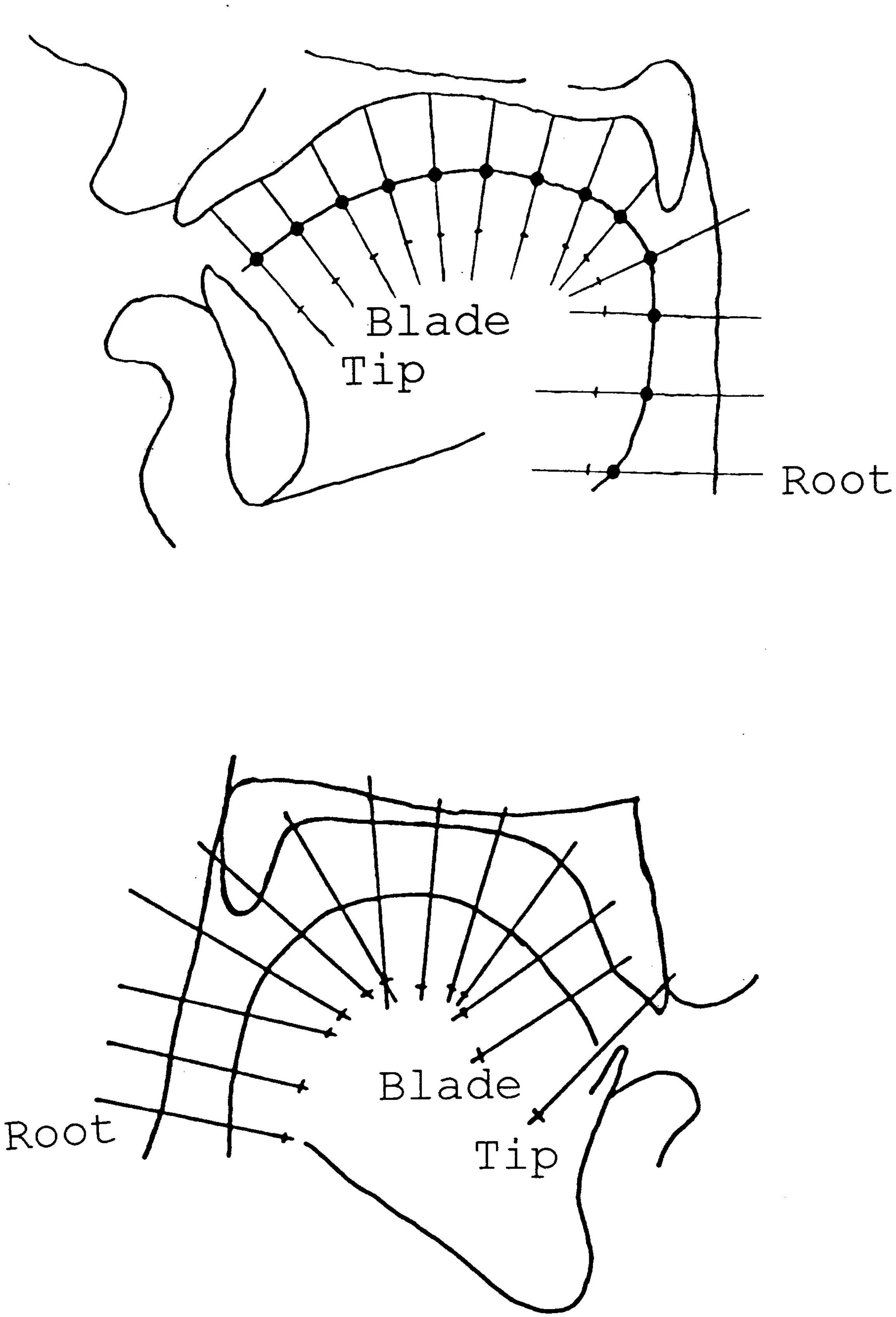


Figure 3: Sample measurement grids.

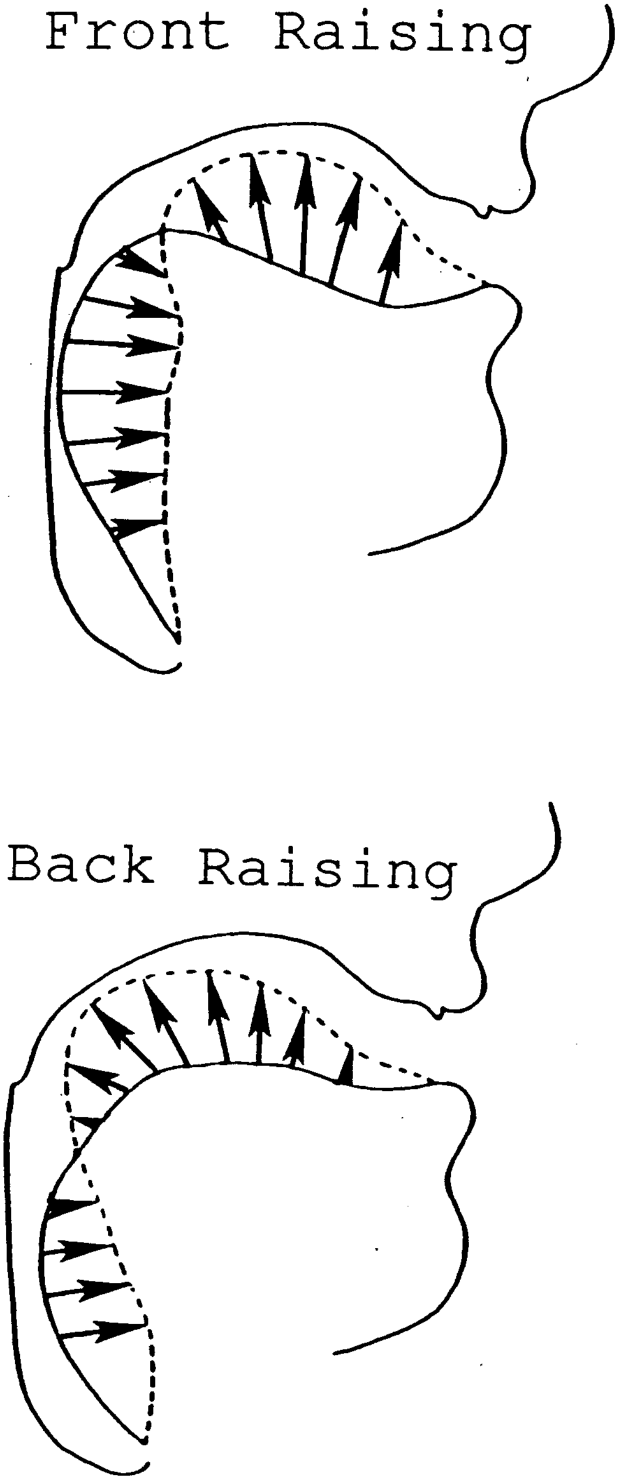


Figure 4: Harshman et al.'s model.

Harshman *et al.* (1977) proposed an articulatory model for the tongue positions observed in their data from AE. As Fig. 4 shows, this model consists of two factors, each of which generates a family of tongue shapes. Each observed vowel's vocal tract shape is generated as a weighted combination of the two factors. The factor labelled 'Front Raising', in partucular, contributes greatly to the tongue shapes of front high vowels.

This model allows us to explicitly test Wood's claim. Wood's claim is that the front vowels of languages with front rounded vowels have a different family of tongue shapes from those of languages without front rounded vowels. One possible different shape is suggested in Fig. 5, where a tongue shape that would create a different constriction location is sketched in. If the tongue shapes in languages with front rounded vowels are in fact different, then the articulatory factors from the AE data should not fit the tongue shapes in those languages as well as they fit tongue shapes in languages that do not have front rounded vowels.

Furthermore, as Fig. 5 suggests, when the tongue shapes from these languages are fit to the best comparable tongue shape generated by the model, there should be a systematic pattern of differences. We'll use several methods for finding systematic patterns. First, these systematic differences can be detected by applying analysis of variance (ANOVA) to the residuals of the fit, which reflect the deviations of the given tongue shapes from the tongue shapes generated by the model. Second, we can detect systematic differences by testing the hypothesis that the mean values of the deviations are significantly different from zero. This is because random deviations arising from measurement error will tend to have a mean of zero, whereas deviations arising from systematic differences should systematically tend to one direction or another and thus have a non-zero mean.

Finally, we can factor-analyze the residuals in the same way that the original AE tongue measurement data was. Since it is usually possible to fit some factor to this kind of articulatory data, due to the presence of correlated errors in the measurements, a randomization test is necessary to determine whether or not the factor in the residual data accounts for significantly more variation than it would account for if it were just fitting correlated errors.

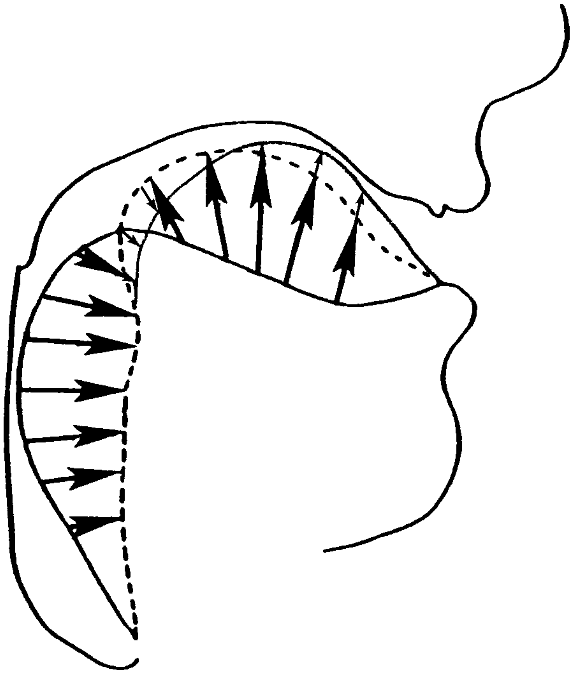


Figure 5: A possible tongue shape creating a prepalatal constric
Small arrows indicate the kind of residual that would be expecte er fitting
Harshman et al.'s factors to tongue shapes from a language wit epalatal
constrictions.

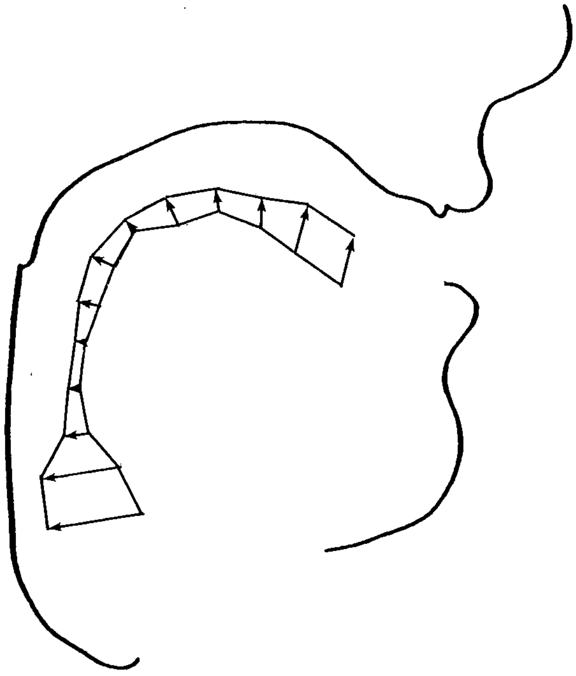


Figure 6: Residual factor found in Akan.

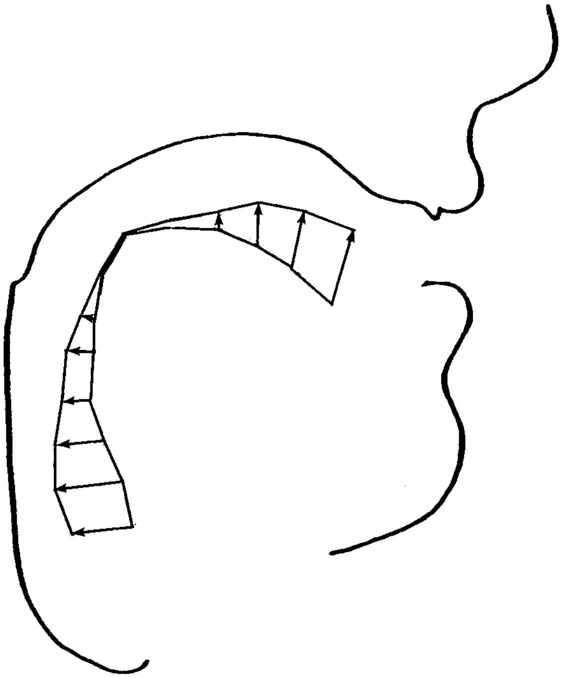


Figure 7: Residual factor found in French.

Before setting out the statistical results, let's consider the specific predictions of Wood's hypothesis in this context. First, the value for frontness of a vowel, *by itself*, should have no predictive power with respect to the residuals (the deviations of the actual articulatory measurements from the factors based on AE vowels). A front vowel could come from either a language with, or without, front rounded vowels. Similarly, whether or not the vowel comes from a language with front rounded vowels *by itself* should have no predictive power, since there is no reason to belive that the non-front vowels pattern differently in the two kinds of languages. Only the *combination* of frontness and coming from a language with front rounded vowels should predict a systematic pattern of deviations.

Preliminary fits

We now turn to the results. Table III shows the overall goodness of fit, on a per-language basis, of the AE-based model to the tongue shape measurements from the other languages. The fit values were calculated by the PARAFAC program (Harshman & Lundy 1985). The languages without front rounded vowels are shown in boldface. They fall in the middle of the pack, so this measure certainly does not separate the two kinds of languages.

Table III: Overall goodness-of-fit of the AE-based model to articulatory data from other languages

Language	Fit value (R^2)
Icelandic	.8902
Akan	**.8559**
Swedish	.8426
Chinese	.8340
Spanish	**.8313**
French	.7924

Analyses of variance

The next step is to analyze the deviations of the measured tongue shapes from those generated by the model to see if there are systematic differences between the tongue shapes found in the data set and the tongue shapes generated by the model. The following analyses of the residuals were performed using the SAS GLM procedure (SAS Institute Inc., 1985). The GLM (General Linear Models) procedure provides both ANOVA and regression models.

The ANOVA of the deviations (summarized in Table IV) reveals that the value for the frontness of a vowel *does* make a difference in how well a vowel fits into the families of tongue shapes use in AE. In other words, there is a probably significant main effect of the frontness of the vowel on its deviation from the tongue shapes of AE vowels. However, contrary to the predictions discussed above, this effect is not further affected by whether or not the vowel comes from a language with front rounded vowels. The combined (or interaction effect) effect of frontness and vowel inventory (having or not having front rounded vowels) is only possibly significant at the tongue root.

Table IV: ANOVA of the deviations from tongue-positions generated by the AE-based model

Effect	Tongue root	blade	tip
Frontness	*		**
Inventory			
Frontness*Inventory	*		

*: $p < 5\%$, possibly significant
**: $p < 1\%$, probably significant

Analysis of the mean values of the deviations confirms these observations (Table V). The GLM procedure provides a least-squares estimate of the mean residual deviations within each category of vowels. This estimated mean, together with the standard error of the estimate, can be used to calculate the probability of the mean being different from zero like a two-tailed *t*-test.

By this test, both front and non-front vowels in this data sample have tongue shapes that deviate significantly from the tongue shapes generable by the model. In particular, both the front and non-front vowels from languages that do not have front rounded vowels deviate. These results are not consistent with Wood's strongest hypothesis, since apparently the languages that have front rounded vowels are using tongue shapes pretty much like those used in AE.

Table V: Mean deviations by frontness and inventory

Category	root	Tongue blade	tip
+front	*		**
-front			**
+front rounded Vs in inventory			
-front rounded Vs in inventory			
+front, +front rounded Vs in inventory			
+front, -front rounded Vs in inventory	*		**
-front, +front rounded Vs in inventory			
-front, -front rounded Vs in inventory			**

*: p < 5%, possibly significantly different from zero
**: p < 1 %, probably significantly different from zero

A closer look at the data seems to be in order. The results suggest that there is substantial cross-language variation in front vowels, although this variation is not a function of whether or not the language has front rounded vowels in its inventory. Perhaps it is simply a matter of choice for any particular language. If this is the case, then the identity of the language, rather than just the property of having front rounded vowels, in combination with the value for frontness, should be able to predict whether or not the tongue shape of the vowel fits into one of the families of tongue shapes that are used in AE.

ANOVA of the deviations, as categorized by frontness and language, again shows a significant main effect of frontness (Table VI). Language by itself is not a significant predictor of how well a vowel's tongue shape is fit by the AE based model - it does not have a significant main effect. However, the interaction effect of frontness and language is a strongly significant predictor of whether or not a tongue shape deviates from those generated by the model.

Table VI: ANOVA of the deviations from tongue-positions generated by the AE-based model

Effect	root	Tongue blade	tip
Frontness	*		**
Language			
Frontness*Language	**	*	**

*: p < 5%, possibly significant
**: p < 1 %, probably significant

A means test parallel to the one summarized in Table V is presented in Table VII. Recall that a mean deviation significantly different from zero in a particular set of vowels indicates a systematic difference between the tongue shapes of those vowels and the AE vowels used to construct the original model.

Table VII: Mean deviations by frontness and language

		Tongue	
Category	root	blade	tip
+front			
-front			*
Akan			
Chinese			
French			
Icelandic			
Spanish			
Swedish			
+front, Akan	**	**	**
-front, Akan			**
+front, Chinese			
-front, Chinese			
+front, French			
-front, French			
+front, Icelandic			*
-front, Icelandic			**
+front, Spanish			
-front, Spanish	*		*
+front, Swedish			
-front, Swedish			

*: p < 5%, possibly significantly different from zero
**: p < 1 %, probably significantly different from zero

Among the languages without front rounded vowels, we can see that Akan front vowel tongue shapes are significantly different from AE tongue shapes at both the root and the tip of the tongue. Akan non-front vowels are also significantly different.

Among the languages that *do* have front rounded vowels, only Icelandic has tongue shapes that are significantly different from the families of tongue shapes generable by the model. In particular, Icelandic non-front vowels show significant systematic deviation from the tongue shapes generated by the model. But neither Chinese nor French nor Swedish shows significant differences from AE.

Randomization tests

A final way of detecting systematic differences between the measurements of tongue shapes in the six languages and AE is to factor-analyze the deviations that result from fitting the AE-derived model to the articulatory measurements from the other languages. The factor resulting from such an analysis represents an independent mode of tongue displacement similar to the two tongue displacement factors of Harshman *et al.* (1977).

The residual factor is orthogonal to the factors in the original model (Harshman 1984, pp. 579-580; i.e., the dot products of the residual factor's contributions to the vowels with Front-Raising's and Back-Raising's contributions to the same vowels are zero). This orthogonalization is due to the fact (familiar from regression analysis and common to all least-squares models) that the residuals themselves are orthogonal to the part of the data fitted by the model. Orthogonalization of a factor in PARAFAC models may result in some distortion of the factor (Harshman & Lundy 1984, p. 275-276), and therefore, some care in interpreting the residual factors obtained below is necessary.

To guard against the possibility that the residual factor is simply capitalizing on correlated errors due to errors in tracing the x-rays or errors in positioning the measurment grid, we must use a randomization test of the sort discussed in Harshman (1984). We can test for remaining significant systematic variation in among the tongue position measures in different vowels by randomly permuting the vowels produced by each speaker in the input to the PARAFAC program. This permutation retains speaker-specific correlated errors (such as those that might be produced by applying a poor measurement grid to all the vowels produced by a specific speaker) and errors that are correlated across measurement gridlines (such as those produced by misplacement of the measurement grid in a specific vowel), while destroying systematic correlations among the vowels.

If the data are permuted and factor-analyzed 20 times, then under the null hypothesis that there is no systematic residual variation in the data (after fitting to the AE-based mode), these 20 random permutations should not differ systematically from the original unrandomized data. Thus, the null hypothesis of no systematic variation can be rejected at the 5% level if the original unpermuted data produces a higher fit value than the 20 permuted ones, since this ranking of the fit values has a probability of 5% under the null hypothesis.

20 pseudorandom permutations of the data from Akan, Chinese, French, Icelandic, Spanish, and Swedish were generated using a random number generator with very low sequential correlations (Press, Flannery, Teukolsky, & Vetterling 1986, pp. 196-198). For example, the seventh permutation of the French data ordered the vowels [ɔ̃, œ̃, u, e, a, ɛ̃, ɛ, o, ɑ̃, ɔ, y, ø, i, œ] for speaker one, [u, i, o, e, ɛ, œ, ɑ̃, ɔ, ɛ̃, œ̃, y, ɔ̃, ø, a] for speaker two, [ɔ, ɔ̃, i, a, u, ɑ̃, e, œ, œ̃, ɛ̃, y, ø, ɛ, o] for speaker three, and [ɛ, o, a, e, ɑ̃, ø, u, œ, œ̃, y, ɔ̃, ɛ̃, i, ɔ] for speaker four. A PARAFAC model was fit to each randomization of the residuals as well as to the unrandomized residuals.

The results are summarized in Table VIII. The table contains stem-and-leaf plots of the distributions of fit values (R^2) for each language. These plots summarize both the distribution and the actual of R^2 values of the PARAFAC model of the residuals in each language. The plot is meant to be read as follows: digits left of the bar (•) are read sequentially, digits right of the bar are read as alternates. Thus, ".3•65" is meant to summarize two values, 0.36, and 0.35. The stem-and-leaf plot is thus like a histogram, since the frequency within a bin corresponds to the length of the row to the right of the bar, and also gives actual values.

The R^2s of the PARAFAC fit to the unrandomized residuals from each language is given in Table VIII in parentheses at an approximately appropriate place on the vertical axis of the corresponding stem-and-leaf plot.

Table VIII: Fit values for factor-analysis of randomized and unrandomized residuals

```
    Akan
(.526)•
     1•  9
   .40•  37
     9•  4
     8•
     7•  0
     6•  00
     5•  4805
     4•  2
     3•  36
     2•
     1•  48589
   .30•  6
```

```
  French
(.371)•
      •
     9•
     8•  1
     7•
     6•  739
     5•
     4•  38
     3•  365
     2•  827
     1•  78384
   .20•  96
     9•
     8•
   .17•  8
```

```
    Chinese



         9•
(.289)8•  3
         7•  23
         6•  00
         5•  8944
         4•  62
         3•  4477
         2•  13
         1•  93
       .20•  4
```

```
  Icelandic
          1•  0
        .30•  1
   (.295)9•  515
        .28•  400820048000101
```

Spanish

4•	8
3•	
2•	0
1•	5
.30•	6058014
9•	23
8•	174
(.276)7•	852
6•	
5•	
4•	
3•	
2•	
1•	
.20•	6
9•	
.18•	8

Swedish

.40•	7
9•	
8•	
7•	
(.363)6•	7
5•	
4•	6
3•	7
2•	3
1•	14
.30•	20
9•	
8•	844
7•	3702
6•	46
.25•	97

Table VIII shows that Akan, French, and possibly Chinese show systematic variance not captured in the AE-based model. This multivariate randomization test's results are thus somewhat different from the results of the univariate ANOVA's reported above. However, both the ANOVA's and the randomization test find Akan to be more different from AE than the other languages.

Examination of the patterns of articulatory displacements produced by the factors from the residuals in Akan, Chinese, and French suggested that the Chinese factor was very similar to the Back-Raising factor of Harshman *et al.* (1977). This was confirmed by calculating the coefficient of correlation *r* between the displacements along each gridline produced by the Chinese residual factor and the displacements along each gridlines produced by the AE Back-Raising factor. The correlation was over 0.99, suggesting that the apparent Chinese residual factor has the same pattern of articulatory displacements as the AE Back-Raising factor. By comparision, neither the Akan residual factor nor the French residual factor had a correlation of over 0.55 with either Front-Raising or Back-Raising.

The high degree of correlation between the Chinese residual factor and the Back-Raising factor suggests that it does not capture very much systematic variance, though it may account for some interspeaker variation. Examination of the contribution of the factor to the residuals showed that the factor apparently accounted mostly for speaker-specific differences in the height of the vowel [ə].

The articulatory displacements produced by the residual factors in Akan and French are plotted in Figs. 6 and 7. It can be seen that these two factors are fairly similar (*r* of about 0.71) but also that they are probably fairly noisy. There are zigzags in the profile of tongue displacement that we would not expect to see, since the tongue surface is normally a smooth curve.

However, in both languages, the factor represents a simultaneous tip-raising and lower-pharyngeal contraction gesture. When the tip is lowered and the lower pharynx is expanded, the tongue shape is rather bunched; when the tongue tip is raised and the lower pharynx is constricted, the tongue takes on a flatter overall shape. To the extent that this factor is undistorted by the orthogonalization referred to above, it generates a family of tongue shapes that ranges from those claimed to be characteristic of tense vowels (at the bunched extreme) to those claimed to be characteristic of lax vowels (at the flattened extreme).

In Akan, this factor contributes to the expanded-contracted pharynx distinction, with the expanded pharynx vowels [i, e, o, u] having more negative contributions to their overall tongue shapes that the contracted pharynx vowels [ɩ, ɛ, a, ɔ, ɷ]. In French, this factor appears to be related to height differences, with vowels ordered roughly as [u/o, i/e/y, ø/œ/ɔ/a/ɛ̃/œ̃, ɛ/ɔ̃/ɑ̃]. The high vowels thus tend to have expanded pharynxes, and low vowels tend to have constricted pharynxes (relative to the AE-based model).

This Tongue-Flattening factor, derived from the portion of the articulatory data that the AE-based model could not fit, appears to represent the articulatory correlate of a phonological contrast that does not occur in AE. It's presence in Akan and French is not attributable to the inventory effect hypothesized above, since the presence of Tongue-Flattening is not related to the presence of front rounded vowels in the language. Rather, the presence of the Tongue-Flattening factor in Akan and French simply reflects a phonological difference between these languages and AE.

Conclusion

The majority of languages with front rounded vowels show no difference from AE. Yet there is clearly quite a bit of cross-language variation in tongue positions in front vowels, most of it is due to the Akan vowels. In retrospect, this is no surprise, since we know that Akan has a phonemic expanded-contracted pharynx contrast that does not exist in other languages, nor in AE from which the model of tongue shapes is derived. As we should thus expect, the tongue shapes of Akan vowels at both the root and the tip of the tongue do not fall into the same families as the tongue shapes AE vowels. The fact that the ANOVAs used here are powerful enough to detect that Akan has tongue shapes that AE doesn't have suggests that we are not failing to detect other significant articulatory differences.

We can conclude that the strong hypothesis advanced by Wood (1979) receives no confirmation in this data, and indeed is contradicted by it. In these seven languages, there is no evidence that front vowels in languages with front rounded vowels are produced with tongue shapes and constriction locations systematically different from the ones used by languages without front rounded vowels. Rather it seems as though the same underlying articulatory organization can give rise to different inventories of vowels.

References

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

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

Fant, G. 1969. "Formants and cavities", in E. Zwirner & W. Bethge (eds.), *Proceedings of the 5th International Congress of Phonetic Sciences*, 120-141. Basel: S. Karger.

____. 1973. *Speech Sounds and Features.* MIT Press Current Studies in Lingustics **4**. Cambridge MA: MIT Press.

Harshman, R. A. 1984. "'How can I know if it's "real"?' A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling", in H. G. Law, C. W. Snyder, Jr., J. A. Hattie, and R. P. McDonald (eds.), *Research Methods for Multimode Data Analysis*, 566-591. New York: Praeger.

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

Harshman, R. A., & M. Lundy. 1984. "Data preprocessing and the extended PARAFAC model", in H. G. Law, C. W. Snyder, Jr., J. A. Hattie, and R. P. McDonald (eds.), *Research Methods for Multimode Data Analysis*, 216-284. New York: Praeger.

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

Lindau, M. 1979. "The feature expanded", *J. Phonet.* **7**: 163-176.

Lundy, M. E. & R. A. Harshman. 1985. *Reference Manual for the PARAFAC Analysis Package.* London, Ontario: Scientific Software Associates.

Navarro-Tomás, T. 1916. "Siete vocales españolas", *Revista de Filología Española* **3**: 51-62.

Ohnesorg, K. & O. Svarny. 1955. *Études Expérimentales des Articulations Chinoises.* Ceskosloveské Akademie Ved: Rozpravy **65**: 5. Prague: Czech Academy.

Parmenter, C. & S. N. Treviño. 1932. "An x-ray study of Spanish vowels", *Hispania* **15**: 483-496.

Peterson, G. E. & H. L. Barney. 1952. "Control methods used in a study of vowels", *J. Acoust. Soc. Amer.* **24**: 175-184.

Pétursson, M. 1974a. "Peut-on interpréter les donées de la radiocinématographie en fonction du tube acoustique à section uniforme?", *Phonetica* **29**: 22-79.

____. 1974b. *Les Articulations de l'Islandais à la Lumière de la Radiocinématographie.* Société de Linguistique de Paris **68**. Paris: Klincksieck.

Press, W. H., B. P. Flannery, S. A. Teukolsky, & W. T. Vetterling. 1986. *Numerical Recipes: The Art of Scientific Computing.* New York: Cambridge University Press.

Quilis, A. 1981. *Fonética Acústica de la Lengua Española.* Madrid: Biblioteca Románica Hispánica.

Russel, G. O. 1929-30. "The mechanism of speech", *J. Acoust. Soc. Amer.* **1**: 83-109.

SAS Institute Inc. 1985. *SAS/STAT™ Guide for Personal Computers, Version 6 Edition.* Cary NC: SAS Institute Inc.

Sundberg, J. 1969. "Articulatory differences between spoken and sung vowels in singers", *Stockholm Royal Institute of Techonology, Speech Technology Lab Quarterly Progress and Status Report* **1**: 33-46.

Wood, S. 1979. "A radiographic analysis of constriction location for vowels", *J. Phonet.* **7**: 25-43.

The flap as a contour segment

Susan Banner Inouye

I. INTRODUCTION

The traditional way of describing spoken language has been to divide the continuous speech signal symbolically into individual segments. The segments themselves are further analyzed as bundles of distinctive phonological features, which describe the basic phonetic character of the segments.

These distinctive features play at least three major roles in phonological theory. They distinguish sounds from one another, they group sounds together into natural classes, and they describe the phonetic character of the sounds. Previous to SPE (Chomsky and Halle, 1968), the emphasis had been primarily on describing the phonological contrast between segments, hence the name, DISTINCTIVE features. The SPE framework marked the beginning of a trend toward characterizing segments in terms of features that play a role in phonological rules and describe the phonetic content of the sounds as well as how they contrast with other sounds.

A characteristic of this framework is that the adjacent columns of features representing a sequence of segments do not overlap in any way with each other. The features in these columns or bundles that represent a phonological unit are claimed to be coextensive with no further organization or internal structure being attributed to them. A feature's value is assumed to be maintained throughout the duration of the segment. This type of description was adequate for segments that are phonetically homogeneous, or static, throughout their duration such as [i] or [ʃ]:

(1)

$$\begin{array}{cc} \text{i} & \text{ʃ} \\ \begin{bmatrix} +\text{syll} \\ -\text{cons} \\ +\text{son} \\ -\text{back} \\ -\text{low} \\ +\text{high} \\ -\text{round} \end{bmatrix} & \begin{bmatrix} -\text{syll} \\ +\text{cons} \\ -\text{son} \\ +\text{cont} \\ +\text{cor} \\ +\text{strid} \\ +\text{high} \\ -\text{ant} \end{bmatrix} \end{array}$$

Problems arise within the SPE framework from the recognition that there are some segments that are not homogeneous in time for the purposes of the phonology of a language. In other words, the static-valued SPE features are unable to describe segments for which part of this trajectory is relevant to a phonological process. As we will see later, it is often the case that phonetically dynamic segments exhibit such behavior. For segments such as these, the SPE framework invented ad hoc features.

A classic example of this use of ad hoc features is the case of the affricates and prenasalized segments which are distinguished from stops and non-prenasalized segments, respectively, by the feature [delayed release]. ([Delayed Release] as a feature of prenasalized stops is presented very tentatively, however; Chomsky and Halle, 1968:317). Using this feature to describe both the gradual way in which the cluster of an affricate is released, and in the prenasalized consonant, the releasing of the velum before the release of the oral closure is a problem As Anderson (1976:332) notes, the problem lies in "constructing a coherent definition for a feature like [delayed releases which will allow it to characterize both the oral articulatory distinctions for which it was designed and the nasal timing distinction to which it is to be extended". This static feature, [delayed release], is an attempt to describe the internal timing of inherently dynamic segments. In addition, extending this feature to the class of post-nasalized segments will obviously fail. The more common features used to describe pre- and post-nasalized segments are [+prenasal, +nasal] and [-prenasal, +nasal], respectively (Anderson, 1974:270). These combinations of features fail to capture the generalization that the nasal value is not constant for the duration of the segment.

Describing phonetically dynamic segments with features that are defined in such a way as to describe the gesture involved is not the only problem with the SPE approach. The use of such ad hoc features is only as constrained as the articulatory or acoustic possibilities of the segment themselves. The problem lies in that these static-valued features are unable to describe subsegmental elements that interact in some way with the phonology.

In this paper, I will look at the nature of some dynamic segments, and the question of how their internal structure should be represented in a phonological framework. We will see that current theories of representation (post-SPE) are able to handle those segments that are recognized as being phonologically dynamic, such as affricates and prenasalized segments (which are also phonetically dynamic), but in this paper, I will build on the findings of current work by positing significant internal structure for certain segments that have not previously been regarded as phonologically dynamic. We will see that certain phonological processes involving these segments can be captured more satisfactorily when expressed in terms of the interaction of their internal subelements and the phonology. The claims here will be based largely on the behavior of flaps, with a brief look at stops.

I first will consider the SPE framework's treatment of the alveolar flap, [ɾ]. We will see that this is another example of the use of ad hoc features to describe a segment that is not phonetically and phonologically homogeneous. SPE does not explicitly take a stand on what the featural description of this segment should be, but Chomsky and Halle suggest that the alveolar flap "is produced by essentially the same muscular activity that is found in the dental stop articulation, except that in the case of the tongue flap the movement is executed with great rapidity and without tension" (Chomsky and Halle, 1968:318). From this description, we might deduce the relevant features of the flap to be the same as the /t/ or /d/, with an added ad hoc feature [+rapid] (and possibly [-tense]) to describe what this framework deems to be the only difference between a flap and the dental/alveolar stops. The (partial) feature matrix of [ɾ] would look like this:

(2)

$$\begin{bmatrix} -\text{cont} \\ +\text{cor} \\ +\text{rapid} \\ -\text{tense} \\ +\text{voice} \end{bmatrix}$$

Given the above featural description of flap, then in the American English Flapping Rule, where alveolar stops are flapped in certain environments, the alternation between /t/ and [ɾ] would take the following form.

(3)

$$\begin{bmatrix} -\text{cont} \\ +\text{cor} \end{bmatrix} \text{--->} \begin{bmatrix} +\text{rapid} \\ -\text{tense} \\ +\text{voice} \end{bmatrix}$$

Another SPE-type featural description of [ɾ] is offered by Kahn (1976) that does not attempt to directly describe the change in articulation between /t/ and [ɾ], but uses instead the feature [+sonorant] to describe the overall change in the status of the vocal tract. The structural change of Kahn's flapping rule is as follows:

(4)

$$\begin{bmatrix} -\text{cont} \\ +\text{cor} \\ -\text{nasal} \end{bmatrix} \text{--->} \begin{bmatrix} +\text{son} \\ -\text{stiff v.c.} \end{bmatrix}$$

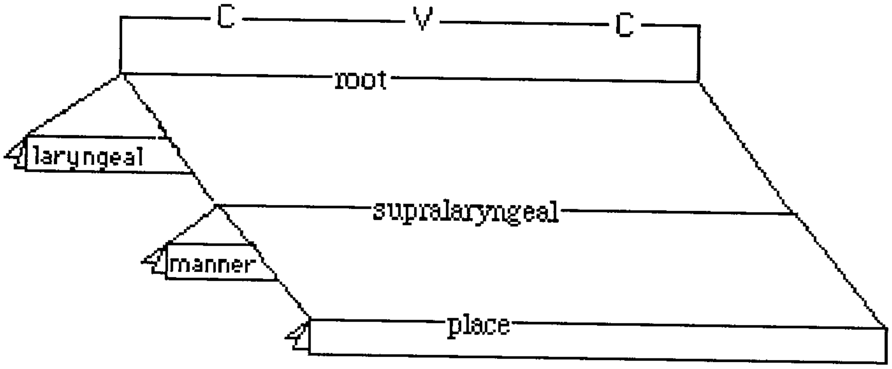
Although the SPE framework is able to describe, to some extent, the phonetically dynamic nature of the flap, we will see later that it is not just the phonetically dynamic characteristic of the flap that warrants a dynamic representation. The flap functions phonologically as a dynamic segment (i.e., a segment with significant internal structure) in English, and therefore must be represented as such in the phonology. The SPE framework is unable to handle any kind of internal structure of segments because it treats segments of columns of features that are uniform throughout the entire segment.

In the years since SPE was published, a number of phonologists have argued that the strictly segmental framework advocated by Chomsky and Halle does not adequately capture many important generalizations regarding phonological processes. A new framework has emerged which still acknowledges the importance of a featural classification system for segments but which, crucially, recognizes the need to represent the relationships that seem to exist among the phonological features and the apparent autonomy of some of the features. This general approach is called Autosegmental Phonology.

Among the many contributors to the autosegmental theory, Goldsmith (1976) established the need for a tonal tier, independent from but associated with the tier where the traditional segments are situated. He also provided a systematic formal representation for autosegmental phonological processes. Although his system of representation had its foundation mostly in tonal phenomena and nasality, Goldsmith made a more general prediction as to the future direction of the framework by suggesting that the speech signal should be broken down into at least as many independent tiers as there are independent articulators. Since Goldsmith's seminal work on the formal representation of autosegmental phenomena, many others have further developed the theory, such as Clements (1985) and Sagey (1986).

Clements (1985) developed a hierarchy for the organization of phonological features based primarily on accounts of assimilation processes cross-linguistically. He asserts that justification for the organization of phonological features should not come from "a priori considerations of vocal tract anatomy", contrary to Goldsmith's prediction. The groupings of the features into the tree shown in Figure (5) are based on what Clements claims are phonological and phonetic processes. In addition, Clements assumes the temporal organization of CV Phonology proposed in Clements and Keyser (1983).

(5)

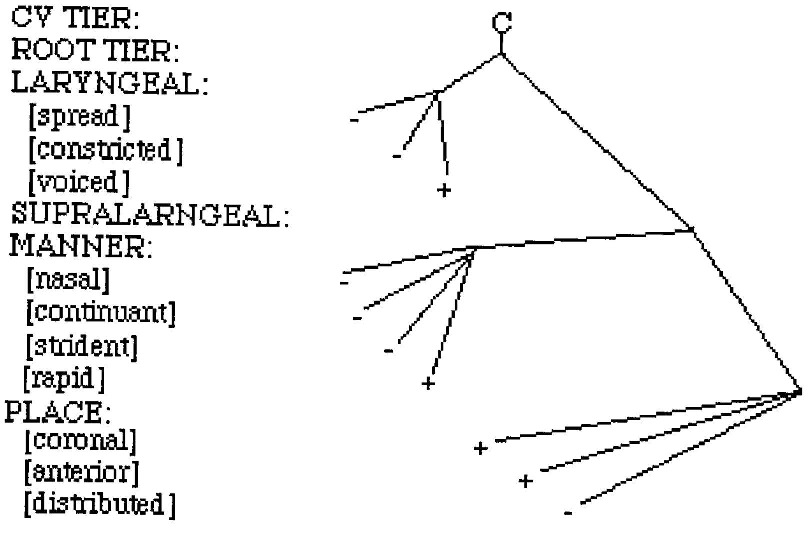


The evidence from phonological assimilation rules seems convincing and the placement of the individual tiers seems basically appropriate at least for those cases that Clements presents in this article.

Clements claims that some of the justification for the organization of the features should come from phonetic processes as well, though he does not provide us with his working definition of what qualifies as a phonetic process. Nevertheless, if it is the case that some of the evidence for this model comes from phonetic data, then the model should be able to handle all (linguistically significant) phonetic processes. One such phonetic process on which we might test the Clementsian model is Alveolar Flapping in English.

Representing the flapping rule in this framework would not look much different from the SPE-type rule because the features utilized in the Clementsian tree are essentially the same static features SPE had to work with. The timing tier (CV-tier or X-tier) and root tier consist of units that are instantaneous points on a number line (where the units are like the integers rather than the real numbers.) The geometry of features put forth by Clements provides the logical organization of features, but does not attempt to manipulate the internal timing of the features or segments. The plane that the feature tree occupies is the logical dimension, not the time dimension. This version of Clementsian feature geometry is not much more enlightening than SPE features in terms of representing the internal structure of the flap. The flap in Clementsian features might look like this:

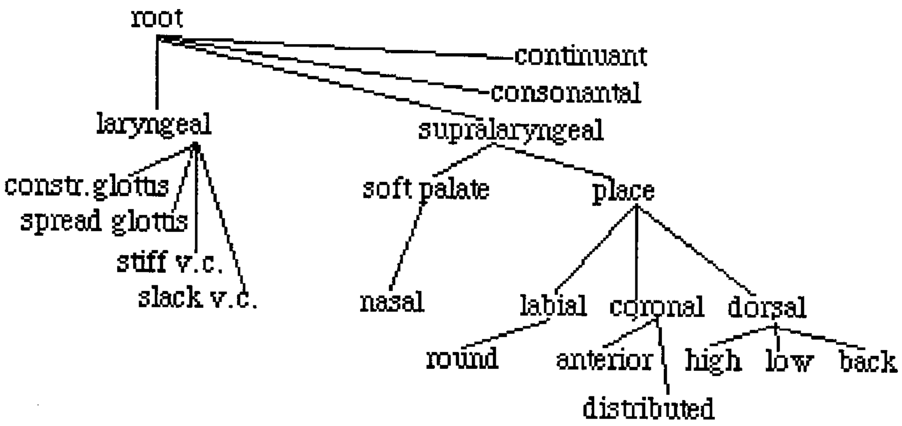
(6)



Even if a more detailed study were to reveal more about the articulatory nature of the flap, it seems unlikely that the Clementsian model would be able to incorporate that new information because of the assertion mentioned above that articulatory phonetics cannot provide justification for the form of a phonological model. The only evidence that this model could take into account, according to Clements, would be proof that the dynamic nature of the flap in involved in a phonological process in the language in question. We will see later that this is precisely the case for the flap in American English, but Clements' model as it is presented here (Clements, 1985) is not developed enough to describe the internal structure of the flap.

Sagey (1986) carries further Clements' work on the organization of phonological features, accounting for many important phonological phenomena. Sagey builds on the hierarchical organization of the features. She reaffirms the separation of the place features onto a separate tier from the manner features based on rules that involve place assimilation with manner assimilation. Sagey uses assimilation processes, and properties of multiply-articulated segments, to arrange the place of articulation tiers into independent tiers that correspond to independent articulators, reminiscent of Goldsmith's prediction. The representation she proposes is shown below:

(7)

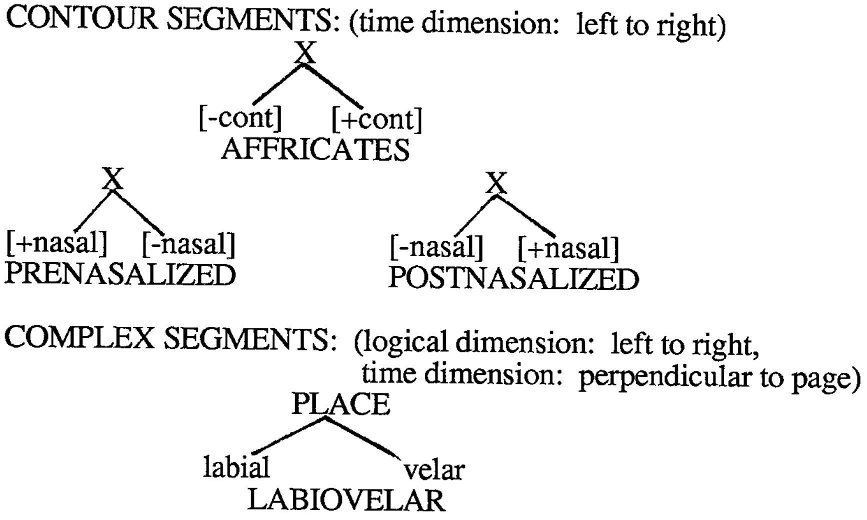


The most significant contribution that Sagey makes to the theory that is most relevant to the issue at hand is to distinguish between simple and multiply-articulated segments. She further distinguishes two classes of multiply-articulated segments that function as single timing units (i.e., are associated with one x-slot, and are syllabified as one segment): complex segments and contour segments.

Complex segments involve more than one articulator, and the articulations are phonologically simultaneous. Complex segments can consist of two or three different articulators being simultaneously associated to one phonological timing unit which is treated as a single segment in phonological rules. Such segments include labiovelars, labiocoronals, clicks and others. The relative temporal ordering of the multiple articulators, according to Sagey, is not specified in the underlying representation of the segment. The temporal ordering of the individual articulations is determined at the level of phonetic interpretation for the particular language. For more details on complex segments and on the percolation of manner features to the different articulators, see Chaps. 2 & 3 in Sagey, 1986.

Contours segments involve a single articulator but consist of a sequence of articulations which are phonologically and temporally ordered. This class of segments consists of affricates and pre- and post-nasalized segments. Where the SPE framework introduced the feature [delayed release] to describe these dynamic segments, Sagey's treatment acknowledges that there are essentially two parts to these segments that are necessarily ordered in time. Because they are ordered in time, they branch at the appropriate node in the time dimension. Complex segments, on the other hand, branch in the logical dimension under the place node.

(8)



The relative temporal ordering of the subparts of a contour segments is specified in the structural description of the segment. One can determine the ordering of these subparts by their interaction with their surrounding environment. These interactions are called "edge effects". Sagey cites data from Sierra Popoluca, a Zoquean language, to demonstrate edge effects of affricates that establish the relative ordering of their parts. Her observation is that stops are aspirated at the end of a syllable, but fricatives and affricates are not.

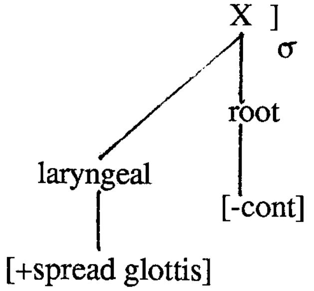
(9)

STOPS	/həp/	[həpʰ]	'mouth'	
	/ʔampat/	[ʔampatʰ]	'I met'	
	/mək/	[məkʰ]	'fog'	
AFFRICATES	/mac/	[mac]	'grasp'	(*macʰ)
	/ʔapič/	[ʔapič]	'thorn'	(*ʔapičʰ)
FRICATIVES	/wəstən/	[wəstən]	'two'	(* wəsʰtən)
	/pištək/	[pištək]	'flea'	(* pišʰtək)

Sagey's analysis of this pattern is that an aspiration rule applies to a [-cont] at the end of a syllable, as in the rule below, so the rule will not apply to affricates which "although they contain a specification [-cont], are phonologically [+cont] on their right edge, to which the rule is sensitive". (Sagey, 1986:95-96). Her rule is shown in (10):[1]

(10)

SIERRA POPLUCA ASPIRATION RULE



(Sagey, 1986: 96)

Another claimed example of contour segments exhibiting edge effects is in Land Dayak. In this language, vowels are nasalized after simple nasals, but not after prenasalized stops.

(11)

[mãlu]	'strike'	[sam͡pe:]	'extending to'
[nə̃bur]	'sow'	[sun͡tɔk]	'in need of '
[ənə̃k]	'child'	[suŋ͡koi]	'cooked rice'

Specifying these prenasalized segments as [+prenasal] does not allow us to express the generalization that these segments are [+nasal] on the left and [-nasal] on the right, so the vowel to the right does not become nasalized. (Sagey, 1986:97, Anderson, 1976).

This treatment of affricates and prenasalized stops captures the phonetic and phonological dynamic aspect of these segments better than the SPE static features. The SPE features [+del rel] and [+prenasal] encode no information as the the identity of the feature values at different points within the segment. The contoured structures allow us to express the differing feature values within the segments as they are relevant to the phonological process in question. The single-valued, static feature [delayed release] has been translated into two discrete phonologically significant events that operate as a single phonological unit. The segments that were traditionally characterized as [del rel] have been, in effect, disassembled or "unbundled" to reveal the internal structure of these segments. This is true for the static features [+prenasal] and [-prenasal] (with [+nasal), as well.

With the introduction for the concept of temporally branching contour segments, it is tempting to impose a contoured representation on the phonetically dynamic flap. Sagey would argue that if the phonology does not need to access the internal structure of a segment, the phonological description should not include this structure (Sagey, 1986:21). There is merit in this viewpoint. It certainly seems to be the case that there is phonetically redundant and predictable information in spoken language that we should not include in an abstract representation. However, in the case of the alveolar flap, there is evidence that American English phonology accesses some as yet unidentified internal structure of this segment. In the next section, we will see the evidence for positing such internal structure which a fully explicit phonological theory must be able to express.

II. FLAPPING IN AMERICAN ENGLISH : KAHN (1976)

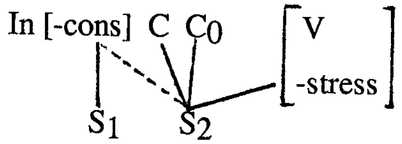
In this section, we will look closely at two analyses of the phonological process, Flapping in American English, that treat the alveolar flap as an unanalyzable unit. The particular analyses we will examine are Kahn's (1976) classic treatment of the English allophones of /t/, and Selkirk's (1982) account of the same phenomena. Because of the inaccessability of the internal structure of this segment in these SPE-style analyses, the phonological environments of the rule that gives the flap allophone of /t/ look accidental. We will see that by positing that this segment is phonetically and phonologically dynamic, and has internal structure, the conditions on the application of the flapping rule can be expressed as the interaction between the values of the adjacent segmental environments and the values of the edges of the flap segment. Certain aspects of Kahn's and Selkirk's analyses will survive our reanalysis (i.e., resyllabification); however, the featural descriptions of the processes that are offered here miss the generalization that the segmental environments are interacting specifically with the edge values of the flap.

Kahn (1976) gives an account of six allophones of /t/ in American English. We will look primarily at the flap allophone [ɾ], and the glottalized [tʔ] for comparison. Kahn proposes a set of syllable structure assignment rules for English that assign segments to syllables, adhering to syllable phonotactics and the general principle of maximal onsets. (See Kahn, 1976:32 for a summary of these rules.) In normal and faster speech, the syllable boundaries are blurred phonetically and phonologically, according to Kahn. The phonetic boundaries of syllables in normal speech are not instrumentally discernible. Kahn carries this observation into the phonology of syllables, claiming that "There need not correspond to every pair of adjacent syllables a well-defined boundary". (Kahn, 1976:17). From this position, Kahn is able to establish a formalism for assigning syllable membership to phonetic strings that allows a segment to be the final member of one syllable and the initial member of the next syllable, i.e., to be ambisyllabic. Ambisyllabicity is not an underlying characteristic of segments but comes about as a result of the resyllabification of certain segments in connected speech. This ambisyllabicity of segments can act as the conditioning environment of some phonological rules, as we will see.

Kahn posits three separate resyllabification rules to account for a variety of cases, as we will see subsequently. These resyllabification rules are shown below. Notice that the C that resyllabifies maintains its original syllabification; thus, the segment becomes ambisyllabic.

(12)

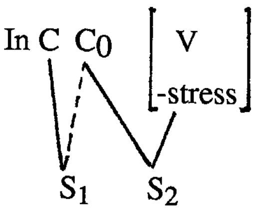
RULE III (from Kahn, 1976): (normal-rate and faster speech only)



associate C and S_1

(Initial member of S_2 becomes final member of S_1 = ambisyllabic)

RULE IV: (normal-rate and faster speech only)

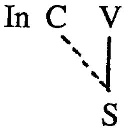


associate C and S_1

Condition: C C_0 must not be a member of the set of universally prohibited clusters; certain highly marked clusters not universally proscribed may be excluded also.

(Final member of S_1 becomes initial member of S_2 = ambisyllabic)

RULE V: (connected speech only)



associate C and S

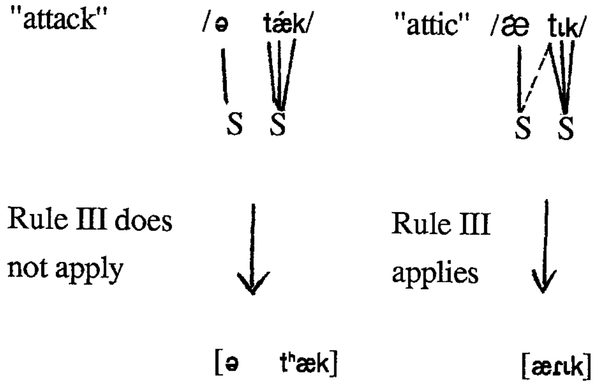
(Extends domain of syllable across word boundary.)

Kahn's analysis attributes the distribution of the allophones of /t/ primarily to differences in syllabification. These resyllabification rules (especially III) are crucial to Kahn's analysis of flapping because he is able to use the ambisyllabicity of /t/ to condition the flapping rule. Rule III enables Kahn to distinguish between the environments that yield the aspirated [tʰ], glottalized [tʔ]

and the flapped [ɾ]. The underlying /t/ becomes aspirated when it is syllable-initial ([+S.I.]), and not syllable-final ([-S.F.]), i.e., not ambisyllabic.

If the conditions are right for a /t/ to become ambisyllabic (as in Rule III above), the environment is right for flapping. Compare the syllabification of the two words "attack" and "attic" which differ primarily in their stress patterns:

(13)



The crucial segmental environment for a consonant becoming ambisyllabic is if it is preceded by a [-cons] segment, as is indicated by the statement of RULE III. In order to capture his intuition that the syllable structure is parallel between "after" and "faster" (where sharp syllable juncture after the "f" in "after" and after the "s" in "faster" is lost), Kahn must introduce Rule IV which causes the "f" to become ambisyllabic. These two rules also have the effect of apparently accounting for the fact the [+cons] blocks the flap from occurring, and [-cons] allows the flap. However, this effect is not acknowledged as such. The phenomenon is couched solely in terms of syllabification. For instance, the /t/ in"outer" is ambisyllabic by Rule III, but the /t/ is the ambisyllabic segment in "after", by Rule IV:

(14)

"outer" "after"

aʊt r̩ æft r̩
\| \| |/ |/
S S S S

by Rule III by Rule IV

V V

[aʊɾr̩] [æftr̩]

The domain of application of Rules III and IV is the word. However, in connected speech where there are no pauses between words, Rule V can apply to /t/ across word boundaries, and flapping can result, where [tʔ] would have occurred in careful speech (i.e., with pauses between words), as in:

(15)

CAREFUL CONNECTED
"at it" /æt # ɪt/ "at it" /ætɪt/

Rule V || Rule V ||
does not || applies ||
apply || ||
V V

[ætʔʔɪt]2 [æɾɪt]

Notice that Rule V also ambisyllabifies the /t/ in "kept it" in connected speech. This /t/ is not flapped. This counterexample reveals two important aspects of the behavior of flap in English. First, ambisyllabification can occur without the [-cons] environment (by Rule V), so [-cons] must not be relevant to resyllabification (if Rule III, IV and V are considered to be variants of a single casual speech process). Second, the flap appears to be blocked by a preceding [+cons] segment, because this is the only difference in environment between "at it" and "kept it". Therefore, [-cons] must be part of the flapping rule, not part of resyllabification.

Kahn equates the [ɾ] and [tʔ] by calling them both "sonorized" versions of /t/ (by the Sonorization Rule). They are then distinguished only by their glottal features: [ɾ] = [-stiff v.c] (Voicing Rule) and [tʔ] = [+constricted glottis] (Glottalization Rule). But there is additional evidence that the syllabification of /t/ is not the only conditioning factor that leads to the choice between [ɾ] and [tʔ]. In the following list of words, the syllable structure seems to be identical, but the realization of the underlying /t/ differs for the word "beaten":

(16)

"beetle"	[biyɾl̩]
"beater"	[biyɾr̩]
"beating"	[biyɾɩŋ]
"*beaten"	* [biytʔn̩]
"beat 'em"	[biyɾm̩] (or [biyɾəm])

In all of these cases, the segment to the right of the /t/ is [+syllabic], as the "voicing" rule requires, and the syllabification and the lefthand [-cons]) environment of /t/ are the same. The Sonorization and Voicing/Glottalization rules (Kahn, 1976:63) would produce *[biyɾn̩], as shown in the derivation below:

(17)

```
beating        beaten

b i t ɩ ŋ      b i t n̩
 |   |          |  |        by Rule I
 o   o          o  o        (associate +syll to o)

b i t ɩ ŋ      b i t n̩
 \|  \|         \|  \|      by Rule IIa
 o   o          o   o       (associate permissible
                             initial clusters)

b i t ɩ ŋ      b i t n̩
 \|  \|/        \|  \|      by Rule IIb
 o   o          o   o       (associate permissible
                             final clusters)

b i t ɩ ŋ      b i t n̩
 \|/ \|/        \|/ \|      by Rule III
 o   o          o   o       (ambisyllabify if o2 =
                               [-stress])

b i T ɩ ŋ      b i T n̩
 \|/ \|/        \|/ \|      by Sonorization
 o   o          o   o       (T =   t
                                  [+son])

b i ɾ ɩ ŋ                   Voicing
 \|/ \|/
 o   o
               *b i ɾ n̩     Voicing applies
                \|/ \|      erroneously
                 o  o
               b i t? n̩     Glottalization
               \|/  \|      should apply
                o   o
```

This breakdown of the flapped versus glottalized environments demonstrates that there is some difference in the two allophones other than syllable membership, a difference that causes the syllabic [n̩] to block the flap, and for the [-syll] (or pause) requirement of the righthand side of [tʔ] to be overridden. The segment that blocks the flap from occurring is the syllabic nasal [n̩], which involves alveolar contact. The syllabic [l̩] (in my own speech) tends to be more back (velarized) than a regular post-vocalic [ɫ], resulting in no coronal contact. The [r], in [ɹ] in [ɹŋ] and [m̩] (or [əm]) obviously entail no coronal contact. This would imply that the flap gesture requires a righthand environment of "no alveolar contact" (i.e., [+cont]) in addition to its ambisyllabicity requirement.

An articulatory description of this peculiar blocking effect account for the facts. According to Ladefoged (1982:154), the flap is a ballistic gesture which entails throwing the tongue tip against the alveolar ridge and withdrawing it again. The flapping sound is produced by the tongue tip on its way by the alveolar ridge. The return or release gesture appears to be a crucial part of the ballistic movement. If coronal alveolar contact is expected immediately after the medial /t/, the tongue will tend to economically maintain the alveolar contact it began during the /t/ (or [tʔ]) rather than flap and then return to the alveolar ridge.

The explanation may not be purely articulatory, but the pattern we have observed seems to point in that direction. For instance, what articulatory account, separate from syllabification, can be given for why the flap does not occur when followed by a [-syllabic] segment, such as /r/ (mattress) or glide (Antwerp, amateur), which also has no alveolar contact? It is possible that flapping is blocked by the affrication of the /t/ in these contexts (i.e., the affrication rule applies before flapping.) Flapping also seems to be blocked by a following [+cont] segment (which is also [-syllabic]), as in"Betsy" [bɛt(ʔ)si] and "fatso" [fæt(ʔ)soʊ]. Since there is no final syllabic /s̩/ in English, we cannot test if a [+syllabic, +continuant] segment allows or blocks flapping. Acknowledging the lack of evidence for a [+syll, +cont] righthand environment of Kahn's flapping rule:

(18)

Flapping Rule in English

```
t ----> ɾ  /  [-cons] ____   ┌+syll ┐
                        /\   └-cons ┘
                       /  \
                      S    S
```

Notice that [m̩] allows flapping, as in "beat 'em" or "atom". To capture the generalization that any syllabic noncoronal, even a [-cont] one, allows flapping, we need to add the additional righthand environment [-cor,+syll] to the structural description. This will be discussed in more detail later in this section. The syllable structure seems to set up the environment for flapping (i.e., the /t/ must be ambisyllabic), but the tongue tip status determines whether flapping will ultimately occur. If flapping is blocked, the glottalized [tʔ] is often realized, because the Glottalized rule is ordered after the flapping rule. Both of these rules are ordered after the Aspiration rule, according to Kahn. I have no articulatory explanation for why the glottalized allophone is the default allophone for /t/ from this data. I have avoided stating the flap in features in this rule because we have not yet established the correct representation in features for the flap.

We have just seen that the righthand segmental environment had to be revised to [+syll, -cons] to account for the edge effects it exhibits. Let's look more closely at Kahn's assessment of the lefthand environment. According to Kahn, the basic segmental environment of flapping is:

(19)

t ----> ɾ / [-cons] ___[+syll]

As we saw above, when a /t/ is preceded by a [+consonantal] segment it does not flap (as in "kept it" and "faster"), so at the segmental level (ignoring notion of syllabicity) this environment seems basically right.

However, flapping does appear to occur after some consonants. The segments in question are the liquids and nasals. The English /t/ and /r/ are specified underlyingly as [+consonantal], as are the nasal stops. According to Kahn's flapping rule, these "consonants" should block the flap from occurring (by blocking the ambisyllabification of /t/). However, as Kahn noted, the facts show us that this is not the case.

Flapping almost always occurs after [ɹ] (underlyingly /r/) when the appropriate righthand environment is present:

(20)

"parting" [paɹɾɩŋ] "artery" [aɹɾr̩i]

In addition, flapping seems to occur after alveolar nasal consonants in non-careful speech:

(21)

"enter" [ɛ̃(n)ɾr̩] "anti-freeze" [æ̃(n)ɾifriz]

For some speakers, in some dialects (the particular dialects were not pinpointed by Kahn), flapping occurs after post-vocalic /l/ (i.e., velarized [ɫ]):

(22)

"revolting" [rivoɫɾɪŋ] "alterations" [aɫɾeyšn̩z]

To account for flapping occurring after apparently [+consonantal] segments, Kahn takes a closer look at the phonetic nature of these segments. In the first case, Kahn argues that the English /r/, which is specified in most descriptions of English as consonantal, is not [+consonantal] at the phonetic level. In SPE terms, a consonantal segment is one that involves "a radical obstruction in the midsagittal region of the vocal tract: (SPE:302). Kahn states that the English (American and British) /r/ is different from most r-like segments in that it involves not obstruction at any time during its articulation. The narrow transcription for this sound is actually [ɹ]. Kahn gives the following feature specification of [ɹ], putting it in the natural class of glides.

(23)

English /r/ ([ɹ])

$$\begin{bmatrix} \text{-syllabic} \\ \text{-consonantal} \\ \text{+sonorant} \\ \text{+high} \\ \text{+back} \\ \text{+coronal} \end{bmatrix}$$

Kahn also provides compelling arguments for [ɹ] acting as a glide with regard to its phonotactic distribution in the language (Kahn, 1976:95-97). In view of the glide-like properties of [ɹ], then, it is not surprising that the /t/ flaps when it follows [ɹ], because, as we saw above, flapping takes place after the other true glides in English. Zue and Laferriere (1979) in an acoustic study of a large speech sample report the probability of occurrence of flap in this environment as 0.99 (Zue and Laferriere, 1979:1048).

When flapping occurs in a word like "winter", there would appear to be a consonantal segment preceding the [ɾ]. However, according to Kahn, many speakers have two pronunciations of the /nt/ segment sequence. In careful speech, the /n/ is elided, being realized only as nasalization on the preceding vowel, producing [wɪ̃ɾr̩]. (It was pointed out to me be I. Maddieson that most speakers nasalize the flap as well in this environment., i.e., [wɪ̃ɾ̃r̩].) There is no consonantal obstruction associated with the nasal segment, so the environment preceding the /t/ is [-consonantal] (a nasalized vowel), and flapping obligatorily occurs. The claim is that flapping is impossible if the alveolar contact of the /t/ is maintained, as in *[wɪ̃nɾr̩]. In addition, not flapping is unnatural (but not impossible) if the /n/ has no alveolar contact: *[wɪ̃tʰr̩]. Zue and Laferriere (1979) give the probability of occurrence figures for all of the phonetic realizations of the

/VntV/ sequence that were observed in their study. The breakdown of probabilities is shown below (Zue and Laferriere, 1979:1048):

(24)

$$\text{Vnt} \dashrightarrow \left\{\begin{array}{l} \text{Vnt} \\ \text{Vn} \\ \text{ɾ̰̃} \\ \text{t̰ʰ} \end{array}\right\} / \;\_\_\text{V} \qquad \begin{array}{l} 0.67 \\ 0.14 \\ 0.14 \\ 0.05 \end{array}$$

The coronal alveolar lateral liquid /l/ is specified as [+consonantal] because of its central (midsagittal) contact with the alveolar ridge. If this segment is [+consonantal], why then can it be considered an environment for flapping? Kahn proposes that for those speakers who tend not to flap in words like "shelter", the /l/ maintains its consonantal pronunciation. Speakers who habitually flap in this position produce a non-consonantal /l/, one without alveolar contact, [ɫ]. The extreme version of this allophone would be a velar (unrounded) glide, [ɯ], as in the pronunciation [mɯɯk] for "milk". Kahn suggests that flapping after a consonantal [ɫ] seems impossible. Zue and Laferriere confirm Kahn's intuition. They report that although only a small percentage of post-lateral /t/'s are flapped, "in the case of these flaps, the tip of the tongue probably did not make full alveolar contact in the articulation of /t/, enabling the tongue tip to engage in the flapping motion". (Zue and Laferriere, 1979:1046).

Notice that where there is allophonic variation in the environments in question, the alternations between the set of contextual allophones that permit flapping and the set of contextual allophones that block flapping alternate with respect to one isolatable variable: relative freeness of the tongue tip.

Although Kahn identifies these allophones, and correctly portrays their phonetic realization, he does not explicitly acknowledge the articulatory contribution of these segments in blocking or permitting flapping in the statement of the flapping rule. This is partially due to the limits of the featural description that was available for the flap.

Kahn admits his uncertainty as to how to characterize the flap in phonological features. Kahn classifies [ɾ] as [+sonorant] because trilled /r/'s are generally classified as [+sonorant]. He adds that some additional feature might be needed to distinguish it as a single rather than multiple-tap trill (a feature that is not available in the SPE framework). Based on this, he uses this feature [sonorant] to describe the change that takes place between /t/ and its allophone [ɾ], addition to a later voicing rule to distinguish the [ɾ] from the [tʔ], which he also classifies at [+son]. This choice of features is clearly an ad hoc solution. Capitalizing on our observations concerning the correlation between certain allophonic variants and the appearance of the flap can lead us to a less add hoc, more satisfactory expression of this complicated rule.

The articulatory explanation for the interaction of these adjacent segments is based on the fact that the segments involve the same articulator: they are [+coronal]. The articulatory nature of the flap gesture must be taken into account as well. We have seen that the flap is phonetically dynamic. It has been described as ballistic in the sense that one articulator, the tongue tip in this case, thrown up against (or toward) another, the alveolar ridge (Ladefoged, 1982:281). The tongue tip must start from a certain distance away from, or else be drawn back from, its target in order to give the gesture the momentum required to make it flap against its destination. An analogy that might make this concept clearer is the feature involved in throwing a ball. The toss is more effective (i.e., the ball goes further) if you "cock" your arm back from its neutral position, to give the throwing gesture momentum. The goals of the two gestures are somewhat different. The goal of throwing a ball is for the ball to be propelled a certain distance. The goal of the alveolar flap is for the tongue tip to strike (or at least nearly strike) the alveolar ridge on its way by, the result being a flapping sound. Each gesture needs the "cocking" gesture in order to give it a running start to achieve its goal.

In an environment where a /t/ is a candidate for flapping (where it is ambisyllabic and followed by [+syll]), if the tongue tip is raised for some other sound, like an [s] or a fully articulated [n], whose tongue tip has gone beyond the "cocked" position needed to initiate the flap, the flap will not occur. It would be necessary for the tongue tip to lower itself from the [s] position to the "cocked" flap-initial position in order to produce the flap. The generalization is that segments that cause the tongue tip to be at or beyond a certain position, which is below that threshold position, block the flap from occurring, and segments whose tongue position is below that threshold allow flapping. It also seems that the threshold is different for the cocking and releasing parts of the gesture, as we have seen in the asymmetrical environments of the rule.

This articulatory account of the interaction between the flap and its environment is very attractive, but it does not explain why flapping is blocked in words such as "actor" and "after". In these cases, the tongue tip is not involved in the articulation of the consonant that precedes the /t/, so it cannot be a coronal contact that blocks the flap. We established with the case of "kept it" that we cannot fall back on the notion of ambisyllabicity in the case. That is, we cannot say flapping is blocked from the left when the tongue tip itself is raised beyond the threshold. We have seen that attributing [-cons] to the ambisyllabification rule is incorrect. Therefore, it does seem to be the case that the lefthand environment of the flapping rule is [-cons]. The threshold that appeared to be specific to the degree of stricture of the tongue tip is actually a threshold that applies to the oral cavity in general, and it is this threshold the [consonantal] feature describes. However, the interaction of the coronality of the flap and its righthand environment does seem to be relevant, as we demonstrated with "beaten".

How are we to express the fact that flapping is blocked by [-continuant] coronals from the righthand side? The feature [continuant] must in this case be articulator-specific, as opposed to a feature such as [sonorant] which describes the global status of the vocal tract. Current theories of feature geometry would need to be revised to accommodate this distinction, since all manner features are implied to be global because they all reside under the major class node, MANNER. In order for a feature such as [continuant] to be considered articulator-specific, the feature would have to be percolated down to the active articulator node, or else the [continuant]would reside under the particular articulator node (e.g., CORONAL) in the underlying representation of the segment. This percolation process and/or a revised segment structure tree needs to be more fully developed than I will be able to do here. For the sake of simplicity, in the rules I am about to propose, I will simply specify [+coronal] as well as [continuant], even though these features are not under a single node of a feature tree.

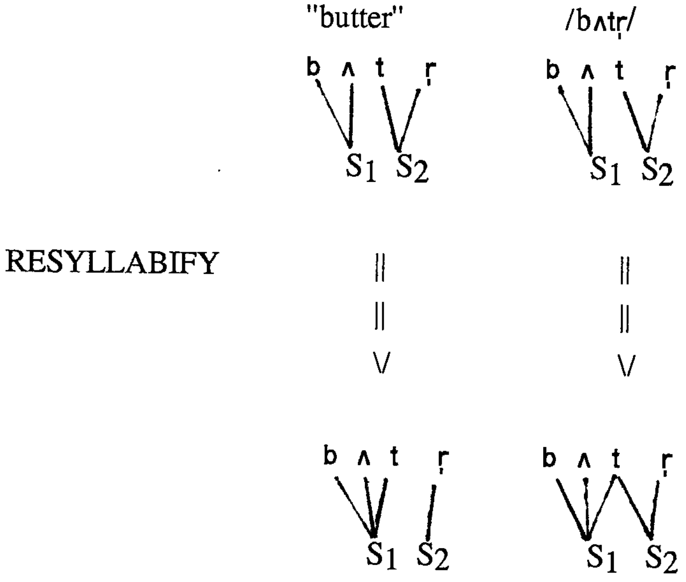
The features [-consonantal] on the left and [-consonantal] (plus [+syll,+cor]) on the right superficially identify the segmental environments that permit flapping, but as I mentioned above, they do not capture the threshold effect that we have observed at the edges of the flap. The reason for this is that the manner of articulation of the flap has not been adequately specified in features. Kahn's attempt to characterize [ɾ] as [+sonorant] fails to capture the complex ways the flap interacts with the segments to its left and right. We have seen that this interaction is similar, though not exactly identical, on both sides of the segment. These edge effects suggest that the phonological (and phonetic) identity of the flap is not uniform throughout its duration. We will pursue this issue further after we look at another analysis of flapping in English, offered by Selkirk (1982).

III. FLAPPING IN AMERICAN ENGLISH : SELKIRK (1982)

In Selkirk's (1982) analysis of the same phenomena (i.e., allophones of /t/ in English), a fundamental assumption is that the syllable has its own hierarchical structure, as opposed to the flat structure assumed by Kahn (1976). The principle of Basic Syllable Composition takes the form of a template and a set of collocational restrictions that are language specific. Selkirk's arguments for this structure are beyond the scope of this paper, so I will not recount her entire analysis here. These syllabification principles act as well-formedness conditions rejecting syllabifications that do not match the template. This is different from Kahn's strategy whereby phonetic strings are assigned their syllable structure by associating segments to particular syllables. In addition, the Maximal Onset Principle is at work to maximize the onset of syllables in conformance with the permissible onsets of the particular language. Kahn also assumes this principle.

Selkirk's analysis entails a Resyllabification stage of the derivation which applies under particular conditions during normal and fast speech. It takes the form of associating "a consonant with a preceding syllable accompanied by a complete dissociation of the consonant from its syllable of origin" (Selkirk, 1982:316). In Kahn's analysis, a similar resyllabification process applies after initial syllabification of an utterance. It produces ambisyllabic segments because it does not dissociate the consonant from its original syllable. This is the crucial difference between Selkirk's and Kahn's approaches. We can see this difference in the figure below, where the word "butter" is resyllabified by Selkirk's strategy in (a) and by Kahn's Rule II in (b). (Note: I am ignoring the internal structure of the syllable in Selkirk's figure, as it is not immediately relevant in this example.):

(25)



(a) Selkirk (1982) (b) Kahn (1976)

Selkirk's claim is that the notion of ambisyllabicity is unnecessary and unsupported, and complicates the expression of the flapping rule. She claims that the same phenomena that Kahn accounts for by invoking ambisyllabicity (distribution of [ɾ] and [tʔ]) can be expressed generally without the notion of ambisyllabicity. Since Kahn's crucial argument for ambisyllabicity comes from his analysis of these allophones, if an equally good or better account of the same facts does not make use of this notion, ambisyllabicity must be abandoned as superfluous, or possibly incorrect.

In addition, Selkirk claims that where Kahn needs two separate resyllabification processes (Rule II and IV) to account for the parallel syllable structure between "after" and faster", her analysis can collapse these two processes into one process, capturing the generalization that these two processes have the same conditions (stress and onset). The collapsed resyllabification rule is shown below:

(26)

$$\begin{array}{llllll}
X & \begin{Bmatrix} [-\text{cons}] \\ [+\text{cons}] \end{Bmatrix} & \underbrace{s\ [-\text{syll}]} & \begin{bmatrix} +\text{syll} \\ +\text{stress} \end{bmatrix} & Y & \begin{matrix} \text{OBLIG} \\ \text{OPT} \end{matrix} \\
 & & & & & \\
1 & 2 & 3 & 4 & 5 & ====> \\
1 & 2+3 & 0 & 4 & 5 &
\end{array}$$

Selkirk assumes that the entire onset "st" in "faster" is resyllabified based on the observation that the /t/ is not aspirated, and is therefore not syllable-initial.

Where Kahn gives syllable structure as the only conditioning distinction between [ɾ] and [tʔ], Selkirk must account for this distinction some other way because both [ɾ] and [tʔ] occur syllable-finally in her analysis. The most important condition, according to Selkirk, is whether or not the consonant is released. She introduces the feature [release] as a feature that, though it is not contrastive in English, can be manipulated by phonological rules. (The feasibility of this feature will be addressed later.) The unmarked specification for English is [+release], and it is the task of the grammar to "capture the conditions for non-release. (Selkirk, 1982:374). Stops in the onset are always released, so the question of release versus nonrelease arises only with syllable final stops. Selkirk does not justify this claim, but it is probably because stops in the onset always occur before segments that are [+continuant]; we will see a phenomenon similar to this in Sierra Popoluca in a later section. The assignment of [-release] to a syllable-final stop is determined by its surrounding (segmental) context. A summary of the relevant contexts compiled by Selkirk is given below:

(27)

Stops ---> [-release] / ...$]_{\sigma}$

a. [-cons]______[-syll] -- OBLIGATORY
(e.g., ra**p**t, atlas, infe**ct**)

b. [-cons]______[Pause] -- OPTIONAL
(e.g., Did you see the ha**t**. That's not ba**d**.

c. [+cons]______[+cons] -- OPTIONAL depending on
which consonant follows (/t/ and /k/ frequently
elide) (e.g., exa**ct**ly, las**t** night, lef**t** me)

d. [+cons]______[-cons] -- POSSIBLE but not preferred
(e.g., las**t** year, ac**t** really nice)

e. [+cons]______[Pause] -- POSSIBLE but not preferred
(e.g., ac**t**, was**p**, as**k**, fas**t**)

Where [-release] is assigned to a voiceless stop, a glottalization rule applies, resulting in closure at the glottis. She does not "trouble to formulate" (HER words) the rule that assigns [-release] to segments, stating that the environments are obvious. It is certainly trivial to write a rule that just lists all of the above context if one does not attempt to unify the occurrences of [-release] (and thus [tʔ]).

Selkirk does not attempt to unify these environments. In addition, her assessment of the glottalization ([-release]) environments seems incorrect, in that the environments she posits as possible for glottalization in English are unnatural at best. For instance, in (27c), (27d) and (27e), glottalization of the /t/ in natural speech seems quite unlikely, if not impossible in my own idiolect and that of the speakers I have consulted. In (27c), the /t/ is most naturally elided; in (27d), the glottalization of /t/ seems to be blocked by affrication caused by the following glide; and in (27e), released [t] is most natural with [tʔ] feeling very awkward. We should regard these cases cautiously as being extremely unnatural environments for [tʔ]. The only remaining cases are (27a) and (27b) where syllable-final /t/ preceded by [-cons] and followed by [-syll] or Pause is assigned [-release] (i.e., is glottalized).

(28)

$$\text{Stops ---> [-release] / [-cons]\_\_\_\_}]_{\sigma} \left\{ \begin{array}{ll} \text{[-syll]} & \text{OBLIG} \\ \text{[Pause]} & \text{OPT} \end{array} \right\}$$

According to Selkirk, stops that have not been assigned [-release] have the default feature value [+release]. This default is realized for /t/'s as [t] or [ɾ]. There are three conditions that must be met before /t/ will be realized as [ɾ]: the /t/ must be syllable-final, [+release], and preceded by a [-cons] segment. The rule looks like this:

(29)

$$\text{t (and d) ---> ɾ / [-cons]}\left[\left[\begin{array}{c}\_\_\_\_\_\_ \\ \text{+release}\end{array}\right]\right]_{\sigma} \qquad \text{OPT}$$

Notice that this rule does not specify a righthand environment. This is because Selkirk erroneously claims that a voiceless flap can occur before a pause, as in "Did you see the hat?", *[hæɾ̥]. I disagree with this observation. The /t/ does not flap here in my speech or that of any of my consultants. It is either glottalized [tʔ] or released (unaspirated) [t]. Unfortunately, being able to put flap in this environment is a cornerstone of Selkirk's analysis. It provides her with another argument against Kahn's claim that ambisyllabicity is a precondition for flapping, because a prepausal consonant can never be ambisyllabic. It also allows her to state the flapping rule as an optional articulation of a simple released /t/ in syllable-final position. But even if her empirical observations were correct, Selkirk's analysis would still be in trouble. She would be unable to motivate when the [-release] insertion rule applies versus the flapping rule, as they have identical structural descriptions. In addition, Selkirk does not address directly the issue of the optionality of flapping.

The [release] feature is an attempt to capture the phonetically dynamic aspect of flaps and released stops in contrast to apparently static glottalized [tʔ], But Selkirk's analysis has encountered the problem typical of the SPE framework. The problem with Selkirk's use of release is analogous to the problem with [delayed release]. The [release] feature is an SPE-style feature that describes the entirety of the segment. She is unable to express with this feature that the [+release] or [-release] happens at the right edge of the segment, and this is the relevant condition of the flapping rule. In addition, she would have no means of describing similar interactions that we have seen at the left edge of the segment (if she were to acknowledge these effects.) The result is an analysis of flap that is not much more satisfying than previous analyses.

The problem that both the Kahn and Selkirk analysis have in common is that they both are unable to capture the generalization that the application of the English flapping rule is tied to the dynamic nature of the flap segment, as the segment is seen by the phonological rule that is responsible for its distribution. Both authors had a sense of the dynamic nature of the flap but were limited by the feature representation available to them. Kahn utilized the feature [sonorant] in an attempt to describe the change in articulation between the [t] and [ɾ]. Selkirk came closer to capturing the dynamic aspect of [ɾ] by introducing the feature [release], but by attributing [release] to the entirety of the segment, she is unable to express the interaction that is taking place just at its edges.

Based on the edge effects that we have observed in the case of flapping in American English, we will now turn to more current theories of phonological representation in order to capture these dynamic effects.

IV. AUTOSEGMENTAL ANALYSIS OF FLAP

In our examination of the flapping rule in English, we have seen that the left and right edges of the flap interact with their adjacent segmental environments. In this section, I will argue that the phonologically significant internal structure of the English flap is best represented as a contour segment.

IV.1. Justifying a Contour Segment.

The Autosegmental framework demonstrates, to a certain extent, the ability to describe the interaction at the edges of segments, as well as the ability to represent inherently dynamic segments as dynamic. As we have seen (in Section I), contour segments are dynamic segments that function as single timing units are complex segments and contour segments. The contour segments that we have seen so far have been binary branching in that they have consisted of a feature node that branches into two sequential values of that feature. The direction of the branching for the contour segments is in the temporal domain, as opposed to the logical-structure domain (as for complex segments), which implies the chronological ordering of the segments' values for the contour feature. Edge effects, as Sagey (1986) points out, provide evidence for the chronological ordering of the separate values in contour segments. Segments to the left of a contour segment will interact with the left edge value of the branching feature, and segments to the right will interact with the value at the right edge of the segment.

These are arguments for the relative ordering of subsegmental elements, not for the claim that these elements are subsegmental within a single segment. Evidence for the existence of a contour segment comes from the role it plays in the language (i.e., its function as a single timing unit: cf. Sagey,1986:69ff).

The alveolar flap in English most certainly functions as a single segment and also exhibits edge effects. These two factors would seen to qualify the flap as a contour segment. Let us consider what the feature geometry would look like for the flap, utilizing a Sageyan/Clementsian autosegmental approach. There are at least three factors to address in determining the structure of this segment: 1) at what node (feature) does the flap branch, 2) how many branches does the flap consist of, and 3) what are the specific values of the feature on each branch?

IV.2 Which Node Branches

In addressing the question of where the flap branches, let's review some contour segments that are firmly established within the autosegmental framework. Affricates are composed of a stop element, which is [-continuant], and a fricative element, which is [+continuant]. They branch at the feature [continuant], which in Sagey's model is attached to the root node. (In other versions of the feature tree, the manner features reside under a manner node, which is a branch of the PLACE/MANNER node; e.g., Clements, 1985. Pre- (and post-)nasalized segments consist of a nasal element and a non-nasal (oral) element. They branch at the soft palate node.

(30)

AFFRICATE PRENASALIZED

In these contour segments, the branching node is determined by the difference in identity of the subelements. In other words, the elements have a certain feature in common that changes its value from one element to the other. It is at that feature that the contour segment branches.

We have made progress with the autosegmental model in that we are now able to access the individual subsegments of affricates and prenasalized segments that in the SPE approach were totally inaccessible. However, the elements of the flap are not as readily identifiable because it is still regarded as a single articulation. The phonetic boundary between one element and the next is not as clear as it is for the affricates, for example. On a spectrogram of an affricate one can draw a definite boundary between the stop and the fricative elements. In both the affricate and the prenasalized cases, we have a fairly clear idea of what is going on articulatorily. For the affricate, a total stricture is made, then it is released to a fricative (which is usually also strident). For the prenasalized segments, the velum is down for the nasal element, and the velum is raised for the oral element. These articulations correlate nicely with the feature [continuant] and [nasal].

What are the subelements of the flap, and what feature changes value between them? The flap has not traditionally been treated as a segment that is composed of identifiable subelements. There has been no evidence of its internal structure brought to our attention previous to this. This may be partly due to the fact that, for the flap, the spectrographic and articulatory information alone do not reveal its internal phonological structure. A flap in a spectrogram looks like a somewhat shorter, sometime fricated stop. Articulatorily, there are different accounts of what actually constitutes the flap gesture: the more ballistic account (Ladefoged, 1982) and the Bernoulli effect account as offered by Kahn (1976:97). As either phonetic account seems to be compatible with the edge effects we have seen, we must conclude that the phonology does not care so much about whether there are delimitable sections in the segment; it cares only about the beginning and and ending characteristics of the flap. Consequently, we must look to the phonological evidence to identify the subsegments of the flap.

Upon reanalyzing the environments of the flapping rule in English, we hypothesized that at the beginning of the flap gesture, the tongue tip needs to be "cocked" to a position in which the tongue tip is free to gain enough momentum to flap against its target (or be sucked against its target by the Bernoulli effect). A segment with a degree of stricture greater than [-consonantal] would block the flap from the left. This suggests that the tongue tip is cocked back to the [-cons] degree. Thus, the leftmost subsegment of the flap has the value [-cons].

Similarly, the righthand environment of the flapping rule can lead us to the identity of the rightmost subsegment of the flap. We found that the tongue tip for the following segment needs to be relatively free to allow flapping. The syllabic [n̩] with coronal contact blocks flapping. The conclusion we can draw from this is that the tongue tip needs to retreat from the contact to a degree of aperture that is greater than [+consonantal]. This suggests that the rightmost subsegment of the flap also has the feature value of [-cons]. So far we have identified the left and right edges of the flap:

(31)

LEFT EDGE	RIGHT EDGE
[-consonantal]	[-consonantal]

The contour segments we have seen so far (Sagey, 1986) have branched at one terminal feature (or node). In Sagey's model, where the manner features are directly dominated by the root node, the flap contour would branch at the [consonantal] node under the root node. If we turn to a slightly different version of the segment structure tree (Clements, 1985), where the manner features are dominated by a manner node instead of the root node, we can have the flap branch at the manner node for the terminal feature [consonantal].

IV.3. How Many Branches

Now let us consider the issue of how many branches the flap has. Sagey does not explicitly put any constraints on the number of branches a contour segment can be comprised of. The contour segments for which she provides internal representations are all binary branching segments: affricates, prenasalized and postnasalized segments. The reason we have seen only binary branching segments so far is probably due to the nature of edge-effect phonological rules. If a contour segment branches three or more ways, the middle branch(es) will not interact directly with adjacent segment environments so it would be difficult, if not impossible, to justify the relative ordering of one or more middle branches, based on edge-effect phonological rules.

There is evidence, however, of contour segments having more than two branches. This evidence comes from contour nasal segments in Kaingang which exhibit edge effects. The significance of these effects were first pointed out by Anderson (1974, 1976). Kaingang has a set of partially nasalized segments that occur in the following environment:

(32)

a. n / Ṽ___Ṽ
b. n͡d / Ṽ___V
c. d͡n / V___Ṽ
d. d͡n͡d / V___V

The medionasal segment [dnd] appears to consist of three subsegmental elements. Sagey (1986:96) presents this Kaingang case as an example of contour segments in the autosegmental framework, but does not provide a representation of this particular segment. It follows from the other forms of nasal branching contour segments that this [dnd] would have the following ternary branching structure:

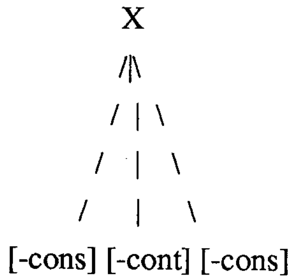
(33)

```
        SOFT PALATE
            /|\
           / | \
          /  |  \
         /   |   \
        /    |    \
[-nasal] [+nasal] [-nasal]
```

The middle element of the Kaingang contour segment [dnd] does not exhibit edge effects but it is still considered part of the contour segment. The point is that the three elements of this segment were not all derived from edge-effects rules. They were initially established by the phonetics of the segment. However, it seems unlikely, as pointed out to me be P. Keating, that a single timing unit could really include within it an entire velum gesture (i.e., raising, lowering and raising again.) We will see in the next section that the nasal values at the edges are probably the result of the realignment of the nasal values of the vowels. Regardless of their derivation, however, these medionasals are evidence of a contour segment with three branches.

In the case of the alveolar flap, there is no doubt that it functions as a single timing unit in English. The two branches of the flap that we have established on the basis of edge effects do not provide a complex phonetic description of the flap. The flap has been described articulatorily as a ballistic gesture that has an approach, contact and release. It has also been described in terms of the Bernoulli effect, where the tongue tip is held away from, drawn toward, then pushed away from the roof of the mouth. Though these descriptions are quite different from each other in terms of the cause of the gesture, they agree on the effect: that the flapping sound comes from the contact or near contact of the tongue tip against the alveolar ridge. The three essential ingredients of both descriptions are that: the tongue tip starts away from the alveolar ridge ([-cons]), the tongue tip strikes (or nearly strikes) the alveolar ridge, and the tongue tip retreats from the alveolar ridge ([-cons]). The edge values that we have established so far describe only the starting and ending gestures. Putting these two values adjacent to each other with no additional feature value to represent the contact portion of the flap does not accurately describe the phonetic essence of the flap. Therefore, I propose that this contour segment consists of three branches, the third being a middle branch, whose value is [-continuant] because this portion of the segment entails contact with the alveolar ridge, similar to a stop:

(34)



IV.4. The Values of the Branches

We have seen that the essential aspect of the alveolar flap is its changing degree of aperture from the beginning through to the end of the segment. In other words, the tongue tip is raised to some degree of aperture greater than its start position ([-cons]). The tongue tip usually makes contact becoming [-cont]. (It can also vary phonetically to a value of [+cont], as we will see in the next section.) Then the tongue tip retreats from its contact climax to some degree of aperture greater than its climax ([-cons]). The SPE features [consonantal] and [continuant] do not express this degree of aperture in a unified way. With this in mind, I propose one multi-valued degree of aperture feature to express the manner of articulation of this segment (and potentially all segments). A similar multi-valued phonetic scale was proposed by Ladefoged (1982:260): stop, fricative, approximant. Also, Browman and Goldstein's theory of Articulatory Phonology utilizes a constriction continuum which consists of seven discrete values which correspond to "categorical approximations" (Browman and Goldstein, 1987:3). In assigning integer values to the degrees, the direction of the values seems to be an arbitrary choice. Perhaps the most intuitive choice is to represent total stricture as the integer "0", i.e., no aperture.

(35)

APERTURE Value	SPE Feature Equivalent	Ladefoged's Scale
0	[-cont,+cons]	stop
1	[+cont,+cons]	fricative
2	[-cont,+cons]	approximant

Utilizing this single feature to encompass SPE manner features that describe aperture allows the flap to be presented as a contour segment that branches at a single terminal node, APERTURE, and also gives a transparent representation of the threshold phenomenon that we have seen at the edges of the flap.. What is not clear from the flapping rule is whether this Degree of APERTURE is an underlying feature or a derived phonetic feature. This issue is much too far-reaching to be decided here, and nothing in the current analysis hinges crucially on this new feature. In addition, it is not clear whether the Degree of Aperture feature should replace the entire manner node, or whether it resides as a terminal feature underneath the class node of MANNER, or whether it is an articulator-specific feature that resides under the active articulator node (e.g., CORONAL). The repercussions of such a proposal need to be considered carefully. We will conservatively assume for now that it resides under the MANNER node. Translating the flap's edge values into degree of APERTURE, we get:

(36)

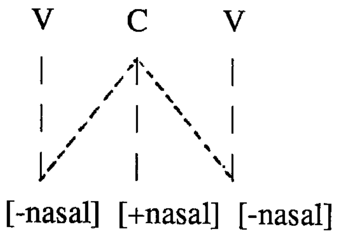
```
           APERTURE
              /|\
             / | \
            /  |  \
           /   |   \
          /    |    \
         2     0     2
    ([-cons] [-cont] [-cons])
  "cocked" "climax" "retreat"
```

V. DERIVATION

V.1. Spreading. We have evidence for edge effects in English flapping rule, and we have established the three branches of this phonetically and phonologically dynamic contour segment. The task at hand now is to write the phonological rule that governs the occurrence of this contour segment.

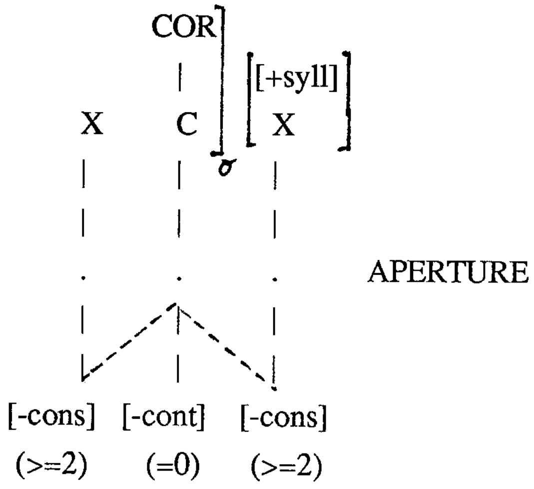
One thing to notice about the flap and its relationship with its segmental environment is that the edges of the flap have the same feature values as the respective adjacent segments. This situation is analogous to the medionasal segments in Kaingang as analyzed by Anderson (1974, 1976). According to Anderson, the medionasals ([dnd], bmb]) are variants of nasal stops between oral vowels. When a prenasalized stop occurs between two oral vowels, the result is a medionasalized stop, as in [V dnd V] versus [Ṽ nd V]. The edges of the consonant are the same value for [nasal] as the adjacent V's. Anderson (1976) suggests that the [-nasal] values from the vowels are extending into the stop segment. This insight, which was made before the development of current autosegmental notations, can be represented in multi-tiered representation as the spreading of independent tiers to adjacent segments, as:

(37)



The same spreading process can account for the edge values of the flap segment in English. Assuming that the underlying /t/ or /d/ has only one branch, in the appropriate syllable environment, certain values of degree of stricture spread onto the /t/. When the lefthand segment has a [-consonantal] (or APERTURE >= 2) value on its aperture tier, that value spreads to the aperture tier of the /t/, as shown below:

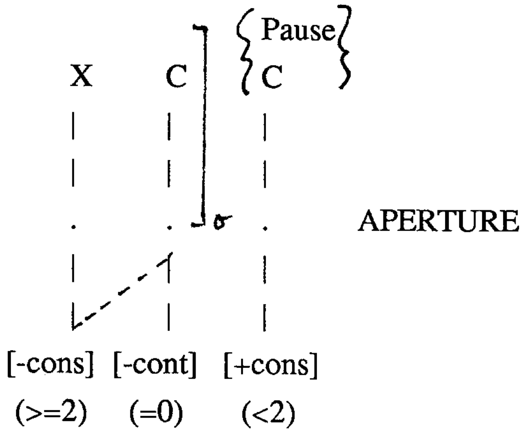
(38) FLAPPING AS SPREADING



Similarly, the stricture values from the righthand environment spread onto the /t/ but only if it is [-cons] or more (APERTURE>=2). The resulting structure is the three-way branching contour segment shown above; the two outer branches arise from the spreading rule. The middle branch is the original feature value of /t/ which remains intact. It is not changed in the flapping rule; it is just encroached upon by the adjacent segments.

Support for this derivation of the flap's contoured structure comes from a number of sources. The spreading analysis provides us with parallel derivation for the [tʔ] allophone of /t/ if, as Selkirk (1982) and traditional descriptions of unreleased [t] suggest, the [tʔ] is a stop without a release (plus glottalization). However, in this analysis, a release is viewed as a bit of [-cons] after the [-cont] of the stop, similar to the contour representation of an affricate. This analysis is a more temporally informative description of the release than Selkirk's [+release] feature. As we have seen,the major difference in the environments of [ɾ] and [tʔ] is a righthand segmental environment. The [tʔ] is realized when syllable-final and when followed by a [-syll,+cons] segment or pause. The lack of release can be seen as the lack of spreading of [-cons] (APERTURE = 2) from the right:

(39) UNRELEASED /t/ AS NO SPREADING FROM RIGHT



Glottalization seems to occur when the lefthand environment is [-cons] (APERTURE = 2), which is the value that spread from the right in the flapping rule, as well. It is tempting to say that the same lefthand spreading occurs with this allophone, but the facts are not as clear-cut as in the case of the flap. For instance, I am unable to detect any difference in the contact of the [l] in "halt" when I pronounce it with glottalized [tʔ] or released [t] ([haltʔ] vs. [halt]). However, for some of my consultants, contacting [l] implies released [t]. Based on the uncertainty of the facts, we cannot say with confidence that there is spreading of the lefthand [APERTURE] value, but the righthand spreading is quite convincing. So the spreading analysis gives us a partial derivation of the [tʔ] allophone. We will have to wait for further evidence in the lefthand spreading.

Flapping is often viewed as a casual or fast speech process. The spreading of feature values across segment boundaries is an intuitive way to represent this temporal readjustment of the

/t/ articulation which occurs with the increased tempo of the utterance. A logical extension of this spreading process is that the faster (or more casual) the speech is, the more extreme the spreading will be. If the non-contact value were to spread right through the stop, the result would be a fricated stop, or a true (but very short) fricative. Zue and Laferriere (1979:1043) report that flaps can often have a fricated realization. This phonetic variant of the flap (and thus of /t/) can be captured elegantly in this spreading model.

However, attributing the spreading process of flapping only to the temporal adjustment that occurs during casual speech does not account for why speakers of American English usually flap even in formal speech. In fact, flapping is quite difficult to suppress for most speakers, especially in more common words, like "butter". From these facts, we can see that although the spreading of feature values across segment boundaries may have originated as a low-level physiological effect of casual speech, in the case of the English flap, the process has become phonologized and applies more generally across all styles of speech.

There is evidence that flapping operates as a late (lower-level) phonological rule. This evidence comes from the types of segmental environments that condition the flap's appearance. These allophones are not the underlying (contrastive) segments. For example, /l/ is underlyingly [APERTURE = 0], but its extremely velarized allophone, [ɫ] has the value [APERTURE = 2]. The same can be said of the /n/, which is underlyingly [APERTURE = 0], but can surface as nasalization on the vowel, [APERTURE >=2]. It seems likely that these allophones are assigned their phonetic identity at a later stage in the derivation. The aperture values that spread to give the flap its structure are not present in the representation until later in the phonology, so the flap is not there until later.

There are also diachronic implications of the spreading of Degree of Aperture values. For instance, many examples of diachronic lenition processes, like intervocalic spirantization and affrication of stops in Germanic (Lehmann, 1973) can be modeled using spreading. English flapping is unusual in spreading a [-cons] degree of APERTURE rather than the more usual [+cont].

V.2 An Alternative Account. The perception or intuition that certain segments apparently BLOCK the insertion of the flap allophone is not being captured directly by the spreading analysis. As mentioned earlier, an approach and a retreat gesture are required to execute the flapping motion. This suggests that there is a well-formedness condition operating at each edge of the flap segment that determines its insertion. These well-formedness conditions, specific to the flap allophone is English, check to see if the adjacent environments are compatible in Degree of Aperture. If not, the flap is not inserted, and a default allophone [t] or [tʔ] is inserted. This suggests that the flap segment exists as a three-way branching contour segment in the inventory of English allophones..

An analysis such as this would allow us to capture the strong intuition shared by many, including Zue and Laferriere (1979:1046), that this blocking effect is what determines the flap's occurrence, and thus, that the dynamic aspect of the flap is inherent to the segment as opposed to being derived from the spreading of stricture values. It has been pointed out to be by B. Hayes that the total substitution of an allophone, as in the above scenario, is a very unnatural and unconstrainable process, and is therefore not a preferred analysis. The spreading analysis, where the flap allophone acquires its branching structure from its context, is preferred.

If the English flap is derived from its context, how then can we justify the existence of apparently autonomous flap gestures in numerous other languages where the flap is in environments that would not be able to contribute a [APERTURE = 2] value for spreading? Just one example of such a language is Ewe, where [ɾ] can be preceded by alveolar and palatal stops and affricates, whose right edges are both [+coronal] and [+consonantal] (Ladefoged, 1968:29) Another example is Russian, where [ɾ] (/r/) occurs in consonant clusters. Waveforms of words with [Cɾ] (/Cr/) sequences show the epenthesis of a short vowel-like segment before the [ɾ][3]. It appears that this ultra-short vowel serves as the "runway" or approach of the flap gesture. In other words, this element is not an inserted vowel per se, but is actually the cocking back of the tongue to the [APERTURE=2] position. The [APERTURE=2] value of this element cannot have spread from the left because the adjacent aperture value there is [APERTURE <=1]. This and the Ewe example above are evidence that the flap can have inherent internal structure. However, notice that in both of these languages, the flap is contrastive with it, and is not subject to the same phonotactic constraints as the allophonic English flap. The Russian flap has no apparent restrictions on the [Degree of Aperture] values it must be adjacent to; this appears to be the case with Ewe as well. This contrastive flap comes with three underlying branches, which introduce extra articulations in some cases (e.g., Russian).[4]

The generalization we can draw from the cross-linguistic behavior of flaps is that there are at least three levels of processes to be distinguished. Starting with the most abstract, we have seen that when a flap is contrastive in a language, as in Ewe and Russian, it must have underlying three-branching structure at the APERTURE node. In the case of the contextual allophonic flap, as in English, phonological spreading of certain [Degrees of Aperture] occurs to create the multiple branching of this allophone of /t/. A third lower level phonetic spreading process (an extension of the above spreading process) can be distinguished in the case where the flap can surface as fricated. What we have seen is that the internal structure of the flap can be manipulated by the phonology at different levels of the grammar and an extension of the existing autosegmental theory can represent these distinctions.

VI. ANOTHER EXAMPLE OF INTERNAL STRUCTURE - SIERRA POPOLUCA

VI.1. Introduction. I offer in this section another example of a class of segments that have phonologically significant internal structure in a particular language. The language is Sierra Popoluca, a Zoquean language. The data below comes from Elson (1947, 1956). Clements (1985) uses the patterns below to demonstrate an effect of rule interaction is his feature geometry. I will argue, however, that he unknowingly discovers a phonological rule that refers to the internal structure of the class of stop segments.

VI.2. The Data. There are two types of transitional elements between consonants in Sierra Popoluca: aspiration between heterorganic voiceless stops, and a "lenis schwa" between nasals and "certain other sounds". There is no audible transition between homorganic consonants. Note that aspiration is not distinctive, as we saw in Sagey's analysis of Sierra Popoluca (cf. Section I.)

(40)

[kɛkʰ.paʔ]	/kɛk.pa/	"it flies"
[miñᵊ.paʔ]	/miñ.pa/	"he comes"
Cf.		
[kɛk.gak.paʔ]	/kɛk.gak.pa/	"it flies again"
[ʔaŋ.kiʔ]	/aŋ.kiʔ/	"yard"

Clements describes the complementary distribution of [ʰ] and [ᵊ] as a floating release feature that is inserted between two stops:

(41)

0 ---> [+cont] / [-cont] ____ [-cont]

This release feature, according to Clements, contrasts with "absence of oral release" (1985:239), meaning that it does not occur adjacent to a sound that is characterized by oral airflow, in Clements' terms, and it picks up its voicing value from "speech physiology" (1985:240). Clements calls upon Steriade's (1982) Shared Features Convention to condition the application of this rule. This proposal states that:

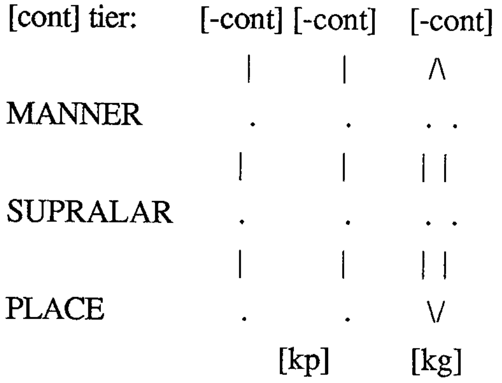
(42)

> "If two root nodes should come to dominate a single feature as the result of a rule, then any other identical features that they dominate are immediately merged into one." (Clements, 1985:240).

Clements' analysis of the Sierra Popoluca data includes an earlier rule that merges identical place nodes to account for a nasal assimilation rule (Clements, 1985:238). Then the Shared

Features Convention links the remaining identical nodes, so that the representation of the heterorganic sequence [kp] as compared to the homorganic sequence [kg] would look like this.:

(43)



Hence, the merged manner tier blocks the insertion of the floating [+cont] feature.

However, a closer look at the data (Elson, 1947) reveals that the lenis schwa is "inserted" between nasals and these "certain other sounds":

(44)

np nk ng nm

ñp ñk ñg ñm

ŋp ŋt ŋt^{y} ŋc ŋč ŋs ŋš ŋm ŋy

(Elson, 1947:16)

Notice that not all of the second sounds in the above sequences are [-cont], as Clements claims in order to write his rule (cf. (43)). The segments /s/, /š/ and /y/ are [+cont][5] . Also, Elson's description of where the "open transition" occurs after oral stops is : "when not followed by a consonant phoneme of the same point of articulation" (Elson, 1947:16). Although the only examples he happens to give are ones where the second segment is a stop, he does not limit the second consonant to this class of segments. It would seem that Clements' analysis handles only a subset of the facts.

VI.3. The Reanalysis. I would like to suggest a different analysis that generalizes across a wider range of facts in Sierra Popoluca in a more intuitive way than the analysis proposed by Clements. We saw in Sagey's analysis of Sierra Popoluca that syllable-final stops exhibit aspiration, while affricates and fricatives do not (See Figure (9)). In addition, Elson observes that nasals have a "lenis voiceless off-glide" (Elson, 1947:15) in utterance-final position.

(45)

/ʔi.’hon/	[ʔi.’hon^{n}]	"his bird"
/ka.mam/	[ka.mam^{m}]	"hard"
/ca:ɲ/	[caɲ^{ɲ}]	"snake"

If we combine all these observations, a pattern emerges:

(46)

a. syllable-final oral stops are aspirated
b. utterance-final nasal stops have lenis voiceless off-glide
c. aspiration between oral stops and heterorganic stops, fricatives, affricates
d. lenis schwa between nasal stops and heterorganic stops, fricative, affricates

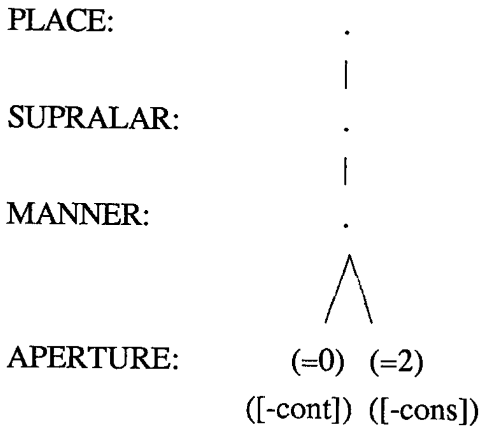
The generalization that we can draw from these additional observations is that the class of stops [-cont] (0 APERTURE), both oral and nasal, have a release element that is phonologically significant in syllable-final position. This release subsegment is audibly realized when a heterorganic stop or fricative follows, and word-finally.[6] The release feature is "deleted" when followed by a homorganic stop. The releases of stops occurring before sounds characterized by oral airflow are masked by the airflow of the following segment.

A plausible articulatory explanation for this comes to mind immediately. When two stops of different places of articulation are adjacent, the articulator releases from the first stop in order to prepare for the different articulation of the second consonant. The moment between the different degrees or places of contact is audible. The release is realized as voiced or voiceless depending on the voicing value of the segment as a whole.

When homorganic stops are adjacent, the reason there is no audible release of the first stop is that for ease of articulation, the contact (or near contact, for fricatives) is maintained throughout the second segment. This can be seem as a low-level (phonological) rule of coarticulation, where rather than releasing contact with an articulator only to make immediate contact (or near contact) in the same place, the speech mechanism conserves effort and the articulator stays put. This articulatory account might have been the original motivation of this rule, but now the process is phonologized.[7] Recognizing this release feature as an inherent part of the stop segment's internal structure allows us to explain the distribution of this release (and its allophones) simply and naturally.

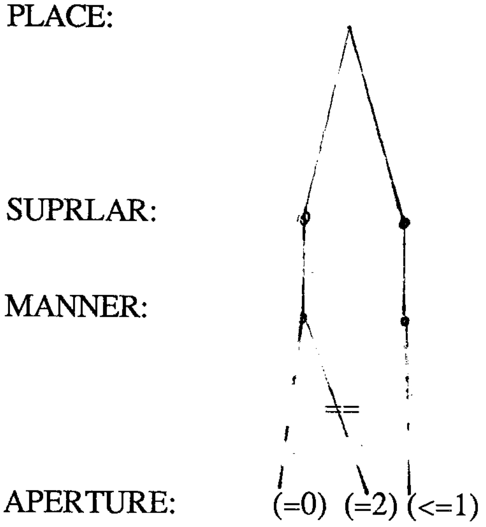
What would the representation of this release feature and its corresponding deletion rule look like? The release of a stop necessarily follows its oral occlusion. In the case of Sierra Popoluca, at least, this release exhibits edge effects similar to those seen for affricates. As we have seen, affricates have been described as a stop element followed by a fricative element. One possible representation for these released stops, then, is to have [-cons] (APERTURE = 2) at the right edge, like an affricate, but with a greater degree of opening:

(47) RELEASED STOP



The release deletion rule requires that the adjacent segments have the same place of articulation. If we assume that the place nodes are linked by the earlier rule of place assimilation that, according to Clements, allows an earlier nasal assimilation rule, we can state the release deletion rule as:

(48) RELEASE DELETION RULE



In other words, delete the release (AP=2) branch of the APERTURE tier if it is followed by an APERTURE value of 1 or greater.

This reanalysis removes Sagey's argument that syllable-final stops in Sierra Popoluca become aspirated (cf. Section I). With the observation that nasal stops also exhibit an aspiration-like release, this phenomenon can be reanalyzed as the voiceless release of the syllable-final stops (nasal and oral).[8] By positing this internal structure of stops in Sierra Popoluca, we are able to capture the pattern of appearance of syllable-final "open transitions" simply and generally. Consider the question of whether the internal structure of stops in Sierra Popoluca is derived from the environment or whether it is present underlyingly. There is no possibility that the release portion of the stop could have come about from spreading because it occurs when its righthand environment is [+cons] (AP<=1), i.e., there is not an adjacent feature value that could have spread to give the feature value of the release (AP=2). Therefore, the branching structure of stops in Sierra Popoluca is probably underlying. This structure is phonologically redundant in the sense that the released and unreleased stops are not contrastive, but the elements are manipulated by the phonology, so they must be present in the representation.

VII. CONCLUSIONS

We have seen that certain segments (flaps in English and stops in Sierra Popoluca) are phonologically dynamic, i.e., their internal structure is relevant in phonological processes. We have posited internal structure of these two segment types that had previously been treated as temporally unanalyzable. In order to adequately represent this internal structure, we have taken the notion of contour segment in the autosegmental framework and extended it to include a three-way branching segment, the flap, and a two-way branching segment type, stops. We have shown that the flap's structure can be derived from (coronal) stops by the spreading of manner features of neighboring segments, improving on previous SPE-style analyses. We have also suggested that the branching structure of stops can exist at underlying representation. We have shown that either the existing manner features, or a newly proposed APERTURE feature, can represent aspects of dynamic manners of articulation of consonants that have previously been described with the ad hoc features [flap] and [release].

VIII. REFERENCES

Anderson, S.R., 1974. Organization of Phonology. Academic Press.

Anderson, S.R., 1976. Nasal Consonants and the Internal Structure of Segments. Language 52, 326-344.

Banner (Inouye), S., 1987. Casual Speech Rules and Tree Geometry. Ms. UCLA.

Browman, C.P. and Goldstein, L., 1987. Tiers in Articulatory Phonology, with some implementations for Casual Speech. First OSU Conference on Laboratory Phonology, 5-7 June, 1987 (DRAFT).

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

Clements, G.N., 1985. The Geometry of Features. In C.J.Ewen & J.M. Anderson(eds.) Phonology Yearbook 2. Cambridge University Press.

Clements, G.N., 1987. Phonological Feature Representation and the Description of Intrusive Stops. In A. Bosch et al (eds.) Parasession on Autosegmental and Metrical Phonology.

Elson, B. 1947. Sierra Popoluca syllable structure. IJAL 13, 13-17.

Elson, B., 1960. Gramatica del Popoluca de la Sierra. Universidad Veracruzana. Xalapa, Mexico.

Goldsmith, J., 1976. Autosegmental Phonology. Indiana University Linguistics Club.

Hyman, L., 1975. Phonology: Theory and Analysis. Holt, Reinhart and Winston, Inc.

Inouye, S.B., 1987. Articulatory Evolution: A Closer Look. Ms. UCLA

Kahn, D., 1976. Syllable-Based Generalizations in English Phonology. Indiana University Linguistics Club.

Keating, P.A., 1984. Phonetics and Phonological Representation of Consonant Voicing. Language, Vol. 60, No. 2, 286-319.

Keating, P.A., 1985. CV Phonology, Experimental Phonetics, and Coarticulation. Revised Version of paper presented at "Colloque Phonologie Pluri-Lineaire", June, 1985.

Keating, P.A., 1985. Universal Phonetics and the Organization of Grammar. In V.A. Fromkin (ed.) Phonetic Linguistics. Academic Press.

Ladefoged, P., 1968. A Phonetic Study of West African Languages. Cambridge University Press.

Ladefoged, P., 1982. A Course in Phonetics. Harcourt, Brace, Jovanovich, 2nd. ed.

Lehmann, W., 1973. Historical Linguistics, An Introduction. Holt, Reinhart and Winston. 2nd. ed.

Maddieson, I., 1984. Patterns of Sounds. Cambridge University Press.

Oshika, B., Zue, V., Weeks, R., Neu, H. & Arbach, J., 1975. The Role of Phonological Rules in Speech Understanding Research. IEEE Trans. Acous., Speech & Signal Proc. ASSP-23, No. 1.

Port, R., 1986. Translating Linguistic Symbols into Time. Research in Phonetics and Computational Linguistics, No.5, July, 1986. Indiana University Working Papers.

Sagey, E., 1986. The representation of Features and Relations in Non-Linear Phonology. MIT Dissertation.

Selkirk, E.O., 1982. The Syllable. In H. Van der Hulst & N. Smith. The Structure of Phonological Representations, Part II. Foris Publications.

Steriade, D., 1987. Gestures and Autosegments: Comments on Browman and Goldstein's "Gestures in Articulatory Phonology". Ms.

Stevens, K.N., 1985. Evidence of the Role of Acoustic Boundaries in the Perception of Speech Sounds. In V.A. Fromkin (ed.). Phonetic Linguistics. Academic Press.

Zue, V.W. & Laferriere, M. 1979. Acoustic Study of Medial /t,d/ in American English. J. Acoust. Soc. Am. 66(4), Oct. 1979.

---

[1] We will see in a later section that these particular data do not justify her claim that affricates are contour segments, but there are many other examples that do. (cf. Sagey 1986: 92).

[2] The [ʔ] in [ʔɩt] occurs by the application of the English Glottal Stop Epenthesis rule that inserts a glottal stop before a vowel that is preceded by a Pause.

[3] B. Hayes, unpublished acoustic data.

[4] It is interesting to note that the Russian [ɾ] alternates with [r]. How does flap derive from trill? I will address this question in a later paper.

[5] Elson does not give a phonetic description of /y/, but it seems possible that it is actually a palatal fricative corresponding to /t^y/ /d^y/. This would explain its being among these other fricatives in this rule.

[6] Voiced oral stops /b, d, d g/ do not occur syllable-finally, seemingly the result of a general constraint for final voicelessness, showing up in the voiceless release of final nasal.

[7] A possible reason that we see a release of the /t/ before /s, š and y/ which have oral airflow, is that the degree of aperture to which the articulators are released is greater than the aperture of fricatives.

[8] This reanalysis of Sierra Popoluca will not adversely affect Sagey's argument for the ordering of subelements in contour segments, as Sierra Popoluca was just one of numerous examples of contour segments and their internal ordering. For more examples, see Sagey (1986:92ff).

Methodological studies using an x-ray microbeam system

Mona Lindau-Webb & Peter Ladefoged

Introduction

The x-ray microbeam technique allows investigators to track the movements of small gold pellets placed on the articulators (Abbs et al. 1988). It provides an excellent way of obtaining data on articulatory shapes and movements in speech. There are, however, severe limitations to the current system. It provides data from only a few points, which must be spaced at least 1 cm apart if rapid movements of speech are to be accurately recorded. In addition, it has so far not been possible to place pellets on the pharyngeal part of the tongue. Accordingly, some preliminary methodological studies are needed before the system can be used most efficiently for studying the dynamics of speech. In this study we attempt to answer the question: As only a limited set of locations are available, what are the best places for a small number of pellets on the tongue when trying to predict the whole tongue shape?

Method

It is possible to use the x-ray microbeam system to record single image scans of a subject. In order to get data from the outline of the midline of the whole tongue we used a string consisting of 15 gold pellets with their centers about 7 mm apart, joined by surgical thread, lying on the tongue of a speaker producing 11 sustained English vowels. The subject was a male speaker of Midwestern American English with no dental fillings. The sustained vowels were recorded in utterances of the form "bee, bee, beeeeee" with the subject prolonging the vowel in the third word. The string of pellets was fixed by glue at only one point in the middle, and was free to conform to the contour of the tongue. The most anterior pellet recorded was near the tip of the tongue, and the lowest pellet was usually in the epiglottal pocket. The pellets adhered to the tongue down into the epiglottal pocket, thus giving a good, reliable outline of the tongue shape for all the vowels. Additional pellets were affixed to the upper and lower lip, to the front point of the maxilla, just above the front teeth, to the front part of the mandible, just at the lower front teeth, as well as a reference pellet on the bridge of the nose. An example of data recorded this way is shown in figure 1. On this figure the pellets of the gold chain have been highlighted , while the other pellets are left unprocessed.

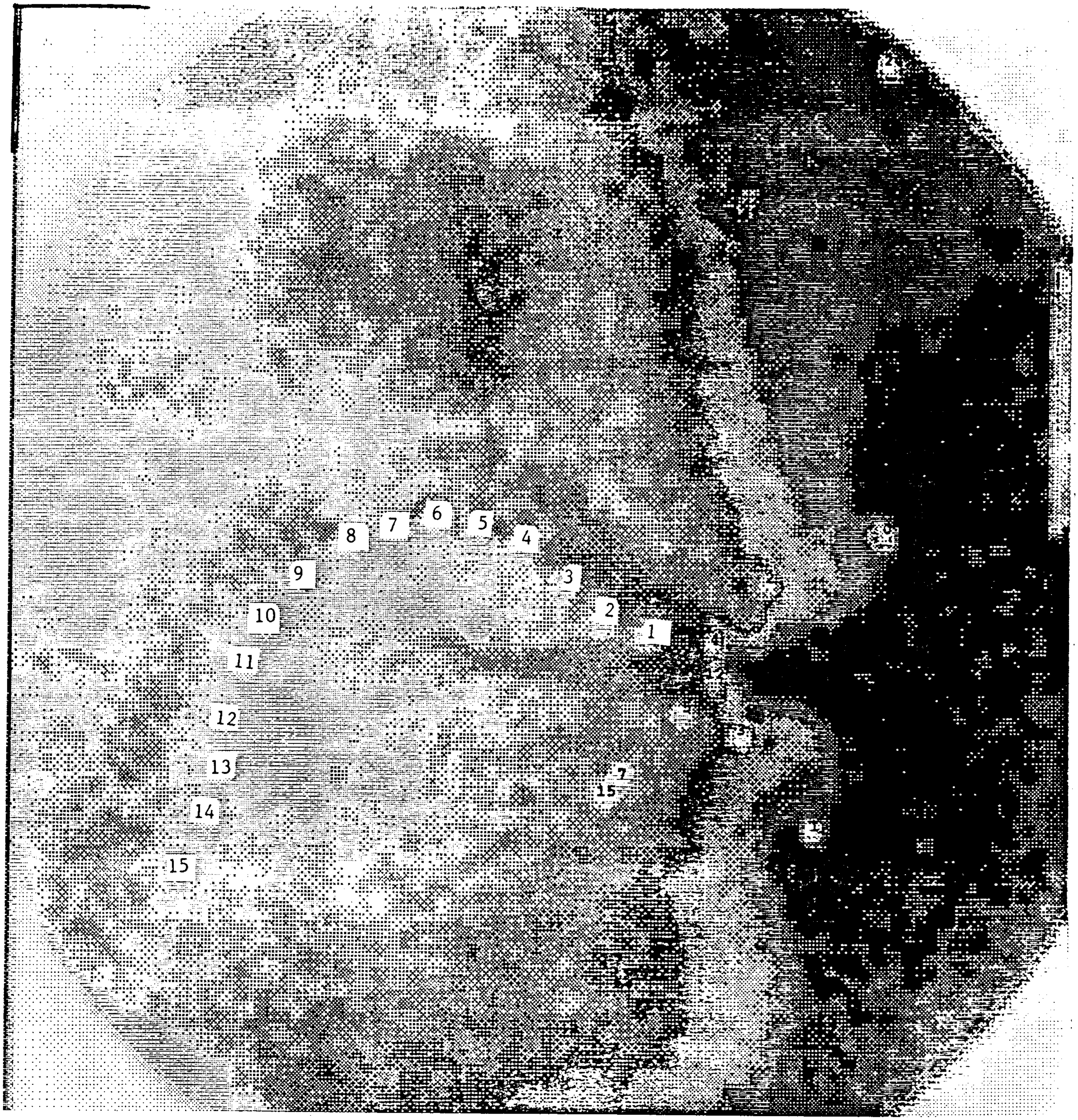


Figure 1. Image scan of a speaker with a chain of gold pellets on the midline of the tongue during the production of the vowel in "bib" with the points 1-15 marked and highlighted. Actual unprocessed image points can be seen for example on the lips.

A simultaneous audio recording was made, and the acoustic vowel space of this speaker's sustained vowels is plotted in figure 2. The same speaker's vowel space from vowels said in short utterances of the form "Say bib between" is plotted in figure 3. Although the position of the vowel in "baw" differs between the two sets of vowel spaces, there is a high correlation ($r=.927$, df 9, $p<0.0001$) between the two sets of formant frequencies, showing that the sustained vowels are very similar to those said in utterances.

Analysis and results

As a means of data reduction the x and y positions of each of the 15 pellet positions for each of the 11 vowels were subjected to a two-mode factor analysis with varimax rotation to determine what factors may underlie the variation in tongue shapes in the different vowels. A three-factor solution converged after 5 iterations, accounting for 97.21 % of the standardized variance in the data. Figure 4 shows the tongue point loadings of each factor plotted as deviations from a mean tongue shape. In the figure of each factor the black dots signify the mean tongue shape, and the open squares and triangels signify the deviations from the mean that can be attributed to that factor. Figure 5 shows plots of the vowel loadings, with the vowel loadings of factor 1 and factor 2 plotted against each other, and those of factor 3 below. Although the interpretation of these factors from a single speaker is not very meaningful, it is interesting to note that the tongue point loadings of factor 1 look similar to those of the Front Raising factor found for the American English speakers by Harshman et al. (1977) with a seesaw motion around the "hump" of the tongue. The factor space formed by plotting the vowel loadings of factor 1 and factor 2 against each other in this way is reminiscent of the conventional vowel space. Factor 1 aids in separating the front from the back vowels, and the tongue root movement of factor 2 contributes to varying vowel height. Factor 3 below is a kind of bunching that contributes to distinguishing between tense and lax vowels.

The question of which of the 15 points, or which combination of 2 or 3 points, is the best predictor of the remaining points can be reformulated as questions of regression. We can determine how successful a given point would be in predicting the tongue shape by taking the values for each vowel for that point and regressing them against the factor values for these vowels. Similarly, we can use multiple regression to determine which combinations of 2 or 3 points are the best predictors of each factor. We correlated each factor with all possible combinations of the positions of 2, and 3 pellets by running multiple regressions with the vowel loadings as the dependent variable, and the x and y values of tongue pellet positions of the 11 vowels as independent variables.

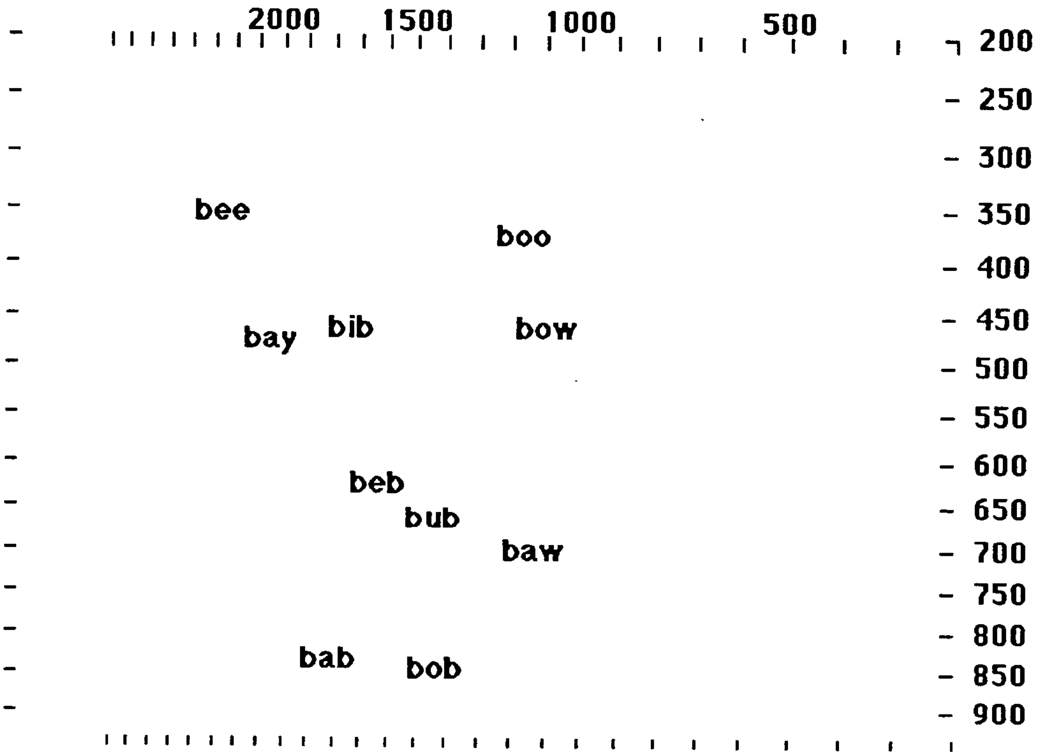


Figure 2. Vowel space of the speaker's vowel space of sustained vowels.

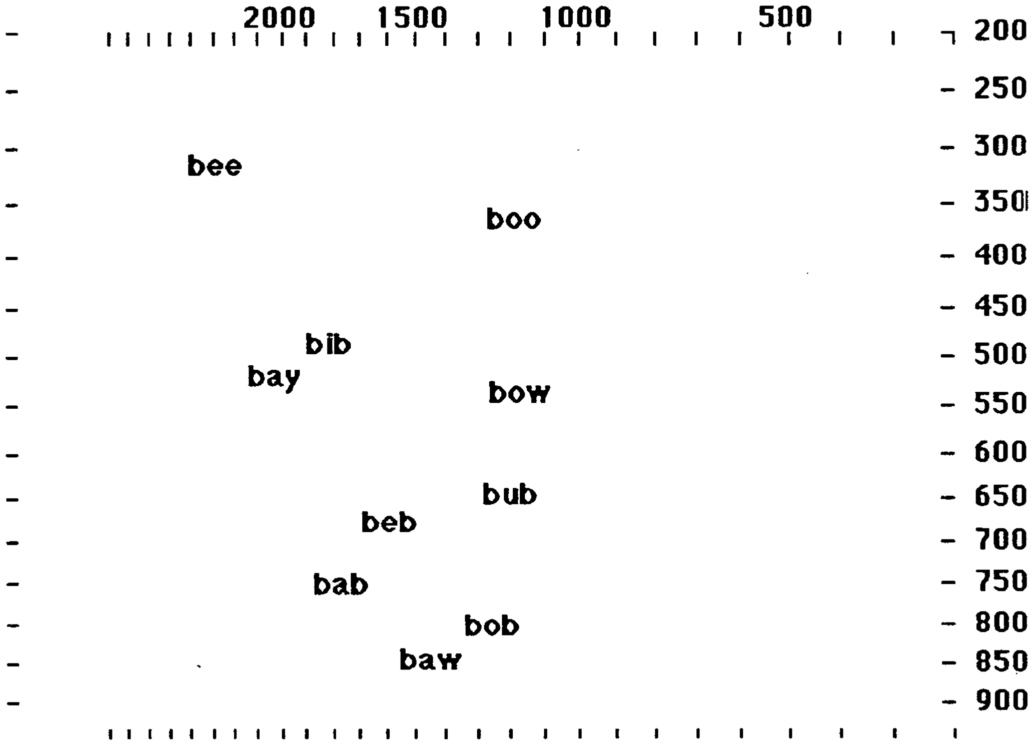


Figure 3. Vowel space of the speaker's vowel space of vowels from short utterances.

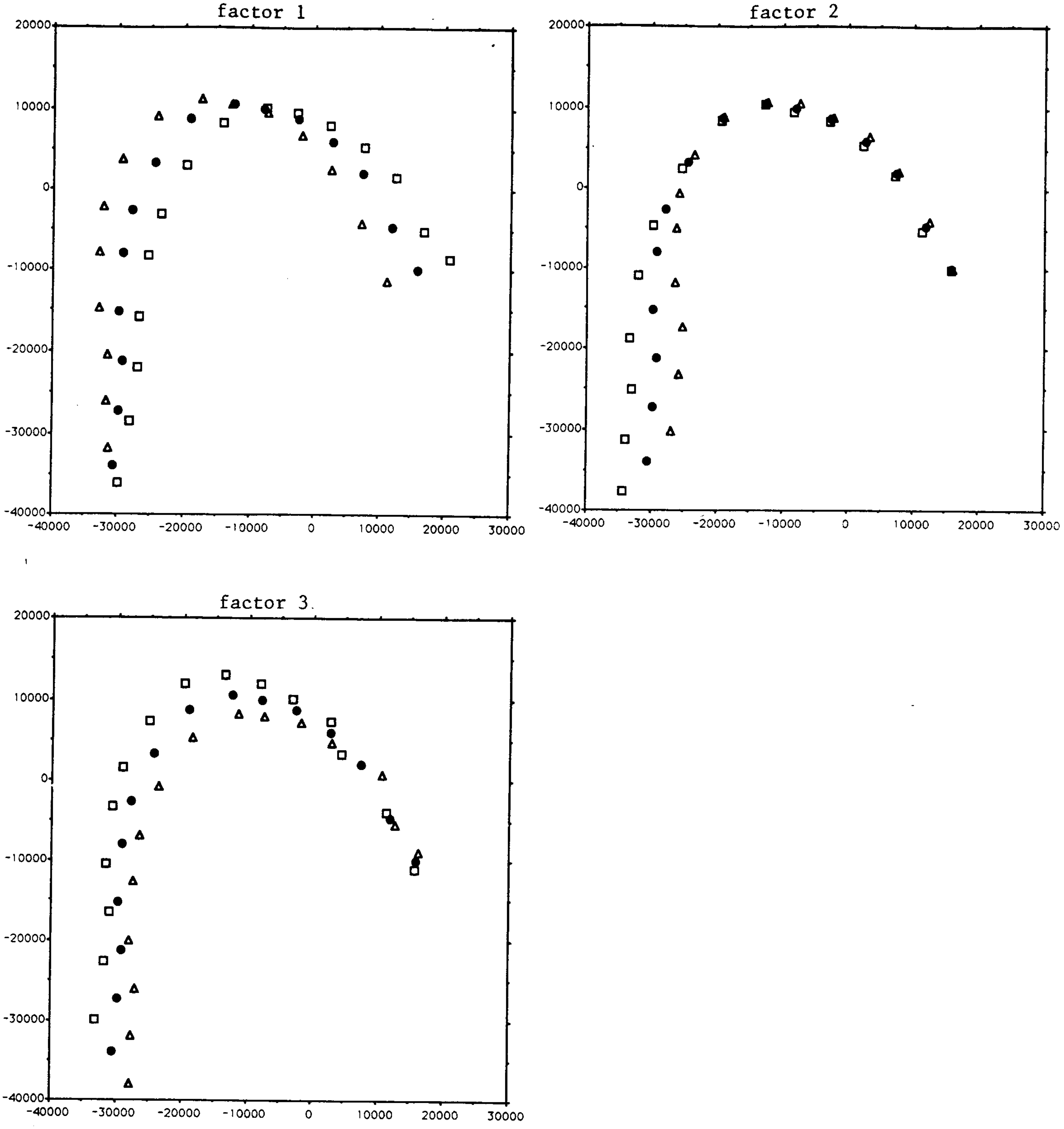


Figure 4. Tongue point loadings of three factors plotted as deviations from a mean tongue shape.

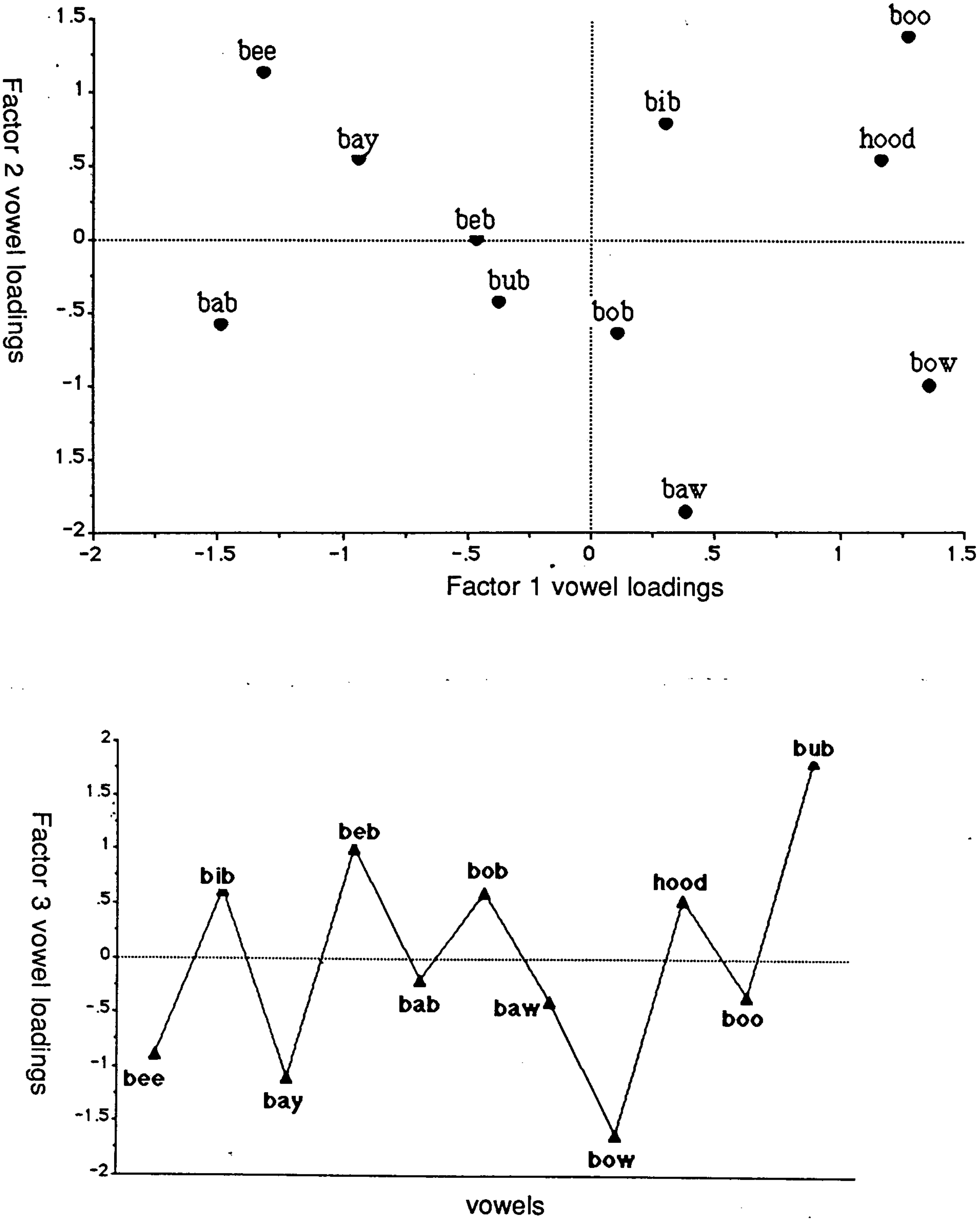


Figure 5. Plots of vowel loadings of the three factors. The vowel loadings of factor 1 is plotted against those of factor 2. The vowel loadings of factor 3 is plotted below.

The results enabled us to determine the best locations for placing pellets, and the extent to which tongue shapes can be determined from a small number of pellets, even without placing pellets at the root of the tongue. If only two pellets are used, there are 5 combinations that resulted in high correlations with the observed shapes (r^2 = >.98), as specified by three factors which themselves determine the tongue shapes within 0.5 mm. The 5 two pellet combinations are shown in figure 6. This figure shows these 5 two pellet combinations plotted on the tongue of a midsagittal section of the vocal tract. The two top combinations have both pellets located above the root of the tongue. Thus it is at least in theory possible to recover the tongue shape from only two pellets, placed where the current microbeam recording system currently allows pellet placement. It is interesting to note that in the top left two pellet combination, the two pellets may be placed very closely together, provided the two pellets are located at the "hump" of the tongue, i.e. the hinge point of factor 1 in figure 4.

Figure 7 shows another way of plotting these 5 two pellet combinations with the pellet numbers along the x-axis. The pellets are located at 7 mm distance from each other. Thus pellet no. 2 is located about 14 mm from the tip of the tongue, pellet no 5 about 35 mm from the tip of the tongue, etc. The root of the tongue is located around pellet no. 11, about 77 mm from the tip of the tongue.

Similar very high correlations (r^2 = >.99) can be achieved with 74 different combinations of three pellets. Most (75%) of these combinations involve one or two points at the root of the tongue. However, there are 19 three pellet combinations, where all three pellets may be placed above the root of the tongue, at 77 mm or less from the tip of the tongue. These 19 combinations are shown in figure 8. There appear to be three locations that are favored in these three pellet combinations above the tongue root, namely pellets no. 2-3, i.e. 14-21 mm from the tip of the tongue, pellets no. 6-7, i.e. 42-49 mm from the tip of the tongue, and pellets no 8-9, i.e. 56-63 mm from the tip of the tongue.

These results are most encouraging. Particularly noteworthy is the fact that we can accurately reconstruct the tongue shape from the tip to just above the root of the epiglottis just from two pellets, placed on the tongue on either of the positions shown in figure 6. Two warnings are, however, necessary: firstly the predictions apply only to vowels, and non-retroflex vowels at that; secondly they are based on the analysis on a single subject. We are currently working on the analyses of data from other subjects, and will report on this soon.

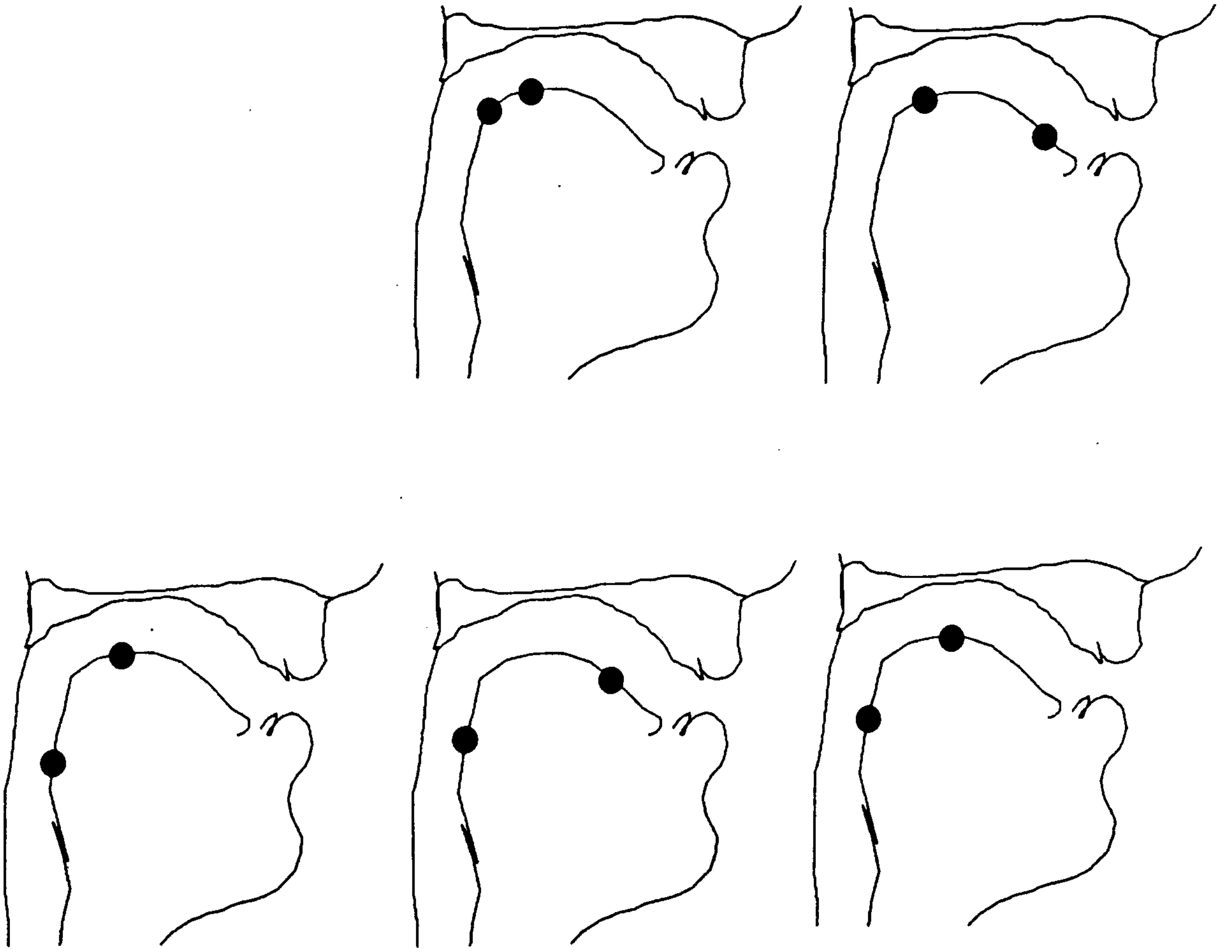


Figure 6. Two pellet combinations with high correlations to observed tongue shapes ($r^2 = .98$), plotted on the tongue. The two top combinations have both pellets located above the root of the tongue.

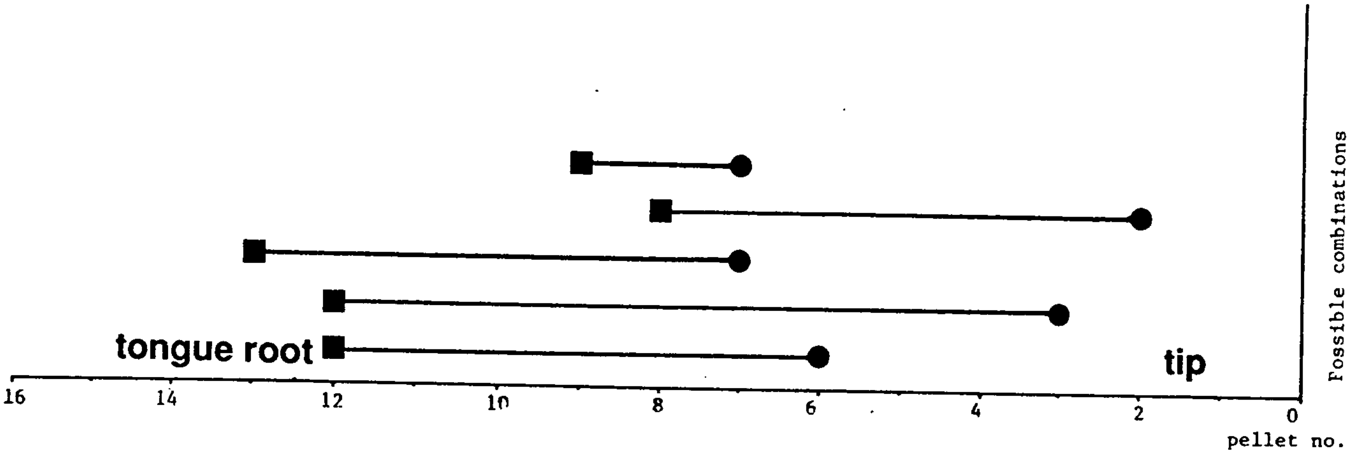


Figure 7. The two pellet combinations from figure 6 plotted as a display of possible combinations with the pellet number on the horizontal axis. The tongue-root is considered to be located below pellet # 11.

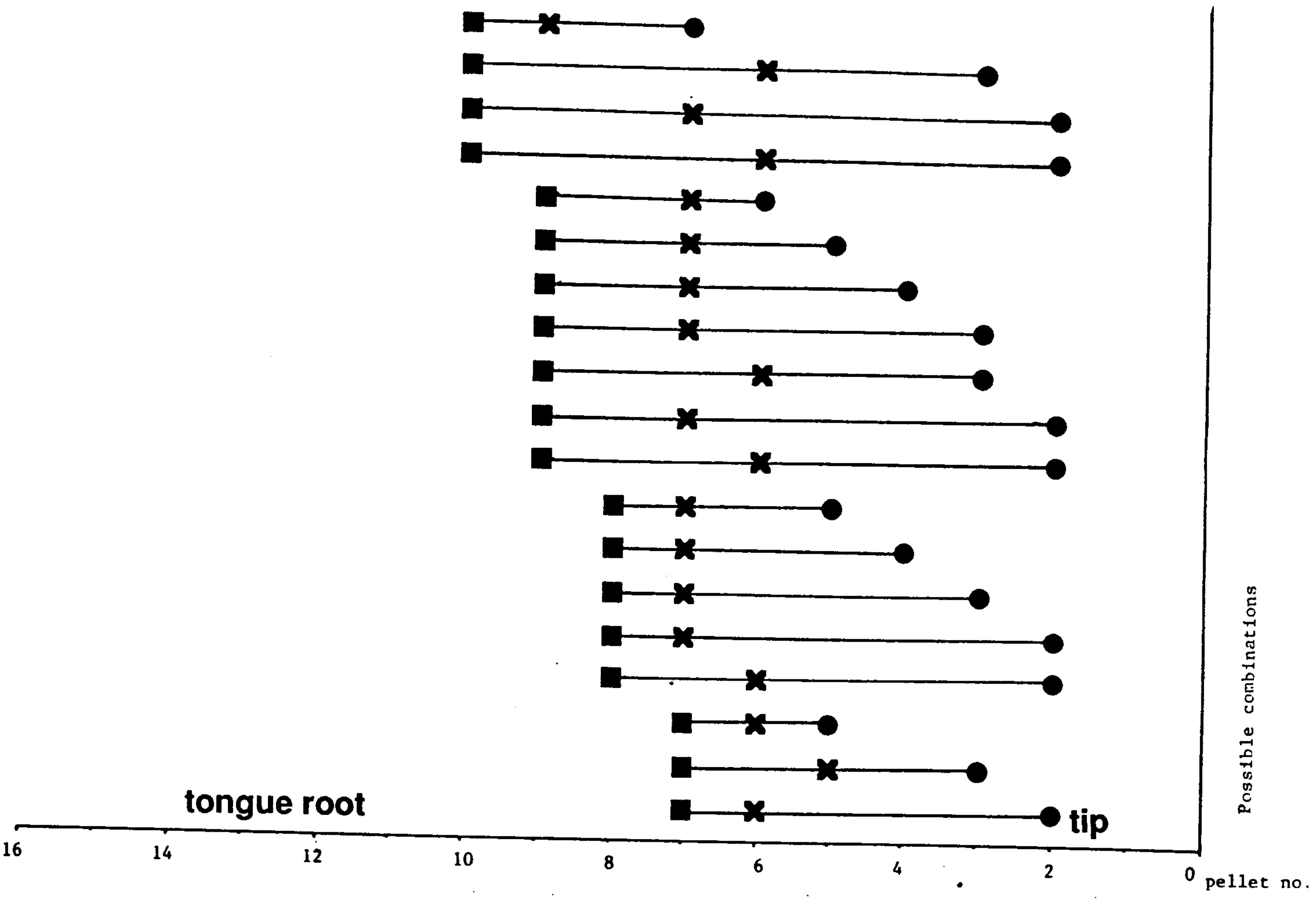


Figure 8. The 19 out of 74 three pellet combinations with all three pellets above the root of the tongue, and a high correlation to observed tongue shapes ($r^2 = .99$) .

References:

Abbs, J.H., R.D.Nadler, & O.Fujimura (1988). "X-ray microbeams track the shape of speech." **Soma**, January 1988: 29-34.

Harshman, R. P. Ladefoged, and L. Goldstein (1977) "Factor analysis of tongue shapes." **Journal of the Acoustical Society of America** 62:693-707.

Aerodynamic constraints on sound change: The case of bilabial trills

Ian Maddieson

Introduction

Trilling of the lips is not in itself an unusual human action. It is observed quite frequently in infant vocalizations; older children and adults trill the lips to imitate the sound of an internal combustion engine; in many cultures a labial trill has a recognized communicative value (like the "raspberry" or "Bronx cheer" familiar in the U.S. as well as the lower-frequency trill, often written "brrr," signifying that the maker is feeling cold); and many musical traditions around the world make use of horn instruments, such as the bugle, in which the vibration of the lips is the source of the instrument's power. Yet this quite familiar gesture is very rarely encountered as a linguistic sound. In fact bilabial trills are so unfamilar as regular speech sounds that many standard phonetic textbooks are silent about their occurrence in language (Abercrombie 1967, Catford 1977) and some even overtly state that "the lips can be made to roll [i.e. trill] .... but this is not found as a regular sound in language" (O'Connor 1973).

In fact, bilabial trills do occur as part of the normal repertoire of sounds in a number of languages. There are principally two groups of languages concerned. One group consists of a small number of Bantoid languages spoken in the Grassfields area of Cameroun in West Africa, including Ngwe (Dunstan 1964), Babanki (Hyman 1980), and some dialects of Kom (Hyman p.c.). The other group contains a number of rather widely scattered Austronesian languages. These include two clusters of languages within the large Oceanic subgrouping of Austronesian languages, one spoken in Manus Province of Papua New Guinea, and the other on the island of Malekula in Vanuatu. The remainder are a few Western Malayo-Polynesian languages spoken in different parts of Indonesia, such as the Nias language spoken on the islands of Nias and Batu off the west coast of Sumatra (Catford 1988a), and Muna, spoken on the island of the same name in S.E. Sulawesi (Nico van den Berg, p.c.). The Manus languages with bilabial trills are listed by Blust (1986) as consisting of the following: Ahus (Hus), Sora, Nali, Papitalai, Lele (Akara), Kele, Titan, Kurti (Kuruti), and Leipon. Ross (1988) names this group the "East Manus network" and adds the names Ponam, Andra, Koro, and Ere, but does not mention Sora and Papitalai. The Malekula languages with bilabial trills include several from the South West Bay area, such as Windua (Ninde) and Na?ahai, and the Uripiv dialect of the Atchin-Wala-Rano language from the northeast of the island. Bilabial trills are not specifically noted in the surveys of Vanuatu languages conducted by

Tryon (1976) and Charpentier (1982) so it is uncertain if there are additional languages with this phenomenon. Locations of the languages named are given on the accompanying map.[1]

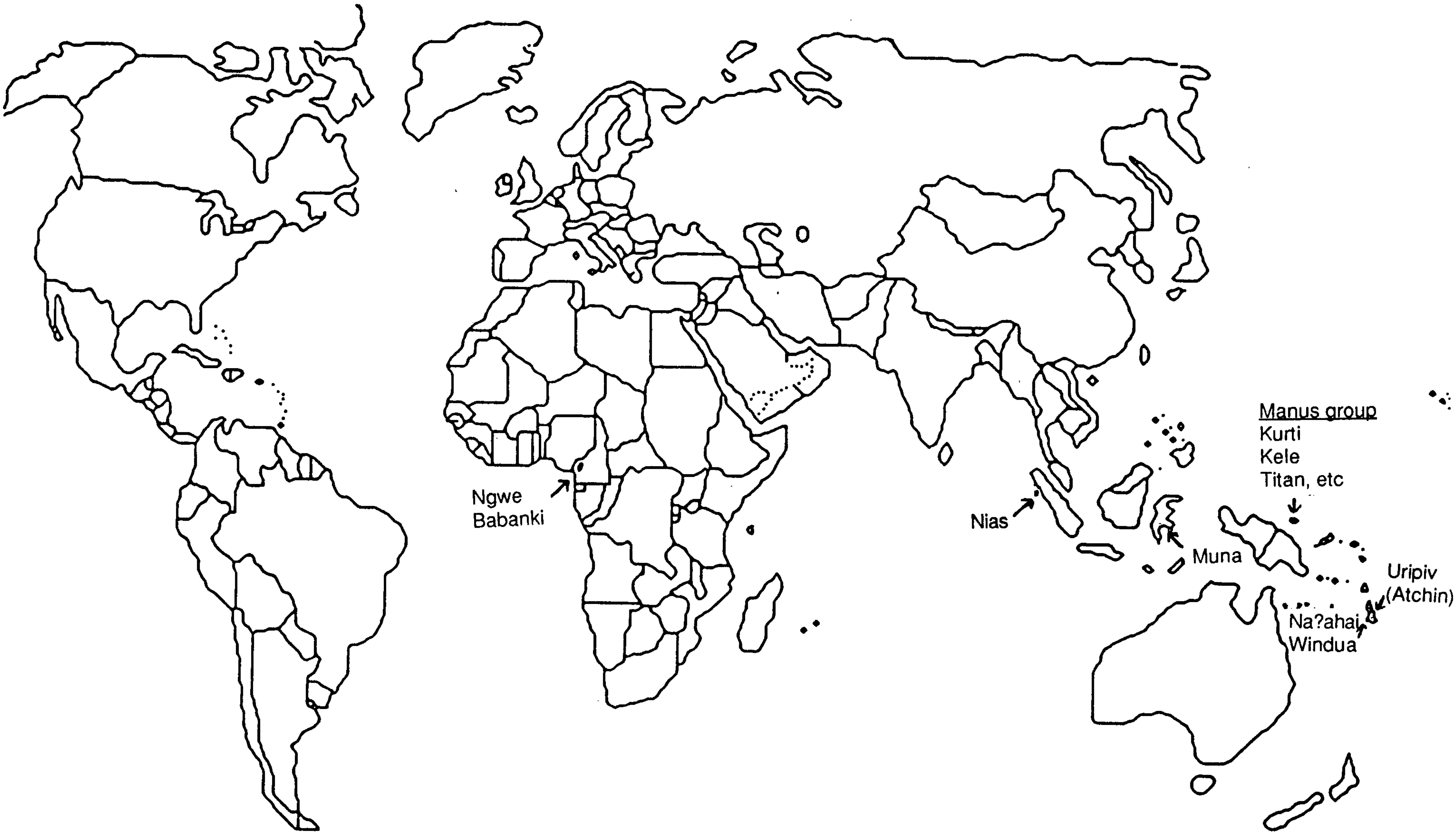
Since it does not seem in itself difficult to make bilabial trills—in fact, Catford (1977, 1988b) actually suggests that they are the easiest kind of trill to make—two questions arise: 1. Why are bilabial trills so rare in language (given that they are common nonlinguistic sounds)? 2. Why do they occur in the languages which do use them?

As a preliminary answer to the first question, we will assume that bilabial trills are rare because they are difficult to integrate into the stream of other sounds: i.e. the articulatory and aerodynamic conditions required to initiate trilling of the lips result in difficult transitions to and from adjoining segments of most types. Hence they are disfavored sounds for ordinary linguistic purposes. With this assumption granted, it is the second question that needs to be answered — why do some languages make use of bilabial trills despite this presumed disadvantage. The answer seems likely to be found in the fact that, with rare exceptions, it is only in a highly specific phonetic environment that bilabial trills have developed. In the two groups of langages mentioned above, the bilabial trills have typically developed from a (historic) sequence of a prenasalized voiced bilabial stop followed by a vowel with a close lip position, that is, in a sequence such as /*mbu/.

The restriction of the development of bilabial trills to this environment suggests that it may contain the specific conditions in which trilling develops *spontaneously*. Subsequently, over a period of time this pronunciation may become established as the norm. Since trilling is an aerodynamic phenomenon it seems most likely that the particular aspects of the environment in question that are relevant are aerodynamic ones, rather than, say, considerations of articulatory simplicity or acoustic distinctiveness. This paper will document the claim that bilabial trills develop in the stated environment, provide phonetic measurements from several languages and discuss the aerodynamic conditions assumed to lead to the occurrence of trills. This latter section includes a simulation of trilling using an aerodynamic model, in order to show that the factors posited as relevant are in fact so.

Distribution and diachrony of bilabial trills

In the first instance, we may note that for several of the languages concerned, the bilabial trill can still be regarded as a conditioned allophone of /mb/, or as a feature of the transition to the following /u/. This is true for Muna and seems true for



Map showing locations of some languages with bilabial trills.

Na?ahai and Windua. Prenasalized bilabial stops without trilling occur before other vowels, and in Muna, even before /u/. The trilled pronunciation is only heard in informal speech in Muna and principally among women. A list of Na?ahai examples showing the distribution of stop and trill in this language is given in Table 1.[2] Following a convention proposed by Ladefoged, Cochran and Disner (1977), a bilabial trill is transcribed with the symbol [B]. In these examples, the symbol should be taken to represent a prestopped trill; that is, the first phase of the trill is a bilabial stop burst and the sequence could be more narrowly transcribed as [mbB].

Table 1. Distribution of bilabial trills in Na?ahai

example	*gloss*	*comments*
trilled release		
ni-mBusin	"tail"	
ni-mBuas	"pig"	
ni-mBuŋ	"manguru fish"	
na-ʔamBur	"unicorn fish"	stem-initial ʔ predictable
na-ʔamBu	"firewood"	ditto, final vowel devoiced
ni-kimBu	(tuber sp.)	final vowel devoiced
ni-mBur	"elephantiasis"	
ni-mBujaŋ	"water-taro"	
ni-mBumBur	"nose ornament"	
ni-mBuri	"clay"	
na-βat mBuŋ	"limestone"	
non-trilled release		
na-mbatuŋ	"my head"	
na-mbɔʔɔŋ	"my mouth"	
na-mbataβamb	"your shoulder"	final stop devoiced, no final vowel
na-βarβiambamb	"wing"	final stop devoiced, no final vowel
ne-βemb	"butterfly"	final stop devoiced, no final vowel
nə-mbɷr	(fungus sp.)	?counterexample; see below.

These examples show that [mb] and [mB] are in complementary distribution: [mB] occurs before /u/, [mb] before other vowels, with the possible exception of the last item, a word for a species of fungus. This item is a little puzzling. All of the examples in Table 1 are in fact noun phrases, and are prefixed by a clitic element whose function is (roughly) to mark them as definite. Historically, this has the form /*na-/, but in Na?ahai it has become subject to a type of vowel harmony governed by the initial vowel of the stem it precedes. Before the high vowels /i/ and /u/ the clitic is /ni-/, as in "tail" and "(tuber)" above, before /e/ it is /ne-/ as in "butterfly" and before /a/ or /ɔ/ it is /na-/. In my field notebook the stem vowel of the "fungus" word is written the same as that in /ni-mBur/ "elephantiasis" where the expected trilling before /u/ occurs. However, no trilling occurred in any repetition of the "fungus" word. Moreover, the prefix has a clearly lower and more centralized vowel in this case, and is transcribed /nə-/ for want of better choices. It seems likely that this word has a different but unrecognized stem vowel which is not /u/ — I have transcribed it with [ɷ] in Table 1 — and hence is not a counterexample to the complementary distribution.

In the Manus languages, bilabial trills remain restricted to the position before /u/, and occur only stem-initially. Historically, all occurrences derive from /*mb/, almost always in initial position. This includes secondary instances in nouns which arose from processes of vowel elision and stop voicing in the sequence /*na-p-/, where /*na-/ is the noun prefix mentioned above in the discussion of Na?ahai and /*p-/ is stem-initial (Ross 1988). Despite the restricted distribution of the trill, it is possible to analyze /mB/ as an independent phoneme, as is proposed by Blust (1986) in his analysis of Nali, and by Barthels (ms) for Lele. This is because /*mb/ in other environments has merged with /*p/ as /p/ (initially) or /ɸ/ (medially). Hence there can be no allophonic variation with [mb], and [mB] is not the only labial stop variant before /u/ since there are a few cases where /p/ appears.

The assertion that bilabial trills only occur before /u/ in Manus languages is based on my partial investigations and on comparisons with the data of other linguists who have more specialized knowledge of these languages. Robert Blust kindly compiled an exhaustive list of the Nali items with bilabial trills known to him. This wordlist is given in Table 2, with some substitutions mainly to make the transcription closer to IPA conventions (such as 'j' for intervocalic 'y' and 'i' and 'u' as the second element of diphthongs in place of 'j' and 'w'.) Inspection of the data in Table 2 shows that with the single exception of "spider", cited as /kamBou/, every instance of [mB] precedes the vowel /u/. And if the exception is not an error, even here the trill is preceding a diphthong containing /u/. In a similar fashion, the entire list of 20 words containing bilabial trills I obtained from Thomas Kele-eh, a Kurti speaker at Aiyura National High School, shows [mB] preceding /u/ with a single exception which has /o/. This is a perhaps morphologically complex form containing a negative, noted as [mBori] "I don't want", [mBorim] "you don't want", and its correct analysis is uncertain. A possibly related form, [mBone], glossed as "don't give it", is cited in the manuscript on Lele by Barthels (n.d.). All other words with bilabial trills cited in this manuscript have /u/. These words were all checked with Bernard Mitol, a Lele speaker at the University of Papua New Guinea, and a few additional ones were noted. Words from these two lists which obviously coincide with items on Blust's Nali list have been added to Table 2, together with a few etymological and other comments. The reconstructed forms for the Proto-Oceanic daughter language of Austronesian (hereafter POC), which is parent to both the Manus and Malekula languages cited here, are based on those given by Ross (1988).

Table 2. Blust's wordlist of Nali items with bilabial trills

	example	gloss	comments
stem initial			
	mBut	"left side"	
	mBut	"Barringtonia"	*Kurti* mBut
	mBua	"testicle"	*Kurti* mBuem "scrotum"
	mBuji	"vulva"	*Kurti* mBuim
	mBupo	"thigh"	
	mBu	"wood"	
	mBul ndrin	"corner"	
	mBuje	"this house (?)"	query on gloss is Blust's
	mBuŋ	"bunch, cluster"	
	mBun	"banana"	< POC na-pudi; *Lele* mBul
	mBuran	"sap (?)"	query on gloss is Blust's
	mBue	"betel nut"	
	mBua kei	"fruit"	< POC na-pua
	mBunei	"rat"	
	mBuai	"crocodile"	< POC na-puqaja *Kurti* mBua, *Lele* mBue
	mBusuhoŋon	"extremity"	
	mBukei	"giant clam"	
	mBua-kop	"hermit crab (?)"	query on gloss is Blust's
	mBuson	"island"	*Kurti* mBuso, *Lele* mBusol
	mBusas	"foam"	*Kurti* mBuses
	mBuroi	"mountain peak"	
	mBulou	"rain; interior (of island)"	
	mBuskau	"small"	
	mBusuii	"small"	
	mBulen	"work"	*Lele* mBulel
	sai mBusukau	"path"	
	mBusi nima	(gloss missing)	

synchronically non-initial

komBura	"navel, ankle"
tumBu	"ancestor"
maremBun	"sea-snake"
kamBou	"spider"
lomBuan	"wet"
komBuje	(gloss missing)

phrases

i-ra-mBusuŋi-i	"he pinched it"
a-mButue-i	"hold it"
a-lomBuli-i	"untie it"
a-mBusi-i	"make it good"

Smaller numbers of examples from the other Manus languages can be found in Blust (1986) and Ross (1988), as well as in the phonetic description of some Kele and Titan words contained in Ladefoged, Cochran & Disner (1977).[3] These confirm the impressions of restricted distribution based on the more extensive materials available from the three languages cited above.

Thus we see that in both the Manus group and in Na?ahai, representing the Malekula languages, bilabial trilling only arises from prenasalized bilabial stops occurring in a particular vowel environment. In Manus languages the development is limited to stem-initial positions. For the most part, the synchronic distribution in these languages remains highly restricted, although it should be noted that in one of the Malekula languages, the Uripiv dialect, it is now possible for bilabial trills to appear in word-final position following loss of some word-final vowels. As a result of this process, Uripiv now has contrastive /mb/ and /mB/ in a limited number of cases.

The Bantoid data will be described before returning to discussion of the remaining Austronesian languages, Nias. In the Bantoid languages we are concerned with, such as Ngwe, there are grounds for believing that the origin of bilabial trills is restricted to word-initial position, and that a similar nasal-stop-vowel pattern is required for their development. In initial position in nouns an /*mb/ sequence can only arise from combination of a nasal prefix, most often the "animal" noun class prefix /n-/ (class 9/10), with a root beginning with /b/ (the nasal becomes homorganic with a root-initial consonant). Ngwe verbs receive a nasal prefix in a form described by Michael Nkemnji, who provided the data, as an 'infinitive'. This is the usual citation form of verbs. Among other uses of this 'infinitive' it is the required form after the continuous aspect marker /sə̀/.

I have (so far) encountered only eight words with bilabial trills in Ngwe, four nouns and four verbs. The words in question are given in Table 3. Five of these have Proto-Bantu reconstructions in Guthrie (1967-71) with rounded vowels, and these roots are all quite widely distributed in the languages of the area. Proto-Bantu is usually assumed to have had a fairly straightforward 7 vowel system (I represent Guthrie's vowels with the symbols /į, ɩ, e, a, o, ω, u/), but many of the Grassfields Bantu languages have developed quite complex vowel inventories with front rounded, back unrounded or central vowels of various kinds. Diachronic processes conditioned by vowel quality may become somewhat opaque as a result of these developments of the vowels. Of the 14 or so distinct vowel qualities in Ngwe, bilabial trills occur before only three in the data to hand, those which I have written as /ɵ/, /ʉ/ and /ɔ/.

Table 3. Words with bilabial trills in Ngwe (tones omitted)

	example	*gloss*	*Guthrie's Proto-Bantu*
Nouns			
	mBɵ	"tadpoles"	
	mBɵɔ	"goat"	*-bωdi
	mBʉa	"nail"	
	aBɵɔ	"ashes"	*-bu
Verbs			
	mBɵ	"get lost"	*-bωd-
	mBɵ	"become rotten"	*-bod-
	mBɵ	"extinguish"	
	mBɔ	"lance, drain (e.g. a boil)"	

In all cases the vowels following the trill have a quite small lip aperture. Six of the eight examples have the vowel /ɵ/, which it seems fair to take as the most conducive to trilling. Dunstan (1964), who cites only the "tadpoles" example, describes this vowel as a "low back vowel with lips close together" (she transcribes the word as /mBɤɤ/.) Ladefoged's x-rays (1968) of an Ngwe speaker show that the vowel is not particularly low, although the speaker he studied did not produce a trill before it in the particular utterance of this word that was recorded. I interpret it as a higher mid rounded central vowel. Before /ɵ/ it would seem possible to regard the trill as a predictable variant of /mb/ since /ɵ/ does not occur after plain [mb], but before /ɔ/ there is contrast between /mBɔ/ "lance (v.)" and /mbɔ/ "Creator".

One of the Ngwe words in table 3, /aBɵɔ/ "ashes", does not have an initial nasal prefix, which raises the question of whether a nasal is a required part of the environment which favors development of a bilabial trill. All of the verbs replace the infinitive nasal prefix with a noun class concord prefix in most constructions, and the noun /mBɵ/ "tadpoles" has a singular with the prefix /lə-/. The trill may be retained in these forms, e.g. [ləBɵ] "tadpole", and hence does appear in an intervocalic position. However, more frequently a voiced bilabial fricative is heard in these positions rather than a trill. In view of the distributional patterns and the data from other languages, we may conclude that presence of the nasal is plausibly implicated in the development of the trill, since trilling is regular only in prenasalized contexts. "Ashes" remains an exception. Yet this word in many languages in the Grasslands group has a nasal prefix, derived from the class 6a prefix /*ma-/, semantically associated with mass nouns and liquids (Voorhoeve 1980). Perhaps this root also had this prefix at a earlier stage in Ngwe and has retained the trilled pronunciation despite shifting to a different prefix.

Hyman reports only one word with a bilabial trill in Babanki, the word for "dog" /mBa/ (cf. Proto-Bantu *n-bua). Kom has a trill in the word for "goat" which has been reduced to a monosyllable, /mBi/ (cf. Proto-Bantu *n-bɷdi). These data are limited but agree with other data in having an original input sequence similar to /mbu/.

We now return to Nias. Catford (1988a) lists 13 words with bilabial trills. The trills are usually pronounced with prenasalization, except apparently in utterance initial position where no nasal portion is perceptible. Most of the reconstructed proto-forms he compares them with have prenasalized bilabial stops. However, there seems to be no limitation to particular following vowel environments. Nor is there any predictability of when the trills will develop based on other characteristics of the reconstructed forms he compares them with (Proto-Austronesian reconstructions from Dempwolff (1938), 'modernized' with the assistance of Robert Blust.) The Nias words in question are listed in Table 4, ordered according to the vowel that follows the trill. The transcription is standardized to show prenasalization in all cases, although Catford's recording of it is variable.

Table 4 shows that all six of the vowels recognized as phonemic can follow a bilabial trill, and trills happen to precede the most open vowel /a/ more often than any other vowel in this sample of words. The vowels are for the most part etymologically conservative, unlike the situation in Ngwe. Furthermore, productive alternation between /b/ and /mB/ can occur in the process Catford calls consonant gradation, apparently also generating a trill before any vowel. For example, the word /baβi/

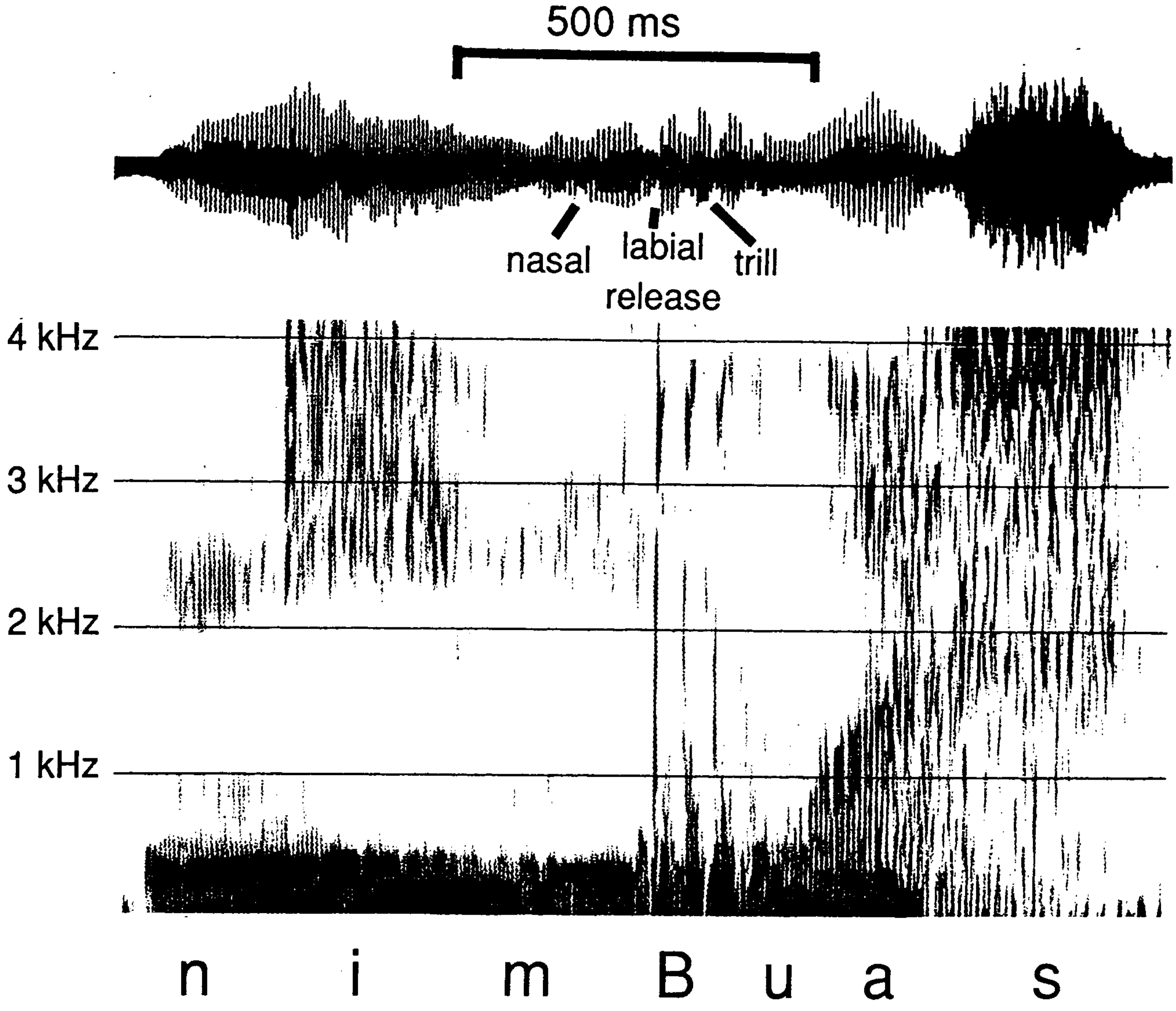


Figure 1. Spectrogram and waveform of the word [nimBuas] "a pig". The language is Na?ahai (Southwest Bay, Malekula Island, Vanuatu).

"pig" exchanges its initial bilabial stop for a prenasalized bilabial trill in the phrase /begu mBaβi/ "feed the pig" (or any other construction of this type). The diachronic source of this gradation phenomenon has not been investigated; its analysis might clarify the Nias developments in general.

Table 4. Words with bilabial trills in Nias (S = Southern dialect, N = Northern dialect)

vowel	*word*	*gloss*	*PAN reconstruction*	
-u	mBua	"fruit"	*mbuaq	"areca nut"
	lɤmBu	"coconut" (S)		
-o	-homBo	"fly (v.)"	*ʃambaw	
	simBo	"steam"	*sebu	
-i	namBi	"goat"	*kambin	
	zimBi	"jaw" (N)		
	simBi	"jaw" (S)		
-e	emBe	"pail"	*timba	
-ɤ	andɤmBɤli	"bargaining"	*-beli	
-a	mBanua	"village"	*mbanua	
	mBanio	"coconut" (N)	(*niuʀ)	
	mBaso	"read"		
	tamBago	"copper"	*tumbaga	
	dzimBao	"egret" (S)		

In summary, Nias limits the development of bilabial trills to prenasalized labial stop environments, but does not have a restriction on the following vowel. The other languages favor development of bilabial trills in environments in which a nasal element precedes and a vowel with a close lip position follows an original bilabial stop.

Phonetic description and aerodynamic considerations

We now turn to the phonetic description of bilabial trills. To illustrate a fairly typical bilabial trill, a spectrogram and waveform of an utterance of the utterance [nimBuas] "the pig" in the Na?ahai language are given in Figure 1. The speech rate is somewhat slow in this token, resulting in rather long segment durations. In the /mBu/ portion there is a relatively long bilabial nasal segment, then the nasal port is closed a short time before the bilabial closure is released. Following this initial release, there seem to be three oscillatory movements of the lips. To judge from the weak bursts associated with the first two of these oscillations, the lips make at least partial contact. The third oscillation is detectable mainly because of the resulting attenuation of amplitude, visible most clearly in the waveform. A short final section of

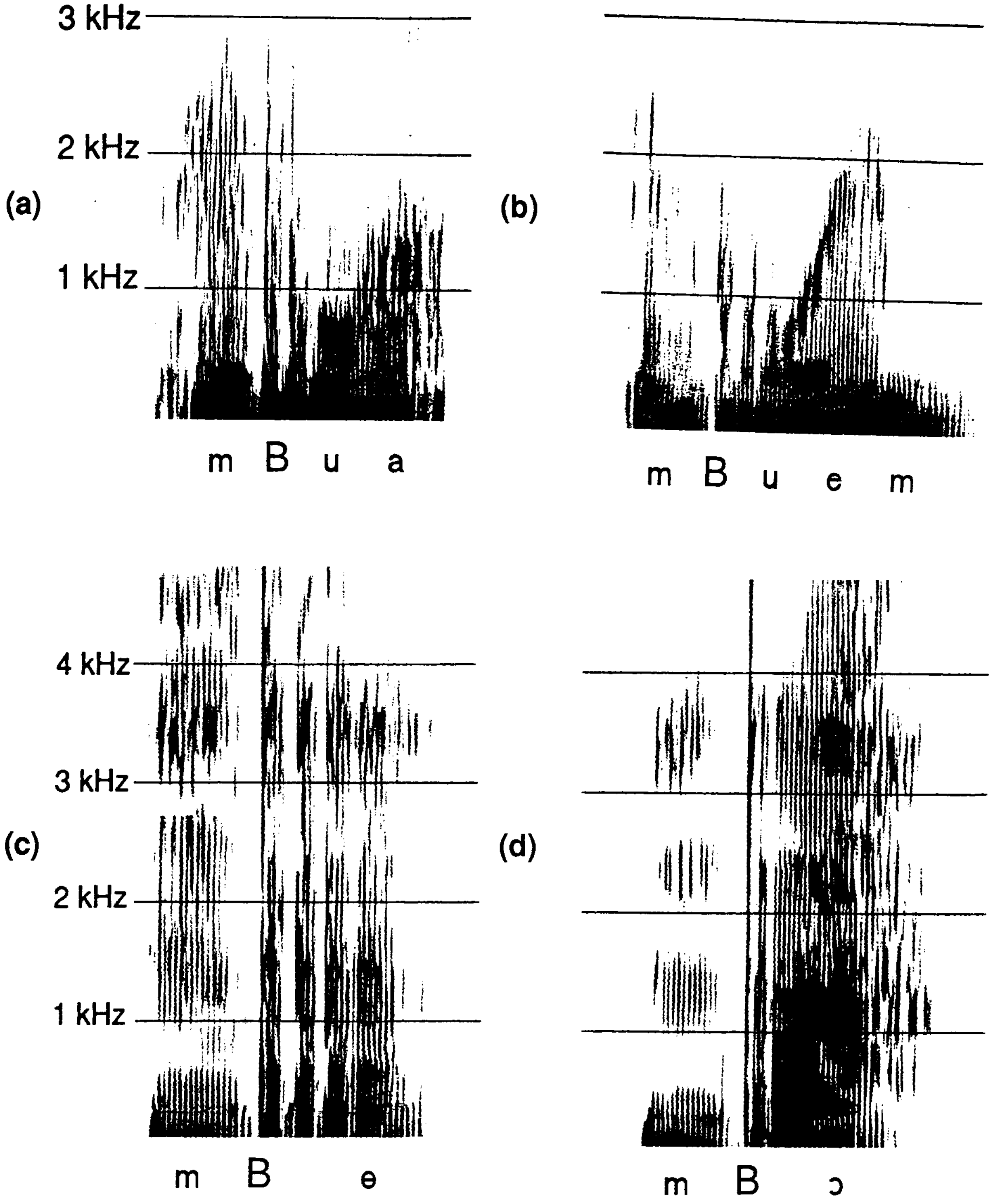


Figure 2. Additional spectrograms illustrating bilabial trills. Items shown are the words (a) [mBua] "crocodile" and (b) [mBuem] "scrotum" from Kurti (East Manus group, Papua New Guinea), and (c) [mBe] "tadpoles" and (d) [mBɔ] "lance (a boil)" from Ngwe (Grassfields, Cameroun). Number of measured trill periods ranges from 1 in (d) to 3 in (c); (a) and (b) both have two reclosures, but an additional weak partial approximation of the lips which reduces amplitude of F2 can be detected in (b).

the /u/ vowel remains unaffected by trilling. The entire [mBu] portion has a duration of about 500 ms, and is voiced throughout: at no time is voicing inhibited during the oral closures, as happens in many instances of tongue-tip trills and frequently in the "voiced" flap allophone of /t, d/ in American English.

Additional spectrograms of trills illustrating some of the range of variation found in the data studied for this paper are shown in Figure 2. Spectrograms like these were made from utterances with bilabial trills spoken by individual speakers of four languages, three Austronesian, one Bantoid. Measurements of the trills are given in Table 5 below. As can be seen, the oral portion of the initial bilabial closure has a mean duration on the order of 30 ms. Subsequent trilling was measured in periods beginning from the release of this closure, with the end of each period judged to occur when there was a sharp rise in amplitude. Since these measurements were made on spectrograms, there were undoubtedly weak trill periods that were missed, since the amplitude rise can be hard to identify. All speakers also produced some tokens of words in which trilling was expected but failed to occur, and tokens in which one or more observable weak partial reclosures of indeterminate duration occurred. These tokens were not measured nor included in the results reported here. The modal number of trill periods observed was 2 for all 3 Austronesian speakers, but is only 1 for the Ngwe speaker studied. The mean duration of the observed trill periods is also reported below. The frequency of the trills was calculated for each token, and averaged across tokens. This procedure may result in an apparent small discrepancy between trill period and trill frequency, since one is a mean per period and the other a mean per token.

Table 5. Mean measurements of prenasalized bilabial trills in 3 Austronesian languages and Ngwe (1 speaker each, no. of tokens in parentheses after language name)

	<u>Duration of non-nasal ([b]) closure</u>	<u>Duration of period in trill</u>	<u>Frequency rate of trilling</u>
<u>Kurti</u> (n = 37)	33.8 ms	37.6 ms	26.8 Hz
s.d.	*5.0*	*3.5*	*2.8*
<u>Na?ahai</u> (n = 22)	30.7 ms	45.6 ms	22.1 Hz
s.d.	*1.2*	*3.7*	*1.8*
<u>Uripiv</u> (n = 14)	33.8 ms	37.3 ms	27.2 Hz
s.d.	*12.8*	*4.5*	*3.8*
<u>Ngwe</u> (n = 15)	33.1 ms	44.1 ms	23.1 Hz
s.d.	*8.12*	*3.73*	*2.1*
Mean across languages	32.85	41.15	24.8

The trill frequencies in the last column of Table 5 are fairly comparable with those reported by Ladefoged, Cochran & Disner (1977) for Kele and Titan, close relatives of the Kurti language, and for Nias reported by Catford (1988a). Ladefoged et al. reported a mean rate for bilabial trills of 29.3 Hz (s.d. 4.8), pooling the data from one speaker of each language. Catford reports a mean of 28.7 Hz (s.d. 3.1) for one speaker of Nias.

We will now consider the aerodynamic conditions and the articulatory movements which would characterize a sequence such as [mbu], and present the argument that suggests why this particular set of circumstances would induce trilling. Initially the nasal port is open and the vocal cords are vibrating. Because of the venting of airflow through the nasal port, there is a minimal increase of intraoral pressure during [m]. Since the stop closure duration is extremely brief in prenasalized stops, the oral pressure during [b] does not have time to increase much from the low baseline established in the nasal portion. Thus intraoral pressure at the time of the labial release is still relatively low, and the force acting to separate the lips due to intraoral pressure is relatively weak. This contributes to a slow initial rate of labial opening. The increase in the size of the aperture between the lips due to the underlying articulatory gesture from consonant to vowel also occurs relatively slowly because the stop is being released into a rounded vowel with a high jaw position. Hence the labial aperture remains constricted for a time following release. We might also assume that the lip tissue is relatively slack at time of release, due to rounding and protrusion of the lips for the vowel. We hypothesize that under these conditions the system remains briefly in a state during which Bernoulli forces created by accelerated flow through the narrow lip aperture may result in involuntary full or partial reclosure of the lips one or more times during the stop-vowel transition. This mechanism is partially similar to the process by which vibration of the vocal cords is initiated when they are positioned with a narrow aperture between them and the subglottal/supraglottal pressure difference is within appropriate ranges.

Each of the considerations laid out above can be examined in more detail. For Bernoulli forces to be effective the labial aperture must not increase too rapidly as the stop is released. Low intraoral pressure and the fact that the stop is being released into a vowel with a narrow target aperture both contribute to this. Although no data on intraoral pressure in prenasalized stops is available at present, values reported for nasals in the literature indicate pressures in the range of 0.5 to no higher than 2.5 cm H_2O (e.g. Brown, McGlone & Proffit 1973). The data in Müller & Brown (1980) indicate that peak pressure during a voiced stop closure is not reached until after 100 ms of closure duration. Interpolating in the curves they provide for carefully spoken

VCV utterances, we may estimate a value of perhaps 3 cm H_2O would have been reached after 30 ms of closure. This duration is fairly typical of a prenasalized stop, and peak pressure in the case of a closure whose total duration is so short is likely to be less.

We argue that unless intraoral pressure is low the impounded pressure will result in too rapid opening of the lips for trilling to occur. On general physical grounds we know that as the lips begin to open one of the forces in effect is one acting in the direction of separating the lips and related to the magnitude of the intraoral pressure and the resistance to airflow between the lips. The effect of this force in the first few milliseconds of a labial release can be inferred from Fujimura's detailed high-speed motion picture study of labial consonants (Fujimura 1961). This study compared speed of lip opening for English /p, b, m/, which are known to be ranked in that order in terms of intraoral pressure. The initial opening rate for /p/ is faster than for /b/, which is in turn faster than for /m/. For example, when /a/ is the following vowel, labial aperture just 5 ms after the explosion is 48 mm^2 for /p/, 31 mm^2 for /b/ and 29 mm^2 for /m/.[4] With /i/ as the following vowel, labial apertures after 5 ms are 28 mm^2 for /p/, 24 mm^2 for /b/ and 11 mm^2 for /m/.

It might be thought that this ordering of opening speeds was the result not of intraoral pressure differences but of differences in articulatory force between /p/, /b/ and /m/. Greater pressure of the lips against each other would produce more tissue compression, stronger elastic restoring forces and hence more rapid opening. However, studies that have measured articulatory pressure and air pressure simultaneously (Lubker & Parris (1970) compared voiced and voiceless labial stops; Brown, McGlone & Proffit (1973) compared alveolar stops and nasals /t , d, n/) have found that air pressure differences between the classes of consonants concerned are more consistent than articulatory pressures and that there is relatively poor correlation between the two. Therefore the different rates of opening for /p, b, m/ in Fujimura are more likely due to the differences in impounded pressure.

The rate at which the lips open is also related to the nature of the following vowel. When the target vowel has a wider labial aperture, then the opening will likewise increase too rapidly for Bernoulli forces to be effective. Fujimura's study again provides data to support this contention. He provides measures of the opening after the release of the stops in the words "prove, spruce, approve," which can be compared with items like "pock, pope, spope: a pope, etc". Labial aperture 5 ms after initial /p/ release is only 5 mm^2 when /ru/ follows, but 28 mm^2 when /i/ or /ou/ follows and 48 mm^2 when /a/ follows. The use of imperfectly matched /u/ examples containing an extra segment, /r/, is unfortunate, but these results are in agreement with the general trend in a number of studies showing that velocity of articulator movement is quite strongly dependent on the distance to the next 'target'. In particular, Kuehn and Moll (1976) showed vertical displacement velocity of the lower lip to be faster in the release gestures for /p, f/ before /a/ than before /u, i/. Kent & Moll (1972) had earlier proposed the general rule that articulatory velocities depend on distance to be traveled. Fujimura's data indicate that the effect of this rule is quite strong in the first few milliseconds following release of a labial stop.

Finally, we assume that unless the lip tissue is relatively relaxed, vibration will be prevented by too much stiffness (or at the very least frequency of vibration will be higher than that observed). This condition is satisfied by a postconsonantal vowel with loosely rounded lip position, since in such a vowel the lip tissue is bundled together rather than stretched out as for a spread vowel.

The general plausibility of the assumptions made above were tested in a preliminary simulation based on modification of an aerodynamic model originally developed for studying vocal cord oscillation. The particular model used is one implemented by Smith & Berke (1988), based on Ishizaka & Isshiki (1976), which is in turn developed from the work of Ishizaka & Flanagan (1972) and Flanagan, Ishizaka & Shipley (1975). A version designed for studying excised larynges was used. This lacks any supraglottal portion, and hence any source-tract interactions, and assumes that the transglottal flow is radiating into the atmosphere. The subglottal pressure parameter can thus be interpreted as intraoral pressure and other variables interpreted as referring to parameters of the lips. This reinterpreted model will be referred to as the labial aerodynamics model.

The model requires user-supplied input values for intraoral pressure, and for lip aperture, lip width, lip thickness, lip mass, and lip stiffness. Based on the intraoral pressure studies cited above, a value of 2 cm H_2O was supplied for intraoral pressure. The model provides for separate specification of inner and outer masses and for asymmetries between upper and lower lip. Since little is known about the actual motion in a labial trill, the upper and lower lips were treated as symmetrical and the inner mass was minimized to eliminate simulating any significant traveling wave effects (parallel to the direction of air flow) in the tissue. Lip thickness was measured in a subject producing an /u/ vowel as about 10 mm, and lip width (horizontal length of opening) was taken from Fujimura (1961) as about 8 mm for the release into /(r)u/. The vibrating mass of each lip was taken to be 1 g, about 5 times the mass of the vocal cords, but the stiffness to be only a quarter of a typical setting for

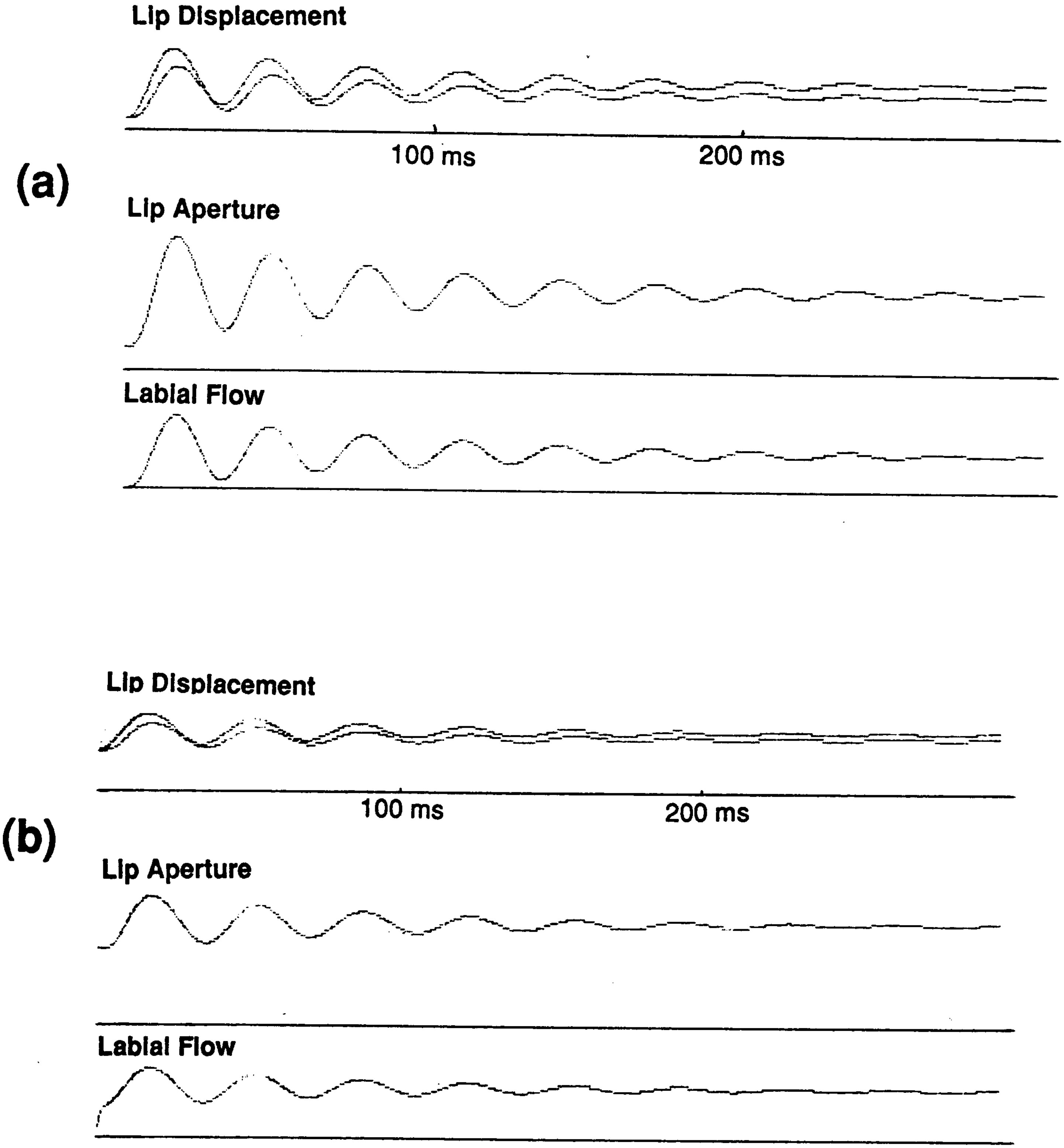


Figure 3. Output of the labial aerodynamic model for (a) initial labial aperture of 1 mm^2, and (b) initial labial aperture of 4 mm^2. Intraoral pressure is set at 2 cm H_2O. Lip length (horizontal measure of opening) is 8 mm for all simulations. The top tracing shows magnitude of displacement of the lip masses in the vertical plane from the midline of closure. The middle tracing shows derived changes in the lip aperture. The third tracing shows variations in calculated airflow though the lips. Scale is proportional to the range of values plotted.

the vocal cords when phonating in the normal range (i.e. 20 kdyne/cm). Lip apertures of 1 mm^2 and 4 mm^2 were investigated to represent the sytem response immediately after release and 5 ms after release. With these values the model produces a damped oscillation of the lips over 200 ms in duration (damping ratios were set .2 to for the outer lip mass and .6 for the inner lip mass). Outputs of the model with these settings, showing displacement of the lip masses (from closure), changes in labial aperture, and airflow between the lips over time, are given in Figure 3 (a) and (b). With the lip aperture set to 1 mm^2 the period was 32 ms and frequency 31 Hz; at 4 mm^2 the period was 35.5 and the frequency 28 Hz. This behavior is quite close to that reflected in the data reported in Table 4 above, except that the oscillations die out sooner in the natural speech data and the frequencies observed are a little lower. These differences can be attributed to the fact that the labial aerodynamics model has static parameter settings, whereas the natural trills occur superimposed on a dynamic gesture of labial opening in the CV transition. If the model is set with an initial labial aperture of even 10 mm^2, oscillation is considerably reduced in both amplitude and duration, as in Figure 4 (a), and the period extends to 44 ms (23 Hz). Note that the measurements in Table 4 reflect the composite response of a system in which the lip area is changing under both articulatory and aerodynamic influences.

Of course, there are many approximations and inaccuracies in this particular model, but it does enable the effect of certain changes of interest on a system that is behaving somewhat like the lips to be examined. Above, it was hypothesized that higher intraoral pressure or a more rapid articulatory lip opening gesture would fail to produce trilling because labial aperture would become too large immediately after release. This can be simulated with the model by further increasing the labial aperture value. Oscillation becomes minimal when labial aperture is set to 20 mm^2, as shown in Figure 4 (b) (intraoral pressure is kept at 2 cm H_2O). This area is less than that measured by Fujimura for /ou/ 5 ms after release. Interestingly, if intraoral pressure is lowered to 1 cm H_2O, then oscillation is barely initiated even at narrow lip aperture, as shown in Figure 5 (a). There is also minimal oscillation if the "tension" parameter of the model ('Q' in the literature) is increased to simulate increased stiffness of the lips, as is shown in Figure 5 (b). These last two conditions indicate that the stop component of the prenasalized stop, and the laxness of the lips in the type of vowel concerned are plausibly functional in inducing trilling.

Discussion and Conclusion

The simulation discussed in the preceding section strengthens the case based on the majority of the linguistic data examined that bilabial trills develop in a quite

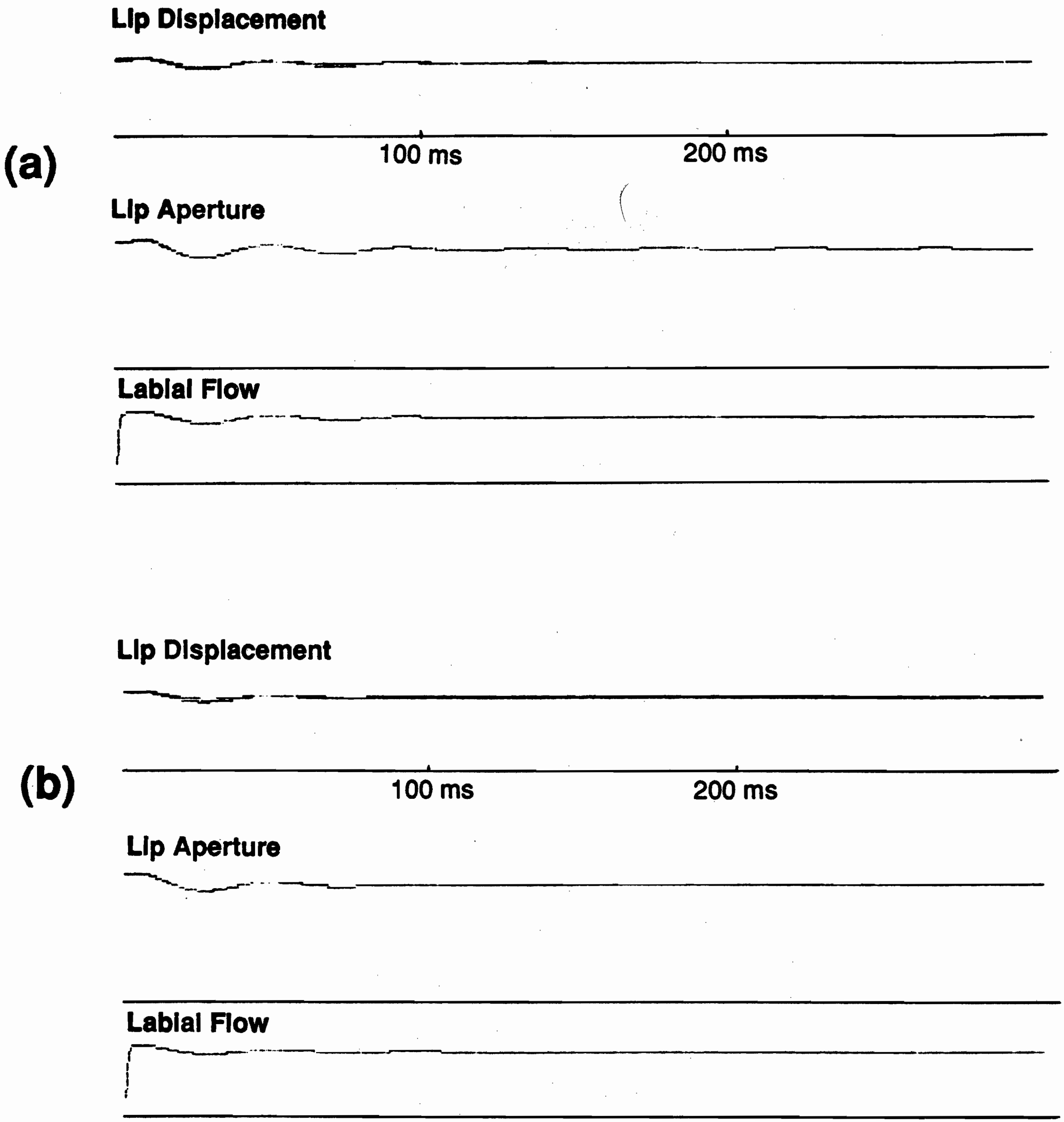


Figure 4. Output of the labial aerodynamic model for (a) initial labial aperture of 10 mm^2, and (b) initial labial aperture of 20 mm^2. Intraoral pressure is set at 2 cm H_2O.

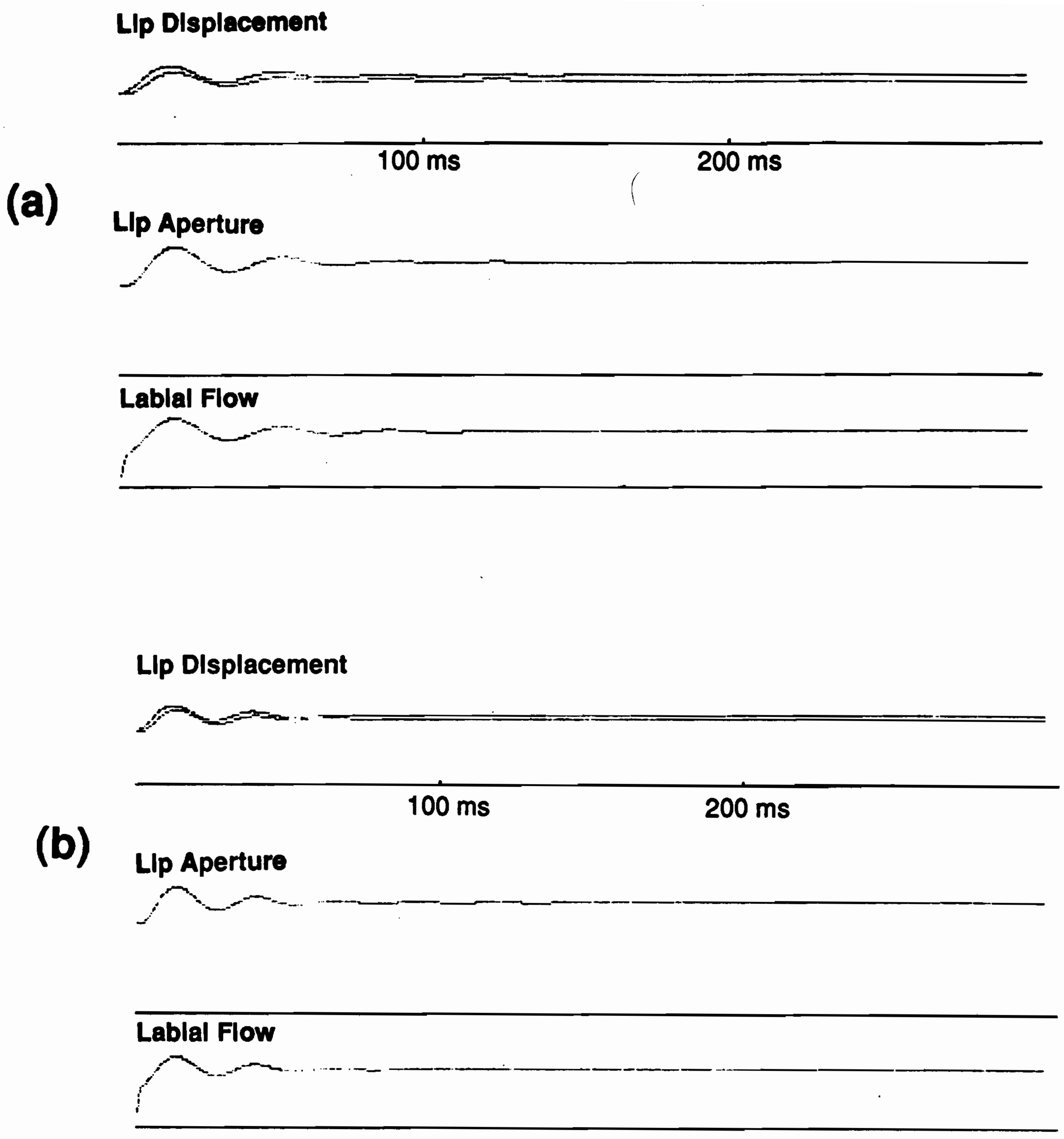


Figure 5. Output of the labial aerodynamic model for (a) initial labial aperture of 4 mm^2 as in Figure 3b, but intraoral pressure of only 1 cm H_2O , and (b) tension parameter 'Q' increased to 1.2 (from 0.9), with initial labial aperture = 4 mm^2 and intraoral pressure = 2 cm H_2O.

specific environment, namely a bilabial prenasalized stop + rounded vowel sequence. Even in these circumstances, the development is comparatively rare as it depends on critical values of several interacting factors, most particularly those which determine the rate at which the labial aperture increases at the CV transition. This makes the occurrence of the bilabial trills in Nias appear all the more surprising. Nias also has prenasalized postalveolar trills (/ṇḍṛ/). Catford (1988a) suggests that the occurrence of these two types of trills in the same language is the result of a common process. In his words:

> "The origin of the bilabial and postalveolar trills is clear enough. They presumably arise from prenasalized stops *mb and *nd as follows. During the nasal phase, when all of the ... air-stream is flowing out through the nose, the bilabial contact for m and the postalveolar contact for n can be quite loose, since they do not have to withstand any appreciable pressure. But at the moment when the escape-route through the nose is shut off ..., the air pressure in the mouth suddenly rises, and if the bilabial and apico-postalveolar articulatory closures are not immediately tightened up, the chances are that the lips and tongue will be set 'flapping in the breeze', as, indeed, we find them doing." (Catford 1988: 154)

The theory implicit here would predict that such trills might be expected to occur in a much larger number of languages than is actually found. Any nasal + stop + vowel sequence would be a candidate for the development of trilling if the articulator concerned permitted it. The other languages that have developed bilabial trills would be expected to have also developed alveolar or postalveolar trills from prenasalized stops. This is not the case. Even though the Manus languages do have a prenasalized alveolar trill /ndr/ this is from POC /*nr/, the prenasalized counterpart of the 'ordinary' alveolar trill /*r/, not an innovation parallel to the bilabial trills in these languages (Ross 1988). As expected, the occurrence of (post)alveolar trills in Oceanic languages is not correlated with occurrence of bilabial trills. Fijian, for example, retains /ndr/ as well as /nd/ and /mb/, neither of which produce trilling.[5] In Nias the majority of the postalveolar trills derive from the Proto-Austronesian segment usually symbolized as 'D'. This was itself probably originally a trill.

Moreover, the presumption that trilling is usually prevented by tightening the articulatory closure seems unlikely to be correct. The available data on articulatory pressure indicates that nasals usually have a force of articulatory contact equal to or greater than that of voiced stops. Brown et al (1973) report mean lingual pressure to be higher for /n/ than for /d/ in /i/ and /u/ contexts, but slightly lower in /a/ contexts for four English speakers. Malécot & Richman (1974) report higher labial pressure for /m/ than for /b/ in /a/ contexts for all five French speakers they studied and for 3 of 5

English speakers. In no case are the differences very large. The amount of lip compression measured in /m/ and in /b/ was also very similar in both languages. These data are not from Austronesian prenasalized stops, of course, but they suggest that there is no reason to expect labial tension to be changed between nasal and stop portions of a normally produced prenasalized stop such as /mb/.

We would agree that labial tension is a factor in allowing the development of labial trills, but would argue that this is associated with the nature of the particular following vowel, not with the production of the /mb/ consonant. However, the factors which jointly control the rate at which the lips are opening in the CV transition are of greater importance than labial tension. In particular, this means that the vowel must be one in which the target labial aperture is relatively small. The situation in Nias thus remains a puzzle, since trilling is not limited to any particular (historical) vowel context. It might be that Nias labials at one time developed rounding, as has happened in some of the Oceanic languages, and the consonantal rounding was the trigger for the labial trilling. Alternatively, rounding may have been an attribute of an earlier overt morphological element which has left its trace in the alternations that Catford calls consonant gradation. Or possibly labial trills did originally develop before /u/, but were analogically spread to other vowel environments.

Nias aside, bilabial trills are principally limited to the highly specific environment described in this paper and their origin can be traced to the aerodynamic conditions that exist in this environment. This makes these sounds of particular interest for constructing a typology of sound change. Most phonetically-motivated sound changes can be traced to auditory or articulatory factors. Only in a comparatively small number of other situations has a sound change been explained as resulting from the aerodynamics of speech production. One such case is the devoicing of geminate stops (as in Cuna) and another is the devoicing of high vowels (as in Japanese). Labial trilling can be added to this short list.

Footnotes.

1. At least two languages of Mexico have a very limited role for bilabial trills. In Amuzgo and Isthmus Zapotec the word for "ant-lion" ends with a bilabial trill. The word seems to function like an ideophone. Bilabial trilling also occurs as an allophonic variant of the labial fricative vowel [y̩] in Lianshang Yi following initial labial and alveolar stops. This vowel has a very close lip position, and is often trilled for all or most of its duration. As many as 8 periods have been measured.

2. These data were compiled with the assistance of Ross McKerras and Longdal Nobel at an S.I.L. translation workshop in Santo.

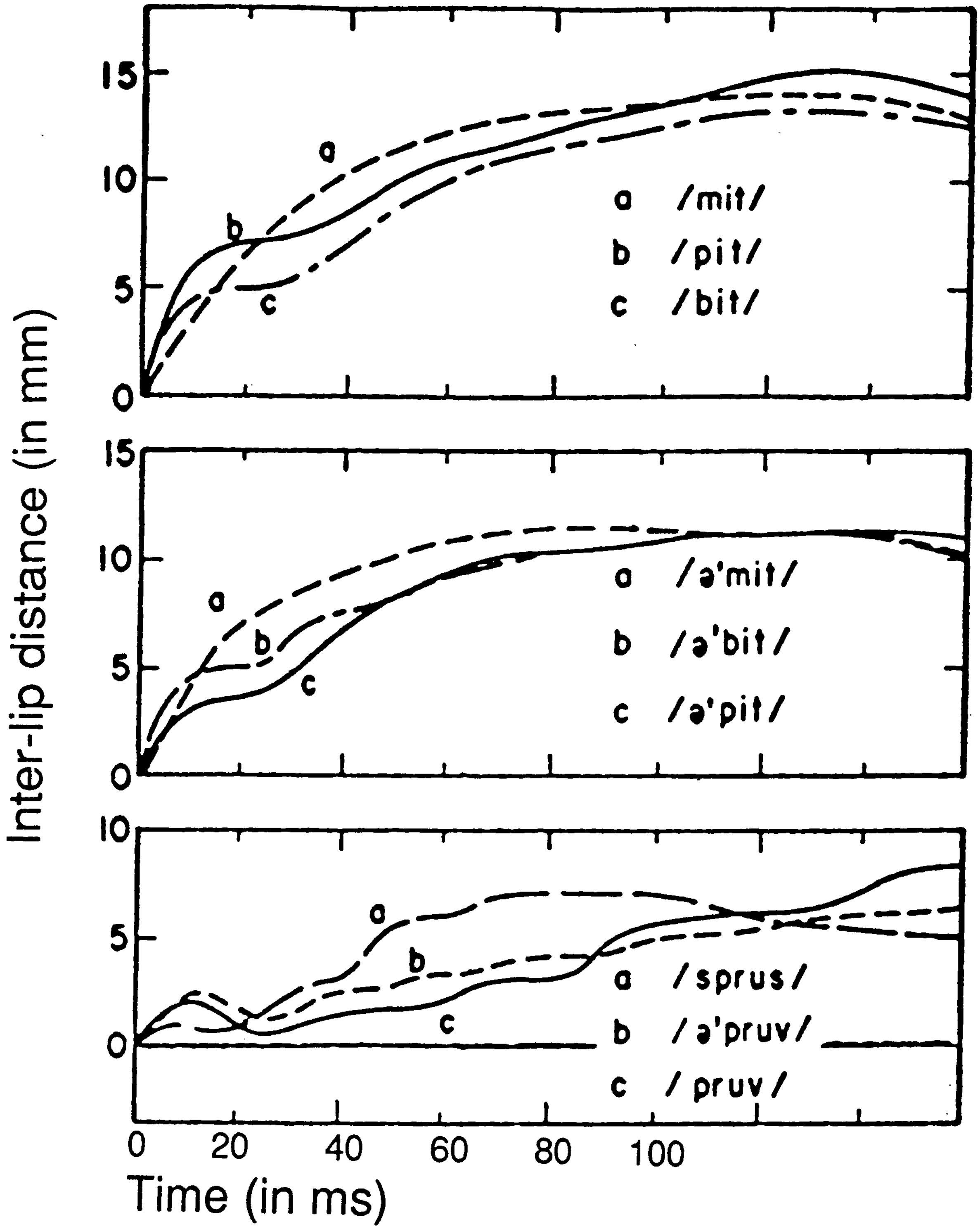


Figure 6. Vertical distance between lips over time, from high speed cinéfilm of an English speaker (after Fujimura 1961). Labial release is at 0. Note in the top and middle panels that stop releases are followed by a more rapid initial opening than nasals preceding /i/, but then the opening movement rapidly decelerates, superimposing an oscillation on the opening trajectory. In the sequence /pru/ (bottom panel) this deceleration results in near-approximation of the lips about 25 ms after the release. This is presumed to result from the fact that the opening gesture is slower in this case because the target aperture is smaller.

3. Ladefoged et al do not note the restriction of [mB] to positions before /u/ in Kele and Titan and cite apparent Kele counterexamples [mBen] "fruit" and [mBin] "vagina". These words actually contain /u/ and the spectrogram of the latter word published in their article clearly shows a much lower onset of F2 than could be accounted for by the labial consonant alone. Cognates of these words can be found in Table 2.

4. These areas are actually approximations calculated by treating the horizontal and vertical measurements of opening as the major and minor axes of an ellipse. Fujimura reports mean area 5 ms after release and at the maximum opening. The time course for the vertical distance measurement is shown in some detail for some "typical examples". The figure reproduced here as Figure 6 is based on that provided by Fujimura (1961: 236).

5. POC *mb and *nd arise from Proto-Austronesian *mp and *nt through voicing assimilation. Many sources continue to write them as *mp and *nt in POC, but Ross notes that voiced reflexes predominate throughout Oceanic and suggests that the voicing assimilation is Pre-POC.

References.

Abercrombie, David. 1967. *Elements of general phonetics.* Edinburgh University Press, Edinburgh.

Blust, Robert. 1986. Admiralty languages. Paper presented at a meeting of the Austronesian Circle, University of Hawaii, November 11th 1986.

Brown, William. S., Robert E. McGlone & William R. Proffit. Relationship of lingual and intraoral pressures during syllable production. *Journal of Speech & Hearing Research* 16: 141-151.

Catford, J. C. 1977. *Fundamental problems in phonetics.* Indiana University Press, Bloomington.

Catford, J. C. 1988a. Notes on the phonetics of Nias. In *Studies in Austronesian Linguistics (Monographs in International Studies, South-East Asia series, No. 76)* ed. by R. McGinn. Ohio University Center for International Studies, Athens, Ohio: 151-172.

Catford, J. C. 1988b. *A Practical Introduction to Phonetics.* Oxford University Press, Oxford.

Charpentier, Jean-Michel. 1982. *Atlas Linguistique du Sud-Malakula (Collection Langues et Cultures du Pacifique, No. 2).* SELAF, Paris.

Flanagan, J. L., K. Ishizaka & K. L. Shipley. 1975. Synthesis of speech from a dynamic model of the vocal cords and the vocal tract. *Bell System Technical Journal* 54: 485-506.

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

Guthrie, Malcolm. 1967-71. *Comparative Bantu (4 vols).* Gregg, Farnborough.

Hyman, Larry M. 1980. Babanki and the Ring group. In *L'Expansion Bantoue I: Les classes nominales dans le bantou des Grassfields,* ed. L.M. Hyman & J. Voorhoeve. SELAF, Paris: 223-258.

Ishizaka, K. & J. L. Flanagan. 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal* 51: 1233-1268. Reprinted in *Speech Synthesis,* ed. J. L. Flanagan & L. R. Rabiner. Dowden, Hutchinson & Ross, Stroudsberg, PA: 148-183.

Ishizaka, K. & N. Isshiki. 1976. Computer simulation of pathological vocal-cord vibration. *Journal of the Acoustical Society of America* 60: 1193-1198.

Kent, Raymond D. & Kenneth L. Moll. 1972. Cinefluorographic analyses of selected lingual consonants. *Journal of Speech & Hearing Research* 15: 453-473.

Kuehn, David P. & Kenneth L. Moll. 1976. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4: 303-320.

Ladefoged, Peter. 1968. *A Phonetic Study of West African Languages* (second edition). Cambridge University Press, Cambridge.

Ladefoged, Peter, Anne Cochran & Sandra F. Disner. 1977. Laterals and trills. *Journal of the International Phonetic Association* 7.2:46-54.

Lubker, James F. & Pamela J. Parris. 1970. Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America* 47: 625-633.

Malécot, André, & Marie Richman. 1974. A labiographic study of bilabial stops in English and French. *Studia Linguistica* 28: 100-108.

O'Connor, J. D. 1973. *Phonetics.* Penguin, Harmondsworth.

Ross, Malcolm D. 1988. *Proto-Oceanic and the Austronesian Languages of Melanesia (Pacific Linguistics, Series C, No. 98) .* Department of Linguistics, Research School of Pacific Studies, Australian National University.

Smith, Marshall & Gerald D. Berke. 1988. Four-mass model of the vocal cords. Computer program implemented at the Speech & Audiology Laboratory, Wadsworth Veteran's Administration Hospital, Los Angeles.

Tryon, Darrell T. 1976. *New Hebrides Languages: An Internal Classification.* Pacific Linguistics, Series C, No 50.

Voorhoeve, Jan. 1980. Kenyang. In *L'Expansion Bantoue I: Les classes nominales dans le bantou des Grassfields* ed. L.M. Hyman & J. Voorhoeve. SELAF, Paris: 274-285.

Multiply Articulated Segments and the Feature Hierarchy

Ian Maddieson and Peter Ladefoged

Expanded version of a paper presented at the 1988 Annual Meeting of the Linguistic Society of America, New Orleans, December 27-30 1988.

Articulator nodes

Recent proposals for the hierarchical arrangement of place of articulation features recognize three major supralaryngeal articulators, represented by Labial, Coronal, and Dorsal nodes, under a place node. A fragment of a hierarchy is shown as figure 1 (after Clements 1985, Sagey 1986, etc).

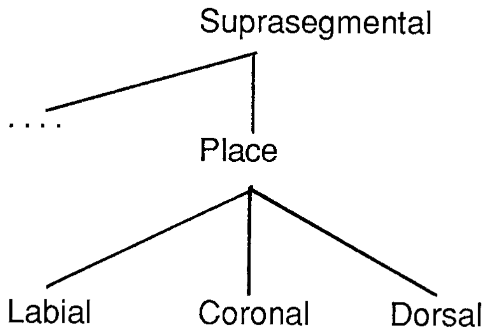


Figure 1. Fragment of a feature hierarchy.

The arrangement of these nodes as sisters expresses the fact that the lips, the tongue tip/blade and the tongue body can be manipulated with a fair amount of independence from each other. The tongue root and the epiglottis form a fourth independent major articulator system, justifying the addition of a Radical node to the tree. This is shown added to the tree in figure 2 (after Ladefoged & Maddieson 1986, Sagey 1987, Ladefoged & Halle 1988, Halle 1988, etc).

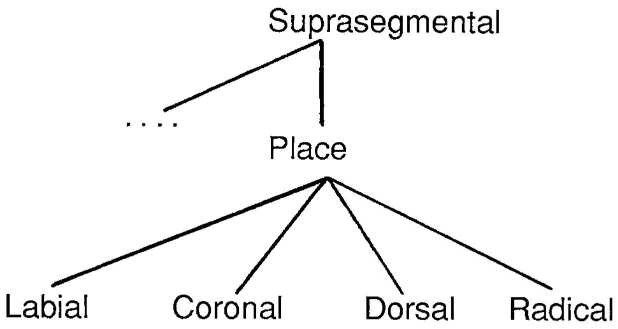


Figure 2. Feature hierarchy with Radical node.

The sisterhood relationship can also be taken as specifically predicting that complex segments combining any major articulator pair or triple, and the combination of all four, are to be expected in human languages. That is, it predicts that all of the categories of double articulations listed in the first column of Table 1 are possible, and that complex segments such as the stops with two simultaneous closures listed in the third column might well be found. We will argue that this much is correct: all double combinations of the major place nodes are likely human speech sounds. Furthermore, the combination of, say, two differing labial articulations is excluded, since the difference between bilabial and labiodental depends on a feature below the Labial node.

Table 1: Double articulations

Double Articulation	*Categorial example*	*Stop representative*
Labial-Coronal	bilabial-alveolar	t͡p
Labial-Dorsal	bilabial-velar	k͡p
Labial-Radical	bilabial-pharyngeal	p͡ʔ
Coronal-Dorsal	alveolar-velar	t͡k
Coronal-Radical	alveolar-pharyngeal	t͡ʔ
Dorsal-Radical	velar-pharyngeal	k͡ʔ

The feature geometry also predicts that the multiple articulations listed in Table 2 can occur. We will argue that this interpretation of the relationship between independent articulators allows for more combinations than are linguistically possible, since robust phonetic cues cannot be provided for multiple articulations. In other words, we claim that no segments with more than two primary linguistically relevant articulations will be found.

Table 2: Multiple articulations

Multiple Articulation	*Categorial example*	*Stop representative*
Labial-Coronal-Dorsal	labial-alveolar-velar	t͡p͡k
Labial-Coronal-Radical	labial-alveolar-pharyngeal	t͡p͡ʔ
Labial-Dorsal-Radical	labial-velar-pharyngeal	k͡p͡ʔ
Coronal-Dorsal-Radical	alveolar-velar-pharyngeal	t͡k͡ʔ
Labial-Coronal-Dorsal-Radical	labial-alveolar-velar-pharyngeal	p͡t͡k͡ʔ

We will support this claim by examining how the presence of a double articulation is signaled at the phonetic level, and showing that the phonetic cues involved cannot be used effectively to signal the presence of three or four simultaneous articulations. Before we proceed, some limitations on what we are claiming should be noted. Our concern is only with linguistically *simultaneous* articulations, not sequential articulations such as /tk/ in Shona (Doke 1931) or Kinyarwanda (Jouannet 1983). Using terms proposed by Sagey (1986), such segments can be distinguished as having a contour place structure rather than a complex one, as affricates and prenasalized stops have contours for [continuant] and [nasal] respectively. Also, we are considering only primary consonantal articulations. We are assuming that secondary articulations, together with vowels and semi-vowels, are represented with features that are not dominated by the articulator nodes we are talking about. We would place the vocalic features under a node that might either be a sister to the major consonantal articulator nodes, as in Figure 3, or might be dominated by major class features that determine degree of stricture, since only true consonants have primary place specifications. In either case, specifying a value for, say, [High] or [Back] does not entail the presence of the Dorsal node in the tree. Nor does [Round] entail the presence of Labial. Thus, in this view a labialized alveolar stop is not an example of a Labial-Coronal double articulation.**footnote 1**

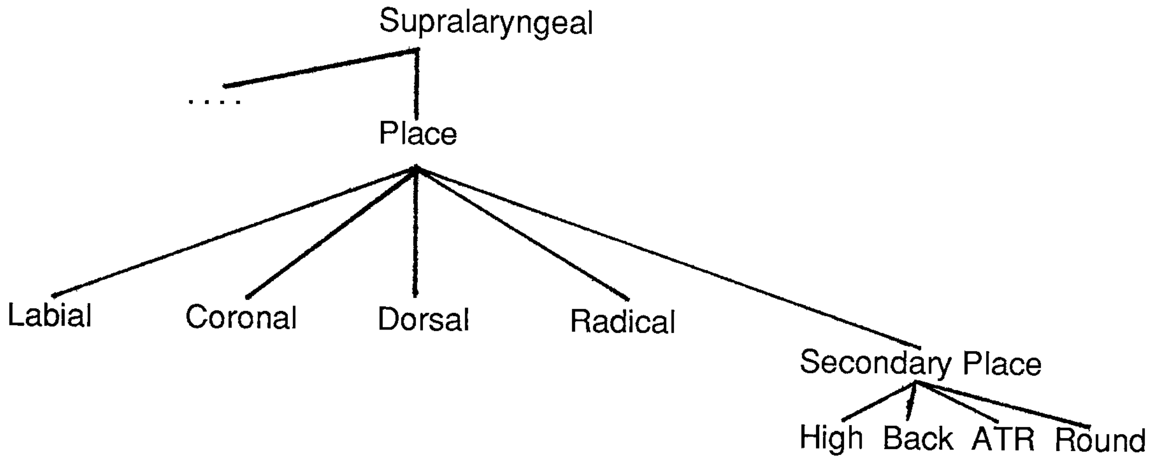


Figure 3. Feature hierarchy with secondary place node.

Phonetic cues for labial-velar articulation

We will present our case based principally on analysis of labial-velar stops, one possible Labial-Dorsal combination, since these are the most common doubly articulated segments in the world's languages. Languages with bilabial-velar stops are especially common in West Africa and northern Central Africa, where they occur in several different families. The languages concerned are numerous and include, for example, Idoma, Yoruba, Eggon, Ibibio (all Niger-Kordofanian), Gwandara (Afro-Asiatic), and Logbara (Nilo-Saharan). Sounds of this type are also found in several New Guinea languages, such as Kate, Ono, Mape, Dedua and Yeletnye. Some of the languages with labial-velar stops also have labial-velar nasals. Examples of the three doubly-articulated labial-velar segments in Idoma are given in Table 3.

Table 3. Labial-velar sounds in Idoma

voiceless stop	àk͡pà	"bridge"
voiced stop	àg͡bà	"jaw"
voiced nasal	aŋ͡màa	"body painting"

We aim to identify those phonetic properties which enable labial-velars to be distinguished both from segments with a **single articulation**, and from a **sequence** of segments with **different** places of articulations. One important factor is segment duration. We have consistently found that doubly-articulated stops and nasals have comparable durations to stops and nasals with single articulations. An example is provided by the measurements given in Table 4 comparing the closure duration of /g͡b/, and /b/ in Yoruba. These data are measured from two repetitions of 9 words each, matched for vowels and tones. The durations are quite similar and the difference is statistically non-significant. Comparable duration to the duration of other single segments of the same manner class is clearly one factor which governs the production and perception of doubly-articulated segments.

Table 4: Durations of /g͡b/, and /b/ in Yoruba.

	/ g͡b/	/ b/
Mean duration	132	128
Standard deviation	12.5	17.6

In contrast, clusters are typically from one-and-a-half to two times the length of single segments of comparable type (Haggard 1973, Catford 1977a). Very few languages include both complex labial-velar stops and sequences of velar and labial stops, but the contrast between single segment and sequence is found in Eggon (Maddieson 1981), as the examples in Table 5 illustrate. Spectrograms of the two verbs /k͡pu/ 'die' and /kpu/ 'kneel' are given in figure 4, preceded by the third person singular subject marker /o-/. The sequence here has about twice the duration of the single segment. Unlike in English words such as 'actor' or 'aptly', the first member of a stop cluster is always released in Eggon; hence the clusters are additionally distinguished from single doubly-articulated stops. In cases where labials, velars and labial-velars are involved the second member of the cluster is frequently lenited in more relaxed speech. These features may also be seen in the words illustrated in the spectrograms in figure 4.[footnote 2]

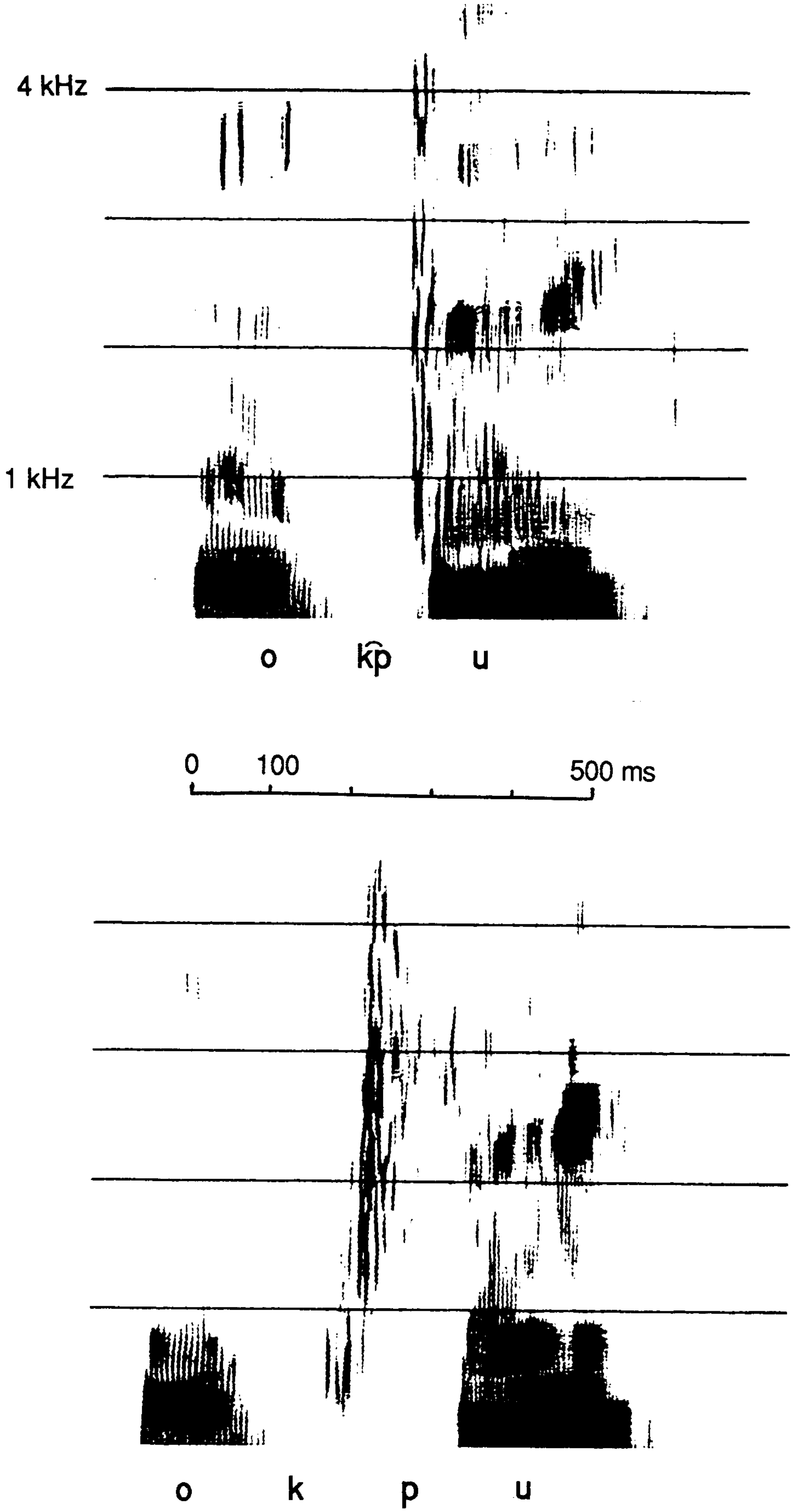


Figure 4. Spectrograms of Eggon utterances /o k͡pu/ and /o kpu/.

Table 5: Labial-velar stops, some single stops and stop sequences in Eggon (tones omitted).

Simple stops		
/p/	/pom/	'pound (v.)'
/b/	/abu/	'dog'
/k/	/aku/	'room'
/g/	/gom/	'break'
Labial-velars		
/k͡p/	/k͡pu/	'die'
/g͡b/	/g͡bu/	'arrive'
Stop sequences		
/kp/	/kpu/	'kneel'
/kb/	/kba/	'dig'
/gb/	/gba/	'divide'
/bg/	/bga/	'beat, kill'
/k͡pk/	/ak͡pki/	'stomach'
/g͡bg/	/g͡bga/	'grind'
/ŋ͡mg͡bb/	/oŋ͡mg͡bbo/	'root (n.)'

Their briefer duration disambiguates labial-velars from sequences but does not distinguish them from single labial or velar articulations. The presence of a complex articulation can be detected by localized auditory cues. In the majority of intervocalic labial-velar stops we have heard, the transition from preceding vowel to stop sounds velar, while the transition from stop to following vowel sounds labial.

The impression is that although the closures overlap for most of their duration, the onset of the velar closure precedes the onset of the labial one by a very brief time, as shown schematically in (a) in Figure 5. It is then released before the labial one, so that labial characteristics dominate the release. This enables the presence of two closures to be detected. Note that if the duration of one closure was contained within the duration of the other, as represented in (b) in Figure 5, the briefer articulation could have minimal acoustic consequences as long as only pulmonic air is involved, and its presence would be unlikely to be detected. Examination of acoustic transitions and burst spectra from a number of languages confirms our auditory impressions. As can be seen in the spectrogram of the Yoruba word [ig͡bi] "storm" in figure 6, the acoustic transitions into and out of the labial-velar stop, instead of being mirror images, are quite different from one another. The onset to the labial-velar has more of the character of a velar closure while its release has labial characteristics. In addition to inferences from acoustic data, the asynchronous timing of the closures can be observed in some of the articulatory and aerodynamic observations that have been made on labial-velars. For example, a difference in the articulatory timing of the closures can be seen in our cine x-ray film of an Idoma speaker. The lowering of the tongue, breaking the velar contact, typically precedes the lip release by one frame. Aerodynamic data indicating earlier velar release in Ibibio /k͡p/ are discussed by Connell (1987).

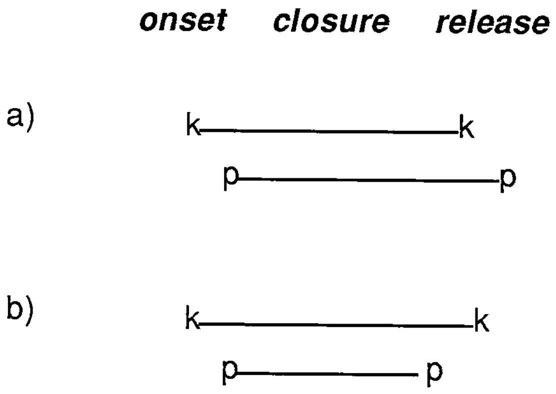


Figure 5. Relative timing possibilities of two closures

This pattern of velar onset and labial release is typical not just of the well-known West African languages with labial-velars, but is found in the New Guinea languages with them as well. Examples from Dedua contrasting /p/, /k/ and /k͡p/ are given in figure 7. This figure includes plots of the spectrum of the release burst of each of the words shown on the spectrograms. The acoustic similarity of the labial-velar /k͡p/ burst to the labial one on the left rather than to the velar one on the right is evident.

Note that the slight asynchrony of the closures we are describing here is a question of phonetic detail. We are not suggesting that labial-velars are phonologically velar at the left edge and labial at the right. Nasal place assimilation shows that this is wrong, since both places are spread to an adjoining nasal in these languages; e.g. the syllabic nasal which marks the progressive aspect in Yoruba is [ŋ͡m] before /k͡p/ and /g͡b/. Rather, the timing asymmetry is a way of providing a robust phonetic cue to the presence of two closures in the segment.

These timing features would seem to be sufficient to distinguish labial-velars from labial + velar sequences on the one hand and from simple labial or velar stops on the other. However, there are frequently also aerodynamic features which distinguish labial-velars from singly articulated stops. In the labial-velar stops of many West African languages the air pressure between the two closures is rarefied, much as it is in a click

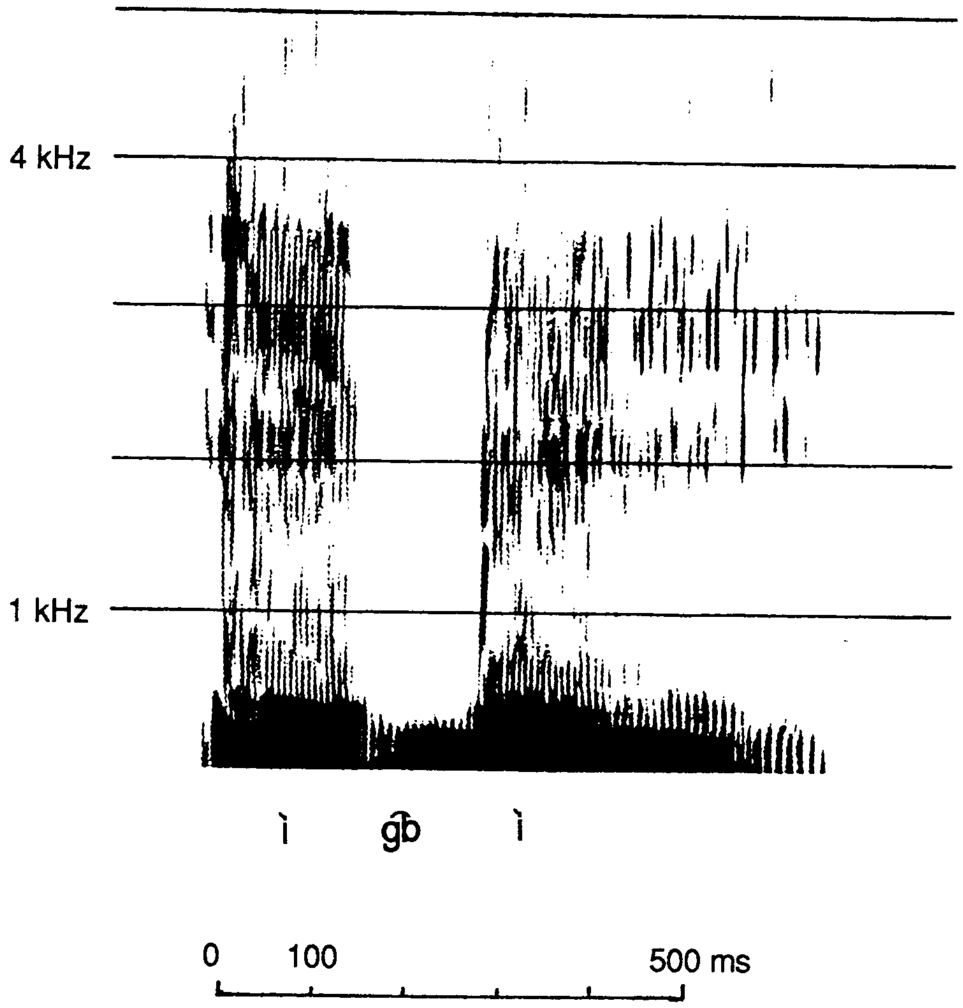


Figure 6. Spectrogram of Yoruba word /ìg͡bì/. Note convergence of second and third formants toward a frequency of about 2400 Hz at time of closure onset, compared with rising transitions from a lower frequency at release.

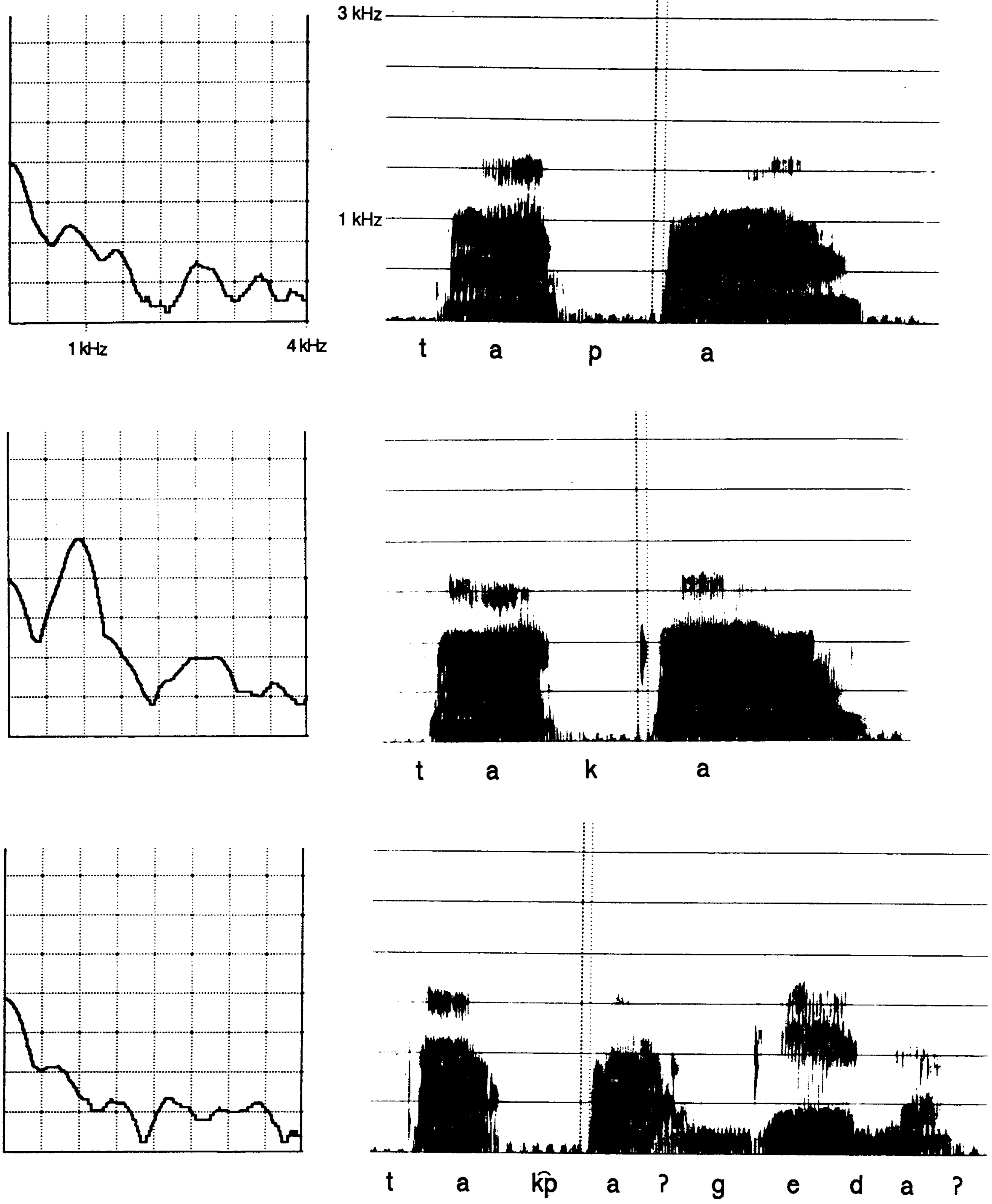


Figure 7. Spectrograms of Dedua words containing /p/, /k/ and /k͡p/ flanked by the vowel /a/. The left panels show the acoustic spectrum of the release burst, calculated over the window enclosed between the dotted vertical cursors on the corresponding spectrogram. Calibration lines are at 500 Hz intervals.

(Ladefoged 1968, Connell 1987). In these cases, if the labial closure is released before the velar one, an ingressive stop is produced, meaning that a different airstream is used for the labial release. In this way a different robust cue for the complex articulation can be made, and the timing **is** as shown in Figure 5(b) above. The click segments of Khoisan languages, and their Bantu neighbours such as Zulu (Doke 1926), are extreme examples of the use of separate airstreams to cue the presence of two different closure locations. The place of the back closure for a click, contrastively velar or uvular for a language such as !Xóõ (Traill 1985), is cued by transitions on the pulmonic airstream. The place of the front closure is apparent from the release involving the velaric suction mechanism.

Other articulator pairings

Cues of the types we have discussed — slight timing asymmetries or use of a supplemental airstream — are potentially available to signal the presence of any pairing of two articulators, except that rarefaction of the air enclosed between a Labial and a Coronal articulation is likely to be unacheivable or of negligible consequence. Interestingly, Labial-Coronal segments are very rare.

Phonemic segments with simultaneous Labial and Coronal closures do occur in Yeletnye, the language of Rossel Island in Papua New Guinea, though not as far as we are aware anywhere else.**footnote 3** In Yeletnye, there are plosives and nasals at bilabial, front alveolar, slightly post-alveolar and velar positions. A bilabial articulation can co-occur with the three other places used for stops, producing labial-alveolar, labial-postalveolar and labial-velar stops and nasals. Examples are given in table 5 based on our own fieldwork (Maddieson 1988) supplemented by Henderson & Henderson (1987). Spectrograms of the words meaning "lung", "horn" and "bag" from this table, prefixed with the first singular possessive marker /a-/, are shown in figure 8.

Table 5: Single and double articulations in Yeletnye.

	bilabial	alveolar	postalveolar	velar
voiceless stop	paa "side"	taa "knife"	ṭoo "tongue"	kaa "spear"
prenasalized voiced stop	mbee "carry"	nde "food"	ṇḍe "firewood"	ŋgaa "(tree)"
voiced nasal	maa "road"	nii "juice"	ṇaa "feast"	ŋa "please"

	labial-alveolar	labial-postalveolar	labial-velar
voiceless stop	t͡pənə "lung"	ṭ͡panə "horn"	k͡pɛnɛ "bag"
prenasalized voiced stop	n͡md͡boo "pulp"	ṇ͡mḍ͡boo "many"	ŋ͡mg͡bo "fog"
voiced nasal	n͡mo "bird"	ṇ͡mo "we"	ŋ͡mo "breast"

As indicated above, we treat clicks as having multiple places of articulation, following Halle (1983), Sagey (1986) and others. Since this view might strike some phoneticians as heretical, it is discussed more fully in an appendix to this paper. Dental, alveolar and postalveolar clicks provide the only examples of Coronal-Dorsal double articulations that we are aware of, but these sounds are very far from being marginal in the languages which have them (cf. Traill 1985).

We will now consider stop combinations involving the Radical articulator. Simple Radical stops are quite rare, though they occur in a few languages such as the Dagestanian language Agul, which has the epiglottal stop [ʡ]. Since relatively few linguists are familiar with sounds of this type, spectrograms of a pairs of words with epiglottal stops are given in Figure 9.

Because Radical stops are rare we would predict that doubly-articulated stops including a Radical closure would be even rarer, and we know of no language containing one. But there **are** a number of languages with stops with a **secondary** constriction of the pharynx. It does not seem improbable that, through an intensification of this secondary pharyngeal articulation accompanying an alveolar stop (as in Arabic) or a bilabial or uvular one (as in Ubykh), a language could develop Labial-Radical, Coronal-Radical or Dorsal-Radical stops. We therefore think that the feature hierarchy is correctly structured to include the occurrence of segments with these double articulations.

Triple articulations

On the other hand, we do not know of any linguistically contrastive segment which is regularly produced with a triple closure, and we would argue that none should be expected, given the nature of the phonetic cues available. Consider what would happen in a stop segment if three closures were involved, as in the putative labial-alveolar-velar stops shown in figure 10. The onsets of the three closures can be staggered as in (a), but this does not produce a clear indication of the complex articulation involved. A closure entirely contained within the duration of a longer closure — as the alveolar closure is contained within the longer duration formed by the overlap of velar and labial ones here — will have no acoustic transitions in adjoining segments. Rescheduling the timing, as in (b) for example, does not help, since now the labial closure is contained within the others. Under no possible arrangement can more than two transitions occur, one at each segment edge. So the timing asymmetry does not provide a transitional cue to the presence of a third articulation.

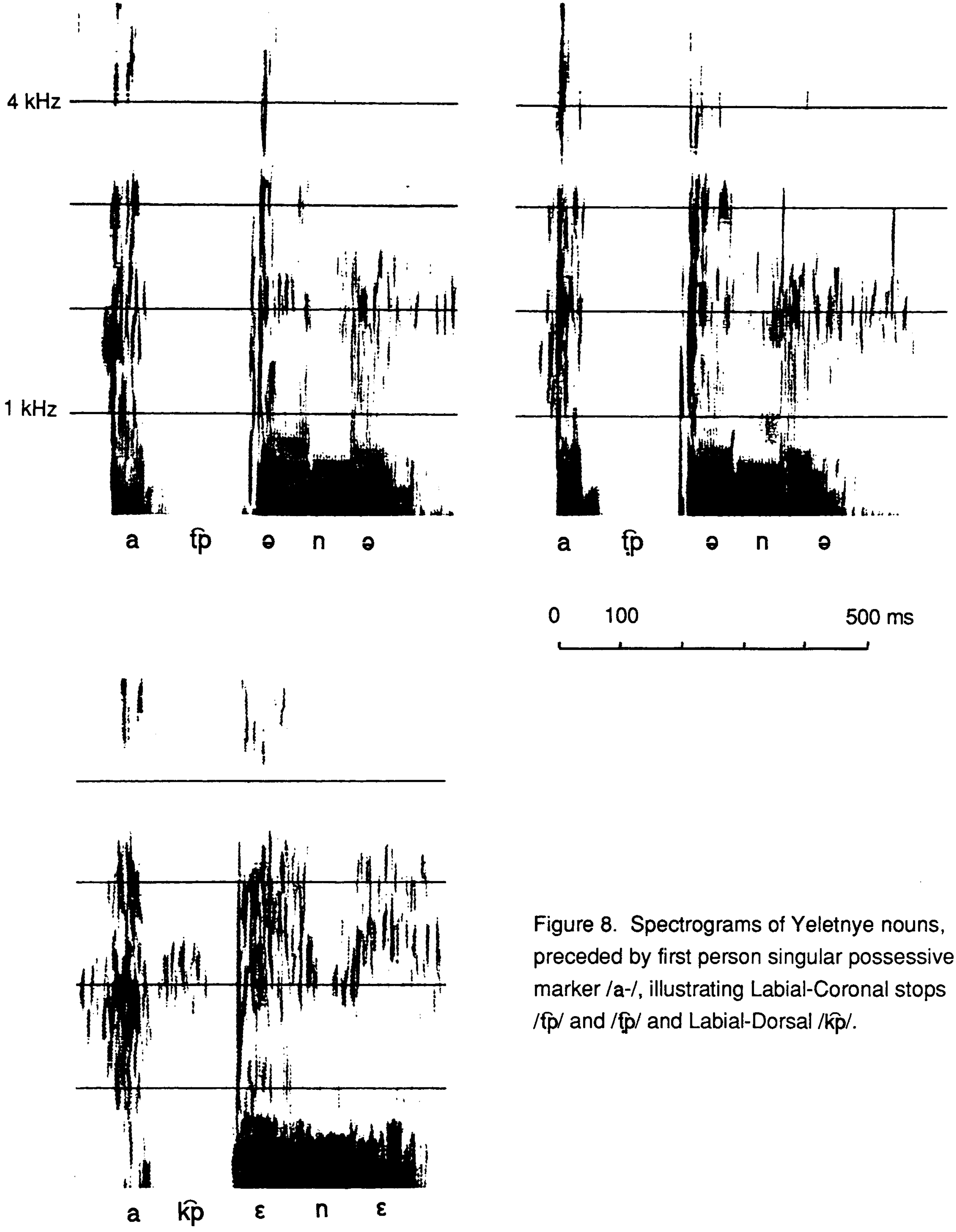


Figure 8. Spectrograms of Yeletnye nouns, preceded by first person singular possessive marker /a-/, illustrating Labial-Coronal stops /t͡p/ and /ṭ͡p/ and Labial-Dorsal /k͡p/.

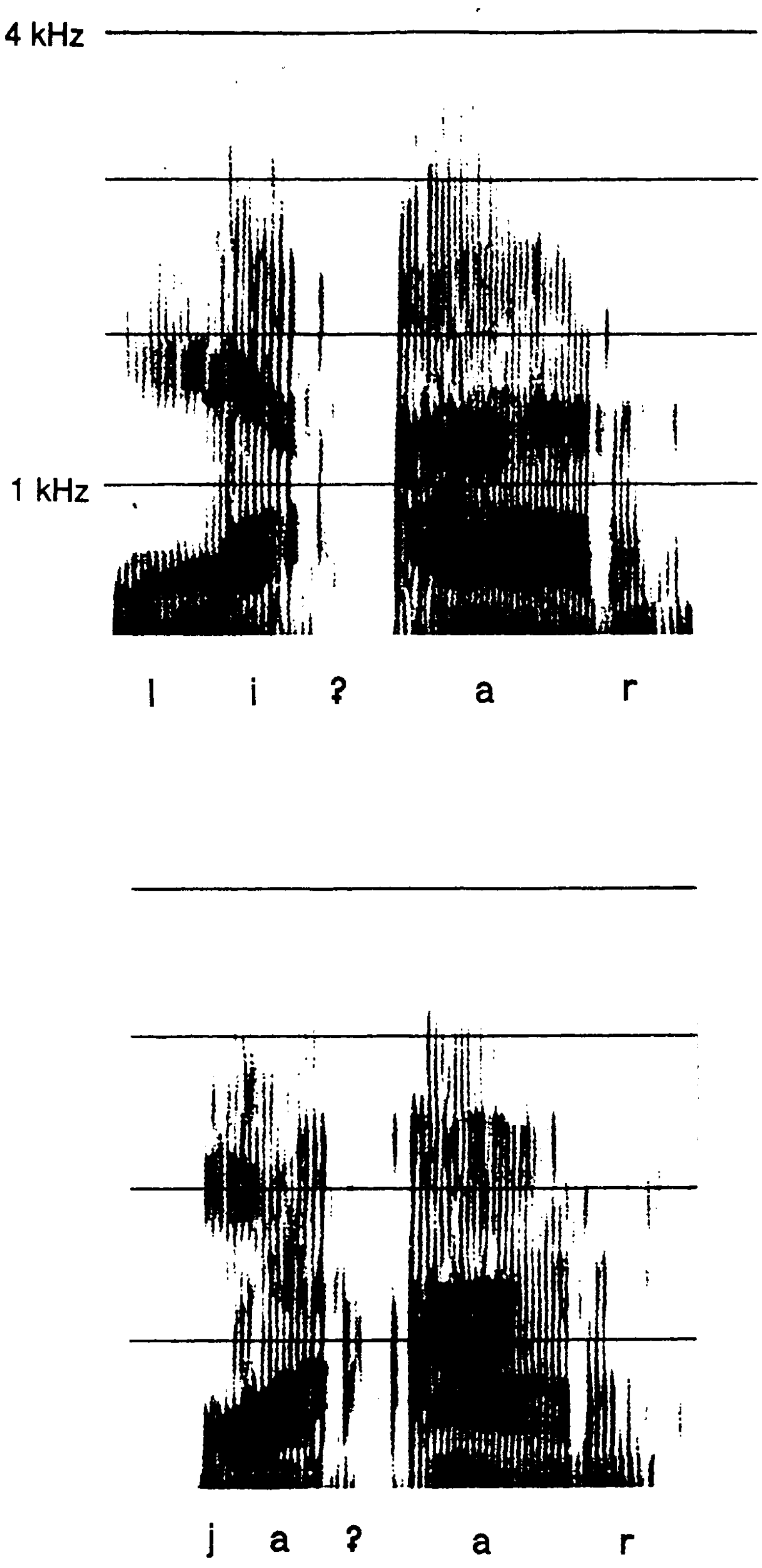


Figure 9. Spectrograms of two Agul words with medial epiglottal stops (from an illustrative tape provided by S. Kodzasov). Note the marked transitional raising of the first formant which is characteristic of this articulation.

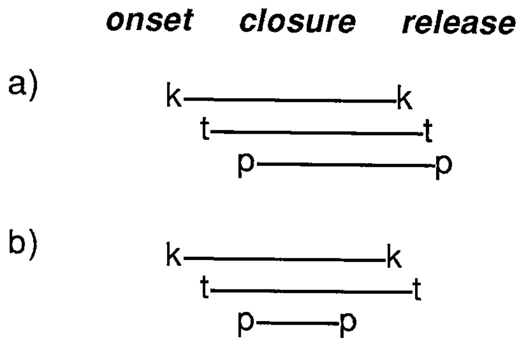


Figure 10. Some relative timing possibilities among multiple closures.

Stops can also be differentiated by characteristics of their release bursts, as we showed with Dedua. When a plosive with multiple closures is produced, only the last closure to be released can have the full character of a burst. A back closure released while another is maintained further forward in the mouth will not radiate the burst energy, whereas a forward one that is released while another is maintained further back in the mouth will not have a pressure build-up behind it: Hence it will have little or no acoustic energy. No arrangement of three overlapping closures can avoid one or other of these situations, so place cues inherent in bursts will be ineffective supplements to the transitional cues.

The limitations on effective clues to multiple simultaneous articulations do not only concern plosives but in varying ways they also apply to other classes of consonants. We will briefly discuss nasals and clicks. In nasals only two transitions are available, as for stops, but place of articulation can be differentiated by characteristics of the nasal murmur itself. The particular quality of the nasal murmur depends principally on the size of the oral cavity in front of the velopharyngeal port. When there are multiple oral closures, the size of this cavity is determined by the closure closest to the velar region. During concurrent articulations there is therefore a cue in the quality of the nasal murmur only for the furthest back closure. But if the onsets and offsets of the closures are staggered, changes in the quality of the nasal murmur might provide cues to multiple articulations as closures are added or released, provided that they are programmed in a sequence that does not produce a masking of a forward closure by a back one. For example, if the timing pattern shown in Figure 10 (a) represents a nasal in which closure 1 is **labial**, closure 2 is **alveolar** and closure 3 is **velar**, then at the onset of the nasal a shifting of the nasal murmur from a quality characteristic of labial place to a quality characteristic of alveolar place and finally to a quality characteristic of velar place might be observed. The brief alveolar portion would potentially provide the cue to the presence of an articulation which would not be cued by transitions to and from surrounding segments (the onset transition would be labial, the offset transition velar.)

We do not know how long this portion would need to be in order for its presence to be noticeable, but, given the well-documented difficulty of discriminating between nasal murmurs (Malécot 1956, Nord 1976), we believe that it would need to be quite long or its presence will be missed. And if it is long enough to be perceived, then a conflict arises with the principle that single complex segments have duration comparable to that of simple segments of the same phonetic class in the same environment. Instead of a single multiply-articulated nasal, a sequence of several nasal segments may be perceived (and production adjusted to that program). In either case, it is easy to see that circumstances do not favor the development and stability of triply-articulated nasals any more than plosives.

As for clicks, there are two ways in which they could be produced with triple closures. One way would be to have two closures, one Labial and the other Coronal, in front of the velar or uvular closure required for the click mechanism. Normally, one of the front closures would be acoustically ineffective. However, these closures could be timed so that the first is formed just before the back closure is made, thus providing a transitional cue on the pulmonic airstream to one front closure. If this closure is also released first, the second forward closure would then be the place of articulation of the velaric airstream, and the back closure would be released into pulmonic airflow. All three closures would then have audible cues to their presence. However, we know of no clicks that are produced without the back closure entirely overlapping the forward oral closures in duration, and we suspect that this is a requirement for their efficient production. When such an overlap occurs, and two forward closures are formed, only the second to be released is likely to produce an audible burst. The presence of the other closure contributes little or nothing.

In fact, Traill (1985: 103) shows that two of five !Xóõ speakers whose production of clicks with bilabial closures he examined using palatography also formed a closure in the dental/alveolar region during the click. Tracings of these palatograms are shown in figure 11. Although he says that these "bilabial-dental" clicks have "three points of articulation" there is no indication that these variants sound any different from the bilabial clicks of other speakers. From his discussion, it seems clear that the dental contact is released well before the labial one is, since cavity expansion for labial clicks is achieved "by a lowering and retraction of the front part of the tongue" (Traill, 1985: 106), which must occur before labial release. Tracings from x-ray films of the same two subjects who show dental closure on the palatograms indicate that the labial closure is indeed maintained when the tongue tip is well away from the upper teeth. These tracings are reproduced as figure 12. The third closure for these clicks is consequently

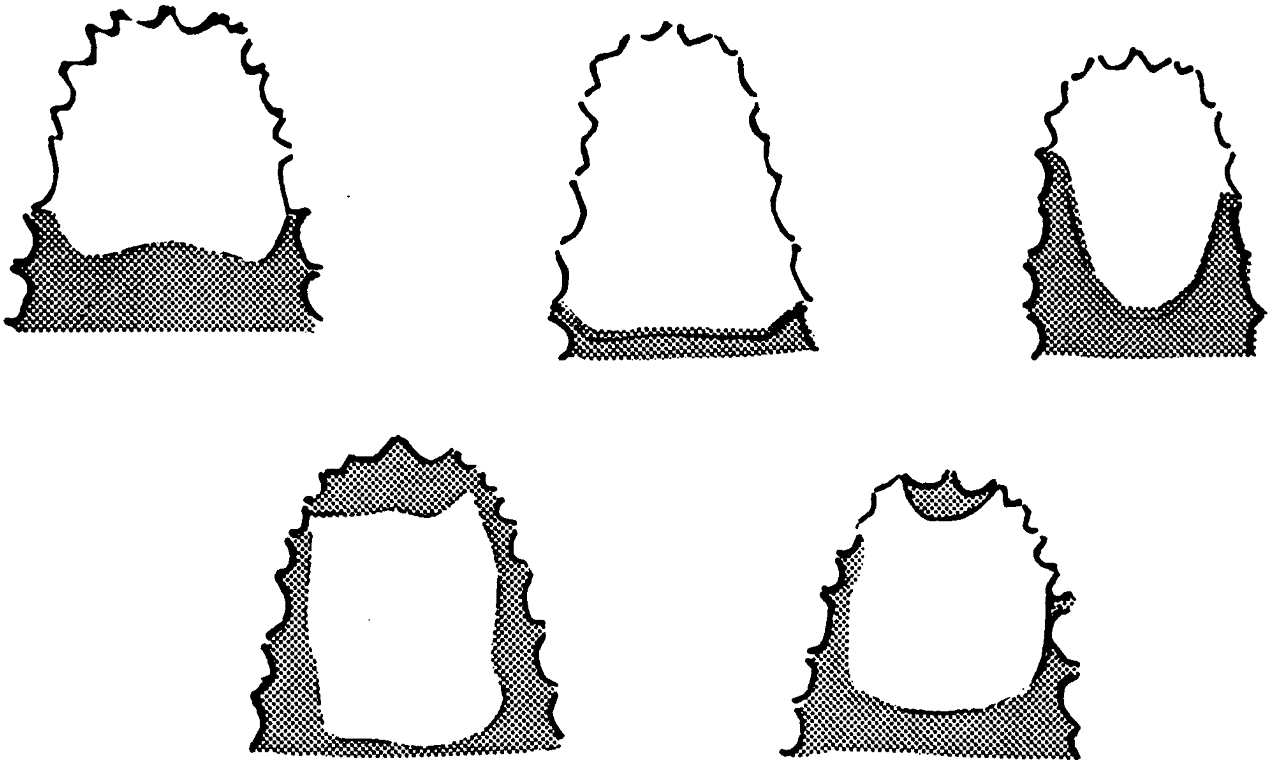


Figure 11. Palatograms showing articulatory contacts (shaded areas) in the dental-to-velar area for five speakers of !Xóõ producing bilabial clicks (after Traill 1985). The three speakers in the top row show no contact of the forward part of the tongue, but the two speakers in the bottom row show contact in the dental region. (Note that the labial closure is not represented in palatograms.)

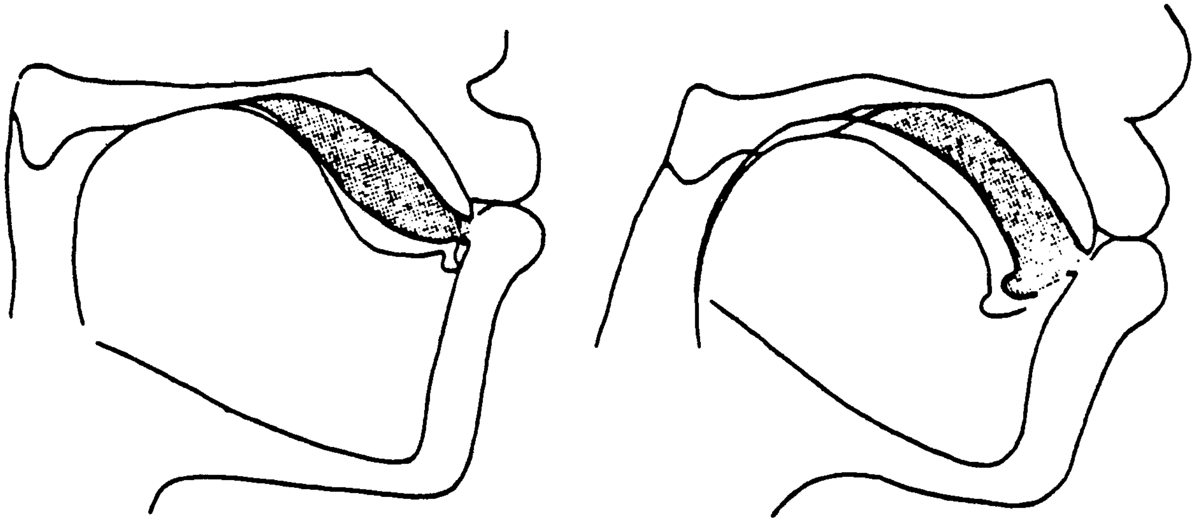


Figure 12. Tracings from x-ray films of bilabial click production by the two !Xóõ speakers with front tongue contacts in the palatograms of Figure 11 (after Traill 1985). The shaded areas show the smallest air pocket between the labial and velar closures observed on any frame in the sequence. The lower tongue outline on these tracings shows the lowering of the tongue in a subsequent frame of the film. This precedes labial release and contributes to the rarefaction of the air between the closures.

of no linguistic significance, as we would predict.

The second way that a click could be formed with three closures is for a Radical closure to be added behind the back closure for the click. We know of no cases with a Radical closure, but if they did exist it is easy to see that potential cues to the place of the velar or uvular closure would be obscured by the earlier formation of a closure further back. If a Radical closure was formed later than the velar or uvular one, no transitional cue to its presence would be heard, neither would sustained nasal airflow be possible as a means of indicating place.

There is not time in this paper to extend this discussion to all possible ways of combining three articulations, nor to consider all the different phonetic factors that would need to be considered for different every type of nasal, click, or other segment class. But in each case we have considered we are satisfied that for reasons of the kind outlined here a satisfactory signal of the complexity of the articulation is unlikely to be achieved. One of the articulations will make no acoustic contribution, or will require excessive duration to be percieved as simultaneous with the others. Hence a contrastive function for segments with three oral closures is not to be expected.**footnote 4** Our arguments will apply *a fortiori* to segments with four simultaneous closures.

Conclusion

In conclusion, then, we propose that the primary place node should have four branches for the four major articulators. There are no physiological limits on the ways that these may be combined. But phonetic constraints require our linguistic theory to stipulate that there may be no more than two simultaneous primary articulations in a single segment.

Acknowledgments

This report forms part of a larger project, funded by the National Science Foundation through grant BNS 87-20098, examining the phonetic basis of phonological features. Thanks to members of the UCLA Phonetics Laboratory group for comments and suggestions during an earlier presentation of this paper.

Footnotes

1. The assumption that there is a secondary place node makes the formulation of the constraint on articulator combinations that we have in mind very straightforward. However, there are obvious connections between the major articulator nodes and some of the secondary articulation features. If these are represented by subordinating secondary articulations to primary articulator features (e.g. [round] dominated by Labial), the constraint can still be formulated, albeit in a more cumbersome fashion.

2. Lenition seems not occur in clusters containing alveolars (e.g. in the words /atku/ 'calabash', /odga/ 'leg'), suggesting that this may be a further strategy for marking the potentially ambiguous clusters.

3. Some other languages have been reported with Labial-Coronal sounds; [t͡p] occurs as an allophone of /k͡p/ before front vowels in Dagbani (Wilson & Bendor-Samuel 1969, Ladefoged 1968) and Nzema (Berry 1955, Chinebuah 1963), and as a variant of /tʷ/ for at least some speakers of Abkhaz (Catford 1972) and Lak (Khaidakov 1955, Catford 1977b). Here Catford describes the labial contact as "light"; moreover, in these Caucasian cases the labial component involves considerable forward protrusion of the lips and the contact is between the inner surfaces. Photographs taken by Catford show this is quite different from the normal contact for /p/, and it might be more justifiable to consider this gesture as phonetically a secondary articulation, since it seems related to lip rounding. Two Chadic languages frequently cited as having labial-alveolars, Margi and Bura (Hoffman 1963, Ladefoged 1968, Halle 1983), on closer examination prove to have labial+alveolar sequences rather than double articulations (Maddieson, 1983, 1987). For example, /bd/ is considerably longer than /b/ or /d/ and a labial release can be detected well before the alveolar one. If a vowel precedes /bd/, the /b/ element of the cluster forms a coda to the preceding syllable, and the vowel is shortened accordingly.

4. Although he did not know of an example, Halle (1983) anticipated that triply-articulated Labial-Lingual-Dorsal segments would be found. Sagey (1986) suggested that Kinyarwanda /tkʷ/ filled this gap. Her arguments that this and similar elements in Kinyarwanda are single segments rest in part on the fact that clusters in loanwords are broken up by epenthesized vowels; since /tkʷ/ and like elements are tolerated they must not be clusters. However, [tk], [dg], [bg], [fk] etc are phonetically speaking sequentially articulated. If they are single segments, it seems necessary to represent them as contour segments, if only to distinguish between e.g. [bg] and [g͡b], i.e. segments with sequential and simultaneous production of the same places of articulation. Evidence that the complex elements of Kinyarwanda are *phonologically* sequential includes such facts as that the nasal preceding one of the (voiced) doubly-articulated stops has only the single articulation of the first element, e.g. [n] before [dg], [m] before [bg]. Jouannet (1985) argues that all these complex elements are best analyzed as sequences largely on the basis of distributional patterns.

Appendix on place of articulation in clicks

Halle (1983) suggested that the Coronal-Dorsal combination predicted by the hierarchical arrangement of place features is to be found in click articulations, such as the dental clicks in the Zulu word /ʇaʇa/ "climb". Clicks, of course, require a double closure for their production. However, the velar closure involved in producing the clicks in this word is usually regarded by phoneticians solely as the basis of the airstream mechanism employed, which is consequently called the velaric airstream mechanism (Pike 1943). The place of articulation given for a click has traditionally referred only to the location of the front closure, and the back closure has not been considered as a place of articulation. From a phonological viewpoint, the back closure in Zulu clicks can be considered to be entailed by whatever feature specifies the airstream mechanism and hence not to require any phonological feature specification.

We agree with those who have considered this traditional phoneticians' view to be mistaken in allocating only one role to the back closure and accept that clicks have two places of articulation, one of which is provided by the back closure. A precursor of this point of view is found in Chomsky & Halle (1968: 319), who argued that clicks are stops produced with "extreme velarization" and assigned the features [+high, +back] to them. These features are shared by velar and velarized sounds in the SPE scheme, and so clicks are described as velar in place. (Although Chomsky & Halle label the back closure in clicks a secondary articulation, there is no mechanism in SPE to distinguish between primary and secondary articulations. Bilabial clicks and labial-velar plosives have the same place of articulation features.) The one argument cited by Chomsky & Halle for their view is that in Nama the back closure can vary in its manner of release, and that the resulting abrupt versus affricated contrast is an articulatory difference. There are at least two other points that can be added. In Bantu languages with clicks, such as Zulu and Xhosa, nasal + stop clusters must be homorganic. Prefixes which contain a final nasal undergo a place assimilation rule when the nasal abuts a stop (or almost any other consonant). When such a prefix abuts a click, the nasal is not assimilated to the dental, alveolar or palato-alveolar place of the front closure of the click but rather to the velar place (Doke 1926: 78). Illustrative examples from Zulu are given in table A1. (Transcription of clicks is with the traditional IPA symbols, rather than those based on the Lepsius alphabet usually used for Khoisan languages.) Doke argues that before a click "with the back of the tongue touching the velum .... the nasal homorganic to the clicks must be the pure velar nasal [ŋ]." While it is true that when a nasal is *simultaneously* articulated with a click the effective closure for that nasal must be the back closure for the click, it is important to note that a nasal *preceding* a click need not necessarily be produced with a velar articulation. The velar closure could be formed later than the front closure, i.e. not until the click itself is being produced. Given what actually happens, it is clear that the velar closure is not acting solely as the initiator of the airstream for the click but is also functioning as a place of articulation whose influence spreads to the adjacent pulmonic nasal segment.

Table A1. Nasal assimilation in Zulu noun class prefix -izin.

	singular	plural	
plosives	u-phaphɛ	izim-paphɛ	"feather(s)"
	u-thi	izin-ti	"stick(s)"
	u-gu	iziŋ-gu	"river-bank(s)"
clicks	u-ʇhuʃɛla	iziŋ-ʇuʃɛla	"sharp instrument(s)"
	u-ʗhududu	iziŋ- ʗhududu	"tall careless person(s)"
	u-ʖhuʃɛla	iziŋ-ʖuʃɛla	"sharp instrument(s)"

The second additional argument is based on a contrast that occurs in some of the Khoisan languages such as !Xóõ, #H´õa and ||Ani. In these languages the back closure itself may differ in place, being either velar or uvular. Contrasts from !Xóõ are illustrated in table A2.

Table A2. Some contrasting clicks in !Xóõ (from Traill 1985).

	velar back closure		uvular back closure	
bilabial	ʘôõ	"dream"	ʘqóu	"wild cat"
laminal dental	ʇâa	"move off"	ʇqàa	"rub with hand"
apical alveolar lateral	ʖaã	"poison"	ʖqàã	"tooth"
laminal palatal	ʗàha	"knock"	ʗqàa	"conceal"

The contrasting positions of the back of the tongue in two !Xóõ syllables with palatal clicks are illustrated by the x-ray film tracings in figure A1 (after Traill 1985: 127-8). In this pair some part of the difference in tongue position is doubtless associated with the difference in the vowel following the clicks. However, in the syllable /ʗqo/ the back of the tongue remains in contact with the uvula from the formation of the closure until the moment of release, whereas it does not contact the uvula at any point in the syllable /ʗe/. Not only does !Xóõ contrast back tongue positions in this way but the language also contrasts abrupt with affricated release of the back closure in both velar and uvular places, and has contrasting velar and uvular nasal accompaniments to the clicks. Hence, the back closure in clicks is an articulatory

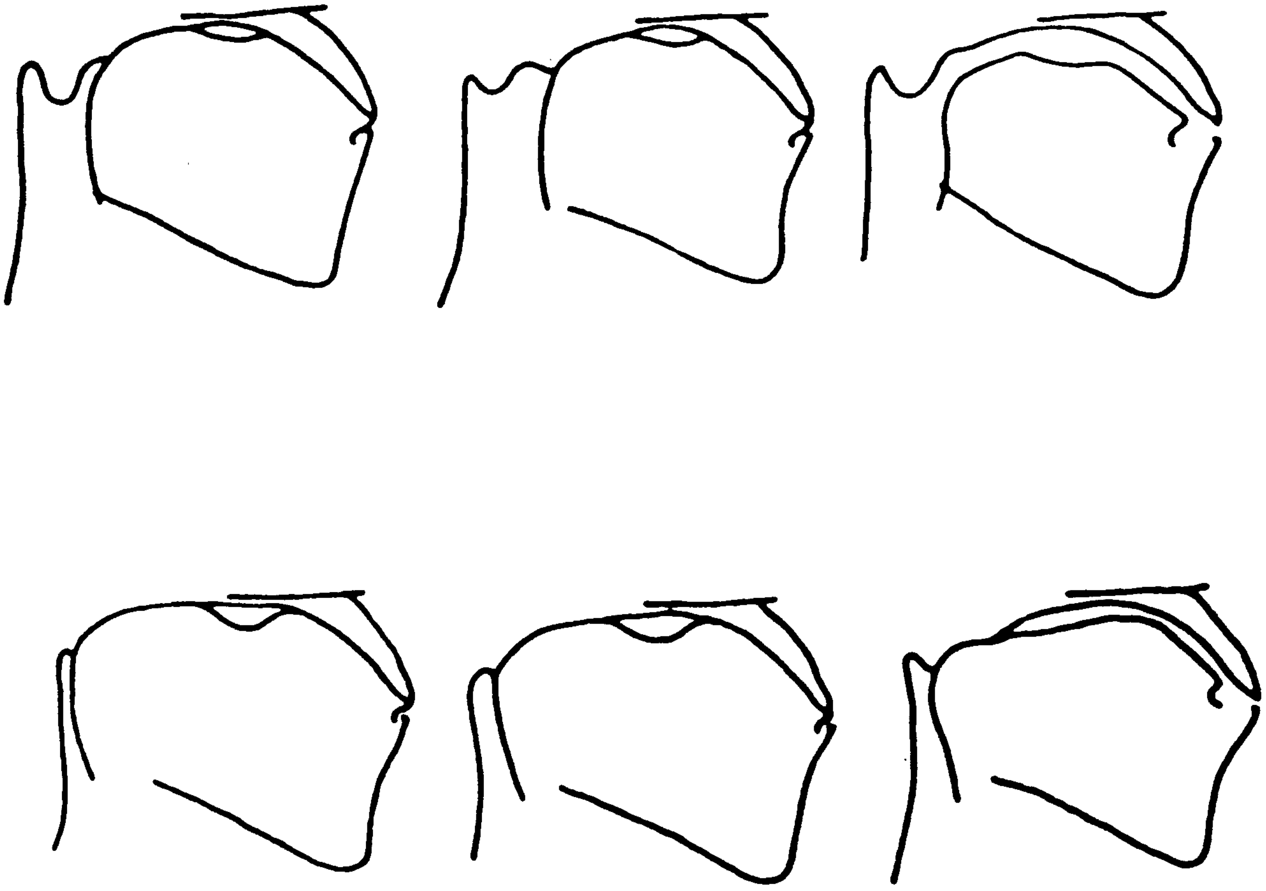


Figure A1. Tracings of frames from x-ray films of a single speaker of !Xóõ producing the syllables /ʗe/ (top) and /ʗqo/ (bottom). Three frames, chosen from the interval which includes the release of the front closure, are shown for each syllable. The frames are 20 ms apart in time. The tracings show the outline of the mandible and do not include the soft tissue of the lips, jaw, etc.

gesture with contrasts in both place and manner dimensions, even though it also forms the basis of the airstream mechanism.

References

Berry, Jack. 1955. Some notes on the phonology of the Nzema and Ahanta dialects. *Bulletin of the School of Oriental & African Studies* 17: 160-165.

Catford, John C. 1972. Labialization in Caucasian languages, with special reference to Abkhaz. *Proceedings of the Seventh International Congress of Phonetic Sciences*, ed. A. Rigault & R. Charbonneau. Mouton, The Hague: 679-682.

Catford, John C. 1977a. *Fundamental problems in Phonetics.* Indiana University Press, Bloomington.

Catford, John C. 1977b. Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology* 6: 293-314.

Chinebuah, Isaac K. 1963. The category of number in Nzema. *Journal of African Languages* 2: 244-259.

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

Connell, Bruce. 1987. Temporal aspects of labiovelar stops. *Work in Progress (Department of Linguistics, University of Edinburgh)* 20: 53-60.

Doke, Clement M. 1926. *The Phonetics of the Zulu Language (Bantu Studies Special Number).* University of the Witwatersrand Press, Johannesburg. Reprinted 1969, Krauss Reprint, Nendeln, Liechtenstein.

Doke, Clement M. 1931. *A Comparative Study in Shona Phonetics.* University of the Witwatersrand, Johannesburg.

Haggard, Mark. 1973. Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics* 1: 9-25.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1: 91-105.

Halle, Morris. 1988. The immanent form of phonemes. In *Giving birth to cognitive science: a festschrift for George A. Miller* (ed. W. Hurst). Cambridge University Press. Cambridge.

Henderson, James, & Anne Henderson. 1987. *Nt:u Kópu Dyuu U Puku Dmi (Rossel Dictionary).* Dictionaries of Papua New Guinea No 9. Summer Institute of Linguistics, Ukarumpa.

Hoffman, Carl. 1963. *A Grammar of Margi.* Oxford University Press, Oxford.

Jouannet, Francis. 1983. Phonétique et phonologie. Le système consonantique du kinyarwanda. In *Le Kinyarwanda: Études Linguistiques* (ed F. Jouannet). SELAF, Paris: 55-74.

Khaidakov, Said M. 1955. Balkharskij Dialekt Lakskogo Jazyka *(Trudy Instituta Jazykoznanija 3).* Akademija Nauk S.S.S.R., Moscow.

Ladefoged, Peter. 1968. *A Phonetic Study of West African Languages (second edition).* Cambridge University Press, Cambridge.

Ladefoged, Peter & Morris Halle. 1988. Some major features of the International Phonetic Alphabet. *Language* 64: 577-582.

Ladefoged, Peter & Ian Maddieson. 1986. *(Some of) The Sounds of the World's Languages (UCLA Working Papers in Phonetics 64).* Phonetics Laboratory, UCLA, Los Angeles.

Maddieson, Ian. 1981. Unusual consonant clusters and complex segments in Eggon. *Studies in African Linguistics, Supplement* 8: 89-92.

Maddieson, Ian. 1983. The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics* 14: 285-310.

Maddieson, Ian. 1987. The Margi vowel system and labiocoronals. *Studies in African Linguistics* 18: 327-355.

Maddieson, Ian. 1988. Fieldwork notes on Papua New Guinea languages.

Malécot, A. 1956. Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique. *Language* 32: 274-

Nord, L. 1976. Perceptual experiments with nasals. *Speech Transmission Laboratory, Quarterly Progress and Status Report* 1976/1-2: 5-

Pike, Kenneth L. 1943. *Phonetics.* University of Michigan Press, Ann Arbor.

Sagey, Elizabeth. 1986. *The representation of features and relations in non-linear phonology.* Ph. D. disssertation, M.I.T., Cambridge, Mass.

Sagey, Elizabeth. 1987. Consonant and vowel processes and the feature heirarchy. Paper presented at Workshop on Distinctive Feature Theory, University of Southern California, Los Angeles, February 1987.

Sibomana, Leo. 1985. A phonological and grammatical outline of Eggon. *Afrika und Übersee* 68: 43-68.

Traill, Anthony. 1985. *Phonetic and Phonological Studies of !Xõõ Bushman.* Helmut Buske, Hamburg.

Wilson, W.A.A. & John T. Bendor-Samuel. 1969. The phonology of the nominal in Dagbani. *Linguistics* 52: 56-82.

An Exploration of Phonation Types in Wu Dialects of Chinese

Cao Jianfen
Institute of Linguistics,
Chinese Academy of Social Sciences

Ian Maddieson
Department of Linguistics
University of California, Los Angeles

Abstract

The three way contrast of initial stops and affricates in terms of voicing and aspiration is well known as the most characteristic feature of Wu dialects of Chinese; however, the phonetic nature of the "voiced" category has been a controversial issue for a long time. The argument earlier concentrated on whether the contrast of "voiced" and voiceless unaspirated consonants was based on the difference between voicing and voicelessness. Recent instrumental studies have revealed that there is no vocal cord vibration during the closure of either type when they occur initially in isolated monosyllabic words or in the stressed position in running speech, so the focus of argument has recently shifted to examining other properties on which the contrast might be based. Against this background, an exploration of the role of breathy phonation has been undertaken. The present study involves four dialects of Wu. Contrasting word pairs with initial "voiced" and voiceless unaspirated stops from a total of 10 speakers were recorded and measurements of the energy difference between harmonics at the onset, middle and offset of the vowel were made. Air flow and air pressure during the closure and release of syllables with bilabial consonants were measured as well. Both these types of measurements have served as tests for phonation type in other research. The results indicate that a phonation type difference does exist between "voiced" and voiceless consonants in Wu dialects. However, this is only when they occur initially in a syllable in isolated monosyllabic words or in the stressed position in running speech. On the other hand, no phonation contrast appears in the case of unstressed position in running speech.

Introduction

The Wu dialects of Chinese are spoken by a population of about 70 million (Yan 1988, Zhengzhang 1988) in a densely populated area covering the southeast of Jiangsu province and the greater part of Zhejiang province. These dialects form one of the major groups of Chinese dialects (Yuan 1960, Norman 1988). The Wu dialects are also phonologically one of the most conservative among these major dialect groups in that they retain the three way contrast of initial obstruents from Middle Chinese (i.e. Chinese of about 7th to 10th centuries B.C. Hereafter we will abbreviate this to M.C.). This three way distinction is reflected in the categories established in the *Qieyun* and can be projected back to even earlier stages of Chinese (Karlgren, 1940; Li, 1971). It has been lost in the other dialect groups of Chinese, which have only two series of obstruents: voiceless aspirated and voiceless unaspirated. The Wu dialects have a series of stops and affricates which are usually labeled "voiced" in addition to these two series. Table 1 shows the close relation beween the stop consonant system of M.C. and Wu dialects.

Table 1. Stop consonant system in M.C. and current Wu dialects.

representative characters	*Middle Chinese* reconstructed segment	*Current Wu dialects* common transcription
幫	*p	p
滂	*p^h	p^h
並	*b	b
端	*t	t
透	*t^h	t^h
定	*d	d
見	*k	k, tɕ
溪	*k^h	k^h, $tɕ^h$
羣	*g	g, dʑ

As in other dialect groups, the "voiced" category is associated with historical tone lowering, but unlike the majority of other dialects, the historical relationship between "voicing" and lowering generally remains fairly transparent in Wu, so that Yin tones—those in voiceless initial syllables—have a higher onset than Yang tones—those in syllables with "voiced" initial. Although the tonal systems vary considerably in number and shapes of tones (Chao 1967), there is usually some redundancy between the voicing categories in obstruents and the tonal categories because of this historical conservatism.

What the actual nature of the distinctions between the three series of obstruents in Middle Chinese were is obscure. There are no contemporary phonetic descriptions, just the division of the *Qieyun* tables into three categories. Though it is certain that there are three phonological distinctions, the phonetic basis of these distinctions must be reconstructed from external or later materials. Studying the three-way contrast in modern Wu dialects can serve as one tool for the study of this aspect of Middle Chinese. These dialects may well preserve phonetic details of the contrast which cannot be recovered any other way. Moreover they may provide more detailed insight into the relationship between consonantal development and tonal development. It has usually been assumed that voicing is the relevant feature which caused the lowering of Yang tones (Hombert, Ohala & Ewan, 1979), but other phonetic factors may also have been involved in the contrast between obstruent categories in Middle Chinese. However, the evidence from Wu dialects does not provide a straightforward answer to the question of what the phonetic factors might be. This is because there has been some dispute over the precise phonetic nature of the "voiced" obstruents in Wu and the extent to which different dialects vary.

In the first place, it has not been agreed that the contrast involves a difference between voicing and voicelessness. Some Chinese linguists believed that the "voiced" consonants in Wu dialects were typical voiced consonants, and this view has appeared in some publications (such as Luo 1956). In most of the common publications, such as Wang (1956), Luo & Wang (1981) and Yuan (1960), this category of consonants is described as being voiceless until the last half of the closure, so these consonants are usually called "half voiceless and half voiced", and transcribed as /**pb**, **td**, **kg**/ to indicate the change in voicing in the middle. More recently Cao (1982) and Shi (1983) have questioned whether there is any voicing at all during the closure of the "voiced" stops, at least when pronounced in isolated words.[1] These studies will be described more fully below.

Many linguists, including those mentioned above, have expressed the view that the "voiced" category is not plain voiced but has a breathy voiced or voiced aspirated component. The details of their descriptions also vary. Karlgren (1915-1926) may have been the first major scholar to suggest that these consonants are accompanied by a weak "aspiration", less strong than that which accompanies the voiced aspirates of northern Indian languages such as Hindi. It was apparently Karlgren's view that the stops were voiced during their closure and their release was characterized by weak voiced aspiration preceding the vowel. On the other hand, Liu (see Chao 1928) and Chao (1928, 1936) reported that vocal cord vibration was not involved during the period of closure, but that some voiced breathiness followed. They considered the "voiced" obstruents in Wu dialects not a true voiced series, but a combined effect of a voiceless stop onset and a murmured release, so they transcribed these consonants as /p^{ɦ}, t^{ɦ}, k^{ɦ}/ etc. Chao (1930) and Li (1986) suggested that this breathiness is riding on the postconsonantal vowel as a kind of synchronous frication. Recently, Ramsey (1987: 91) has described the "voiced" consonants in Shanghai as like the voiced aspirated consonants of Indian languages. In his view:

> "The breathiness is acoustically quite prominent; it pervades the entire syllable, beginning in the initial consonant and lasting throughout the syllable vowel."

It is thus apparent that there is a wide range of opinions concerning the phonetic nature of these "voiced" stops in isolation.

Besides these differences of opinion, there are also possible variations across contexts and across dialects that have been commented on. Forrest (1948: 223) observed that:

> "In the Wu dialects—[Wenzhou, Ningbo, Suzhou, Taizhou, and Shanghai]—the voicing of these sounds is conditional, and appears only in the interior of a word group; when uttered singly, or at the head of such a group, these sounds appear as aspirated and unvoiced"

Similarly, with respect to Shanghai dialect, Sherard (1972) reported that the "voiced" obstruents were accompanied by breathiness in isolation but become plain voiced in non-initial contexts.

The phonetic realization and degree of contextual variation of the "voiced" category is also reported to differ somewhat from region to region with different patterns for northern and southern dialects. Chao (1970) summarized the situation thus:

> "...... for simplicity in notation and terminology, we are calling these voiced stops ...... Actually, they are pure voiced only in unstressed intervocalic positions, whereas when stressed they are voiceless stops followed by a voiced aspiration. This feature is shared by most of the Wu dialects lying in Jiangsu province, whereas those in Zhejiang province have fully voiced stops in such positions".

Several sources suggest that the breathy quality of "voiced" stops in more northerly dialects may be missing from more southerly ones. For example, Zhengzhang (1985) suggested that the M.C. "voiced" obstruents are plain voiced in some dialects of southern Zhejiang. Norman (1988) also reports that in most of the northern Wu dialects, the "voiced" series has a lenis voiceless onset followed by breathy voice or murmur when they occur initially in a phrase, but in the southern part of Zhejiang this series is voiced throughout, and is without any perceptible breathy voice or murmur.

However, all the literature reviewed above is based on listening techniques alone, not on instrumental analyses of speech. During the 1980's, several experimental investigations of Wu dialects have been undertaken, which have provided more objective data. The conclusion reached by both Cao (1982) and Shi (1983) was that the "voiced" stops in the Changyinsha and Suzhou dialects respectively are the same as the voiceless unaspirated ones, except that fully voiced variants occur in unstressed intervocalic positions. When the "voiced" consonants occur in syllables pronounced in isolation or in stressed positions in running speech, they are voiceless during the closure and unaspirated. Spectrograms showed no systematic difference of voice onset time (VOT) between "voiced" and voiceless unaspirated. The only obvious distinction found between "voiced" and voiceless unaspirated cases was the tonal difference between the syllables in which they are located. On the other hand, they do differ from each other when they occur in unstressed positions in running speech. Here the "voiced" category appears as fully voiced and the voiceless category remains voiceless, but the tonal distinctions are neutralized. Moreover, these investigations did not find any evidence for the existence of voiced aspiration between consonantal release and vowel onset in the "voiced" case from examination of the spectrograms. Four pairs of wide and narrow band spectrograms of Changyinsha dialect are given in Figure 1 in which the situation described above can be observed. When /ba/ and /pa/ are pronounced in isolation, neither shows a voice bar during the closure, and there is no obvious VOT difference between them. In isolation, the tones are quite different from one another, as the narrow band spectrograms show. On the other hand, when /ba/ is in the bisyllabic word /tɕiˀ ba/ the spectrogram shows clearly that the stop is voiced, while its voiceless partner in the bisyllabic word /tɕiˀ pa/ remains voiceless. In the bisyllables, the tone contrast between /ba/ and /pa/ is neutralized, their pitch contour and level being very similar to each other, as the narrow band spectrograms show. Note that these spectrograms were made by the same speaker reported in Cao (1982), but at a later date.

On the basis of perceptual tests, Cao (1987a, 1987b) showed that listeners judged syllables to begin with a "voiced" consonant when the onset pitch was low after a voiceless unaspirated consonant. In these tests, isolated syllables were synthesized using the voicing pattern of syllables with initial voiceless unaspirated stops. Some were made with the typical pitch pattern of natural voiceless initial syllables modeled after the Shanghai dialect, while others were modified to have a pitch pattern like that of "voiced" initial syllables. The subjects, a group of Chinese linguists familiar with Wu dialects, listened to the synthesized syllables and were asked to report if they heard voiced or voiceless initial consonants. They reported that the syllables with lower pitch onset began with a voiced consonant. In other words, no actual voicing was required

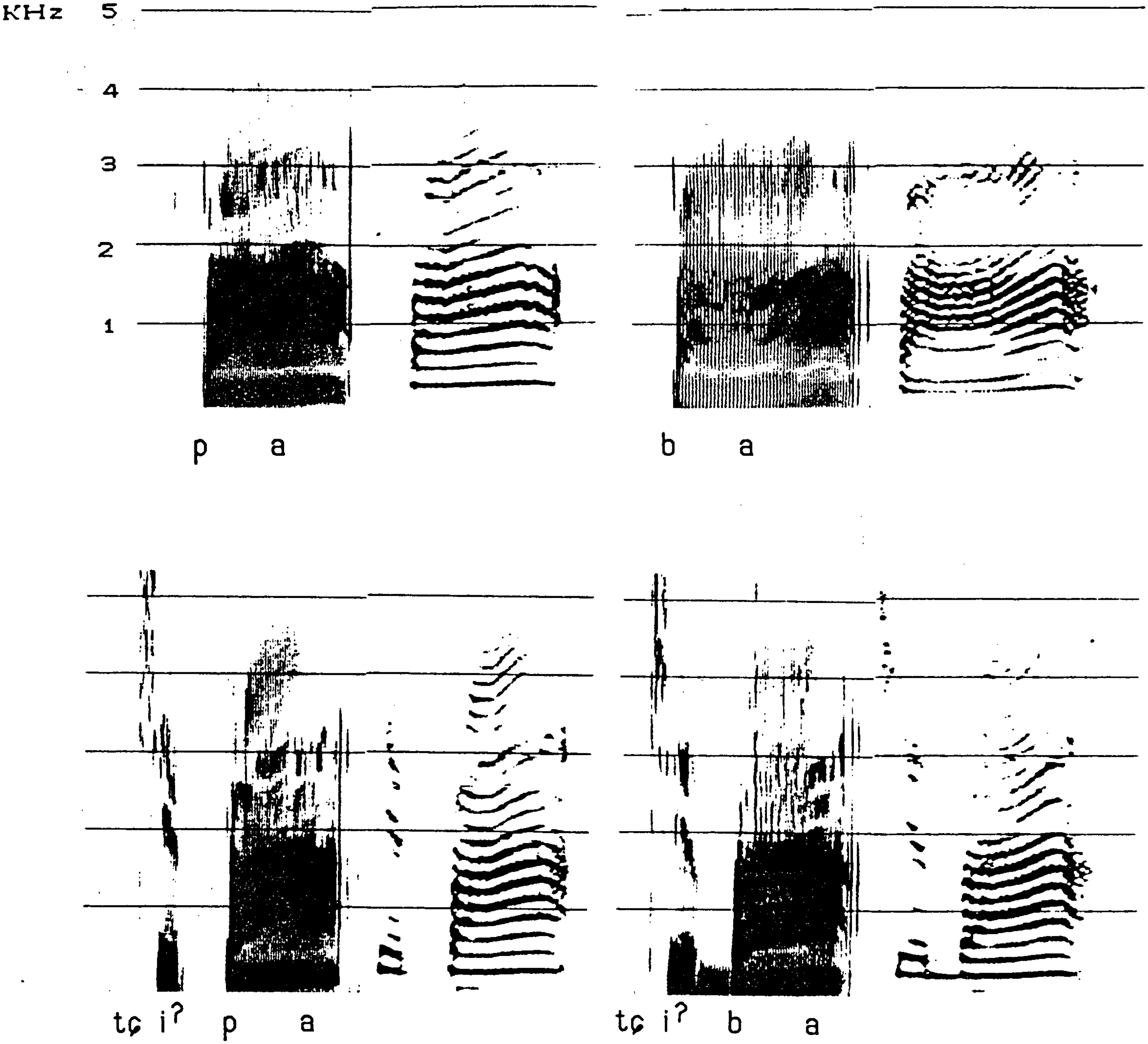


Figure 1. Wide and narrow band spectrograms of 拜/pa/ and 敗/ba/ syllables spoken by a Changyinsha speaker in isolated position (top) and in bisyllabic words (bottom).

to elicit a response that a word contained a "voiced" consonant. Some limited aerodynamic data on the Changyinsha dialect was also reported in Cao (1987a) but at that time no systematic difference between the "voiced" and voiceless unaspirated categories in isolation was observed, other than the pitch difference. However, it was not considered to be the case that breathiness had been ruled out by these results, since some differences in the airflow patterns were noted during the following vowels which might be correlated with differences in phonation type, not just differences in pitch. The airflow pattern during the vowel of the "voiced" initial syllables generally has a lower onset and takes a longer time to reach its maximum than that of the corresponding voiceless initial syllables. Although this airflow pattern difference seems to match the difference in the tonal patterns of the syllables in question, it does not seem that it is a necessary consequence of producing the pitch difference. Hence, Cao suggested that there might be some phonation difference between the syllable types associated with their vowels.

Ren's study of four Shanghai speakers (1987) confirmed little or no VOT difference between "voiced" and voiceless unaspirated stops, but he also used spectral measurements, as proposed by Bickley (1982), to investigate whether there were breathiness features. He found indications that the onset of the vowel following "voiced" stops did have some breathy qualities, but that these had disappeared by the end of the vowel. He regarded this pattern as evidence that this quality was not inherent in the vowels, preferring instead to treat it as some kind of transitional feature between consonant and vowel. Surprisingly, breathiness was reported for both initial and medial stops (where they are truly voiced). Since he also found evidence of breathiness following voiceless aspirated stops, which are followed by higher Yin tones, he argued that there must be relatively independent mechanisms for stop production and tonal distinction.

However, these instrumental studies leave several questions unanswered. For example, are there differences between northern and southern Wu dialects in the realization of the "voiced" stops, and could these differences be responsible for the differing opinions on the nature of this stop series? Secondly, apart from Ren (1987), it is generally agreed that unstressed medial instances of the "voiced" stops are actually plain voiced stops. Is any breathy feature that occurs with isolated "voiced" stops retained in this position? Can the indications of breathiness obtained from spectral measurements be confirmed by aerodynamic data? Against this background, a series of more extensive explorations of phonation type have been undertaken on several dialects within the Wu group.

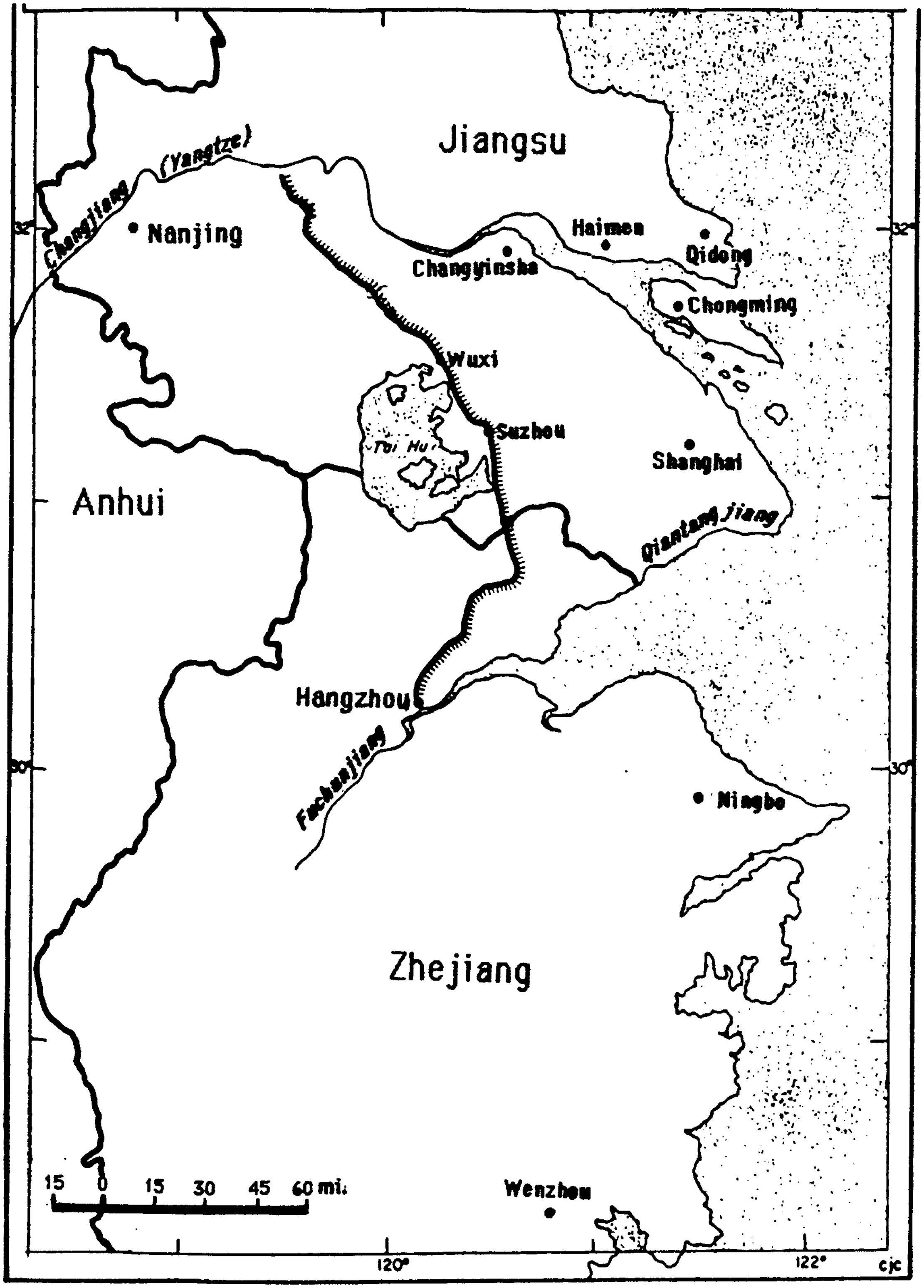


Figure 2. Map of the Wu dialects mentioned in the present investigation.

Dialects investigated

The present exploration involves four dialects of Wu, namely, Shanghai, Changyinsha, Ningbo and Wenzhou. These locations are shown on the map in Figure 2. The dialects were chosen so that two examples from each of the northern and southern groups were included. The Changyinsha dialect is very similar to Chongming, Haimen and Qidong dialects, which are spoken in a wide area of northern Wu, so the Changyinsha dialect can be taken as a good example of northern Wu Dialects. The Shanghai dialect is the most widely known of the Wu dialects and is usually taken as the typical example of them. It is also classified among the northern group, although it has some intermediate features (Norman 1988). The Wenzhou dialect is usually treated as the typical representative of southern Wu. The Ningbo dialect is spoken in northern Zhejiang province near Shanghai, but is commonly regarded as belonging to the southern Wu group, despite similarities to Shanghai. We will describe the relevant phonological differences between these dialects briefly below.

Phonologically, there are 5 underlying tones in Shanghai and in some varieties of the Ningbo dialect (Shi 1979). Changyinsha and Wenzhou, on the other hand have retained the 8 separate tones of late M. C. which arose as a result of a split in each of the original four tone categories of M.C., conditioned by the initial contrast of "voiced" and voiceless consonants.

In Shanghai, tone1 is a high falling tone (HF), tone 2 is a low rising tone (LR), tone 3 is a mid rising tone (MR), and the other two tones are short tones with checked finals, one high (H) and one low (L) (Zee & Maddieson 1980). According to the pitch contours observed in the present investigation, tone 2 can be described more exactly as a low level and rising tone (LL-R), and tone 3 as a mid falling and rising tone (MF-R). Tones 2 and 5 belong to the Yang tone category and they are always accompanied by a "voiced" consonant as the initial; tones 1, 3 and 4 belong to the Yin tone category and they always have a voiceless initial if the syllable in question begins with an obstruent.

The phonological features of the Ningbo dialect are quite close to those of the Shanghai dialect. There are 5 tones in the variety of this dialect spoken by our subjects. Tone 1 is a high falling tone (HF), tone 2 is a low level and rising tone (LL-R), tone 3 is a high level tone (HH), tone 4 is a high short tone (H) and tone 5 is a low short tone (L). The last two have checked finals. There is a similar relationship between tones and consonants to that found in Shanghai.

Changyinsha and Wenzhou are tonally more conservative in that there are four tones which occur with the "voiced" obstruents and four which do not. To illustrate this conservatism, Table 2 gives a summary of the relationship between Middle Chinese and modern Changyinsha dialect in tone categories and initial consonants.

Table 2. Correspondences of tone categories and initial consonants in Middle Chinese and Changyinsha dialect

M.C. tone category	consonant category in M.C. and Changyinsha		Changyinsha tone		
			name	contour	pitch level
平 Ping	/ 清 Qing (voiceless)	阴平	Yinping	HH	54
	\ 浊 Zhuo ("voiced")	阳平	Yangping	MR	35
上 Shang	/ 清 Qing	阴上	Yinshang	HF-R	435
	\ 浊 Zhuo	阳上	Yangshang	LR-F	131
去 Qu	/ 清 Qing	阴去	Yinqu	HR	45
	\ 浊 Zhuo	阳去	Yangqu	LL-R	214
入 Ru	/ 清 Qing	阴入	Yinru (checked)	H	5
	\ 浊 Zhuo	阳入	Yangru (checked)	L	2

The tones we observed in Wenzhou dialect, listed in the order corresponding to the historical categories in Table 2 above, are as follows. Tone 1 is a high level tone (HH), tone 2 a low rising-falling tone (LR-F), tone 3 a high rising tone (HR), tone 4 a low level and rising tone (LL-R), tone 5 a high falling tone (HF), tone 6 a low level tone (LL), tone 7 a short checked mid tone (M) and tone 8 a short checked low tone (L). These contours differ in a few minor details from those reported by Zhengzhang (1980).

The correspondences between the relevant vowels in the four dialects in question are shown in Table 3. The vowel /a/ remains the same in all four dialects but /ɔ/ and /ɛ/ in Shanghai and Ningbo have different corresponding forms in the other two dialects. However, these differences do not affect the contrast that we are discussing, so for convenience we will use the same transcription as for Shanghai in each dialect.

Table 3. Correspondence between vowels used in the four dialects

Example wordpairs (initials p-/b-)	Shanghai	Ningbo	Wenzhou	Changyinsha
拜/敗	a	a	a	a
报/刨	ɔ	ɔ	ɤ	ao
背/倍	ɛ	ɛ	ai	ei

Speech materials and subjects

Audio recordings were made of 4 speakers, 2 males and 2 females, from Shanghai and 2 speakers from each of the other three dialects, one male and one female for Ningbo, 2 males for Wenzhou, and 2 females for Changyinsha. The recorded materials included 9 basic monosyllabic word pairs, given in Table 4, as well as some bisyllabic words and short sentences. The monosyllabic pairs contain a minimal contrast between what are traditionally analyzed as initial "voiced" and voiceless unaspirated bilabial, dental and velar stop consonants in three different vowel environments.

Table 4. Test word pairs

拜	/pa/	fall on one's knees	敗	/ba/	fail
报	/pɔ/	report	刨	/bɔ/	plane
背	/pɛ/	back	倍	/bɛ/	times
带	/ta/	band	汰	/da/	wash
到	/tɔ/	arrive	道	/dɔ/	road
对	/tɛ/	pair	队	/dɛ/	team
嫁	/ka/	marry	挤[2]	/ga/	squeeze
教	/kɔ/	teach	骱	/gɔ/	joint
盖	/kɛ/	cover	依[3]	/gɛ/	lean against

The test words chosen all belong historically to the Qu tone category and show the tonal variants conditioned by the initial consonants. The particular tonal contours found in isolation in each of the dialects is shown in Table 5. The lower onset for all the tones belonging to the "voiced" category can be seen in this table.

Table 5. Variants of Qu tones in dialects examined

Dialect:	Shanghai	Ningbo	Changyinsha	Wenzhou
Qing 清 (voiceless)	MF-R 324	HH 55	HR 45	HF 51
Zhuo 浊 (voiced)	LLR 14	LLR 213	LF-R 214	LL 22

Traditionally, when people discuss the "voiced" consonants in Wu dialects, attention has principally been paid to pronunciation in isolated monosyllabic words. So the present study also focuses on isolated monosyllabic words, but some bisyllabic word pairs and sentences were also included because of the differences between isolated pronunciation and the situation in running speech noted in Wu dialects. In these examples the consonants in question are placed in unstressed intervocalic position. These bisyllabic word pairs and the sentence contexts in which the syllables were placed are listed in Table 6. The syllables of interest are enclosed in slashes.

Table 6. Bisyllabic word pairs and short sentences (as pronounced in Shanghai)

Bisyllables:	voiceless	"voiced"	
结拜 tɕiˀ/pa/	swear brotherhood	击败 tɕiˀ/ba/	defeat
勿答 fəˀ/taˀ/	not answer	勿踏 fəˀ/daˀ/	not trample
老家 lɔ/ka/	home town	老茄 lɔ/ga/	appear mature

Short sentences:

a) 一拍 就 落下来 一百张 雪白的 纸头
jiˀpʰaˀ dʑiu lɔˀɦulɛ jiˀ/paˀ/tsaŋ ɕiˀ/baˀ/gəˀ tsɹdɯˀ
"one hundred sheets of white paper will fall down when you tap them"

b) 杨韬 用 羊刀 削 洋桃
jaŋtʰɔ joŋ jaŋ/tɔ/ ɕiaˀ jaŋ/dɔ/
"Yangto peels the peach using a knife"

c) 西德 的 希特勒 死脱了
ɕi/təˀ/ gəˀ ɕi/dəˀ/ləˀ ɕitʰəˀləˀ
"Hitler of Germany died"

In addition to the audio recordings of the items in Tables 4 and 6, a more limited amount of aerodynamic data was recorded from the same speakers. Oral air flow and intraoral pressure was recorded using a flow mask around the face and a pressure transducer inserted between the lips in the way described by Dart (1987). Only the words in Table 4 beginning with bilabial consonants and the items in Table 6 containing bilabial stops were used in this part of the experiment.

Analysis

Displays of the power spectra of the vowels were made from the audio recordings using an FFT analysis. These spectra were taken from a point in the vowel as close to the release of the consonant as practicable, generally some 30 milliseconds after the burst. The differences between the amplitude of the fundamental or first harmonic (H1) and (a) the second harmonic (H2) or (b) the strongest harmonic in the first formant area (F1) were measured. These differences will be referred to as H2-H1 and F1-H1 respectively. From the aerodynamic records the peak air flow at the consonantal release and the peak air pressure during consonantal closure in the bilabial stops were calculated. The ratio of the air flow to the air pressure in each token was then calculated. This measurement will be referred to as AF/AP. Both spectral and aerodynamic measurements have served as tests of phonation type in other research (Bickley, 1982; Maddieson & Ladefoged, 1985; Kirk, Ladefoged & Ladefoged, 1984; Dart 1987; Ren 1987). When the vocal cords are vibrating in a tenser position there is less energy in the fundamental relative to higher components compared to when the vocal cords are laxer, as they would be in breathier phonation (Fant 1983). If the spectral measurements show a difference right at the onset of the vowel between pairs which differ in the initial consonant class, then we may assume that the vocal cords are actually in somewhat different positions during the consonant closure. The AF/AP measure reflects the rate of airflow out of the mouth for a given intraoral pressure reached during the consonant. By dividing the flow by the pressure, we normalize the flow measurements for variations which may be due to differences in lung pressure. For any given pressure, greater or lesser air flow as the vowel begins can be assumed to be due to differences in the tension of the vocal cords, which again must be already present during the consonantal closure.

Since the discussion concerning "voiced" consonants in Wu has been concerned with understanding what differences exist in the consonants, the main focus of the present study has been on examining phonetic properties during the consonants or as close to their release as possible. However, we do not wish to overlook the possibility that the main differences could occur during the vowels which follow. If the contrastive features lay principally in the vocalic portion of the syllable, we would expect the spectral measures reflecting phonation type to show the strongest difference between the categories in the middle of the vowel. To examine this question, measurements of the spectral differences were made at the onset, middle and offset of the vowel in a subset of the data.

Results

(a) words in isolation

The spectral measurements H2-H1 and F1-H1 and the aerodynamic ratio AF/AP were examined using analysis of variance techniques to search for main effects. Post hoc tests of the differences between the measurements on the "voiced" and voiceless unaspirated cases were then performed to evaluate the significance of the effects. The results from spectral measurements made at the onset of the vowel are summarized in Tables 7 and 8. These tables show the means of the H2-H1 and F1-H1 measures respectively. Values after voiceless unaspirated and "voiced" consonants are given separately, as well as the difference between the means. In the final column the probability that this mean difference between the two consonant classes is due to chance is given. Probability scores were calculated from paired data t-tests. Values less than .01 were considered significant.

Table 7. H2-H1 at vowel onset in the four dialects (in dB)

dialect	no. of pairs	mean H2-H1 in voiceless	mean H2-H1 in voiced	mean diff.	prob.
Shanghai	36	6.881	3.806	3.076	0.0001
Ningbo	18	9.056	4.544	4.511	0.0004
Changyinsha	18	7.439	1.728	5.711	0.0012
Wenzhou	18	8.872	5.256	3.617	0.0001

Table 8. F1-H1 at vowel onset in the four dialects (in dB)

dialect	no. of pairs	mean F1-H1 in voiceless	mean F1-H1 in voiced	mean diff.	prob.
Shanghai	36	13.631	10.753	2.879	0.0033
Ningbo	18	11.828	11.089	0.739	0.4462
Changyinsha	18	13.939	11.856	2.083	0.1901
Wenzhou	18	12.900	11.750	1.150	0.4578

The H2-H1 measure shows a highly significant difference for each dialect. There is a smaller difference between the amplitude of the first and second harmonics in the "voiced" tokens than in the voiceless unaspirated ones. In general, the F1-H1 measurements did not prove to significantly distinguish "voiced" and voiceless categories except in the Shanghai dialect, but the trend in the data is in expected direction. That is, the "voiced" cases show less difference in the amplitudes of F1 and H1 than the voiceless unaspirated ones. So these results indicate that the first harmonic at the onset in "voiced" cases does have relatively more energy in comparison with their voiceless partners.

As a partial check that this spectral difference is associated with the consonant contrast rather than being a property of the vowel as a whole, the H2-H1 measurement was made at three different points in the Shanghai pairs with the vowel /a/. The results of the spectral measurements at the onset, middle and offset are given in Table 9. These indicate that the effect is significant only at the onset of the vowel. In the middle of the vowel the effect is weak and at the vowel offset, it has even reversed its direction. We conclude that the phonation difference is not spread over the entire vowel, but is concentrated at the onset.

Table 9. Spectral measurements of H2-H1 at different points of the vowel in Shanghai dialect.

measurement point	no. pairs	mean diff.	prob.
onset	12	2.633	0.0099
middle	12	1.850	0.1718
offset	12	-2.642	0.0850

The results of the analyses of aerodynamic measurements are given in Table 10. When the dialects are examined separately the mean difference of the ratio of AF/AP between "voiced" and voiceless cases is significant only in the Wenzhou dialect. It is marginally significant in each of the other three dialects. However, since only the words beginning with bilabials were used for aerodynamic study, these results reflect only a small number of measurements. Failure to reach significance might well be due to the small number of tokens involved, rather than to absence of a real difference. For this reason, the AF/AP measures for the four dialects were pooled and a joint analysis of all 30 paired measurements was conducted. In this analysis the "voiced"/voiceless contrast was shown to be highly significant across the four dialects taken together. The pooled results are given in the last line of Table 10.

Table 10. Airflow/pressure ratio (AF/AP) in the four dialects

Dialect	No. of pairs	mean AF/AP in voiceless	mean AF/AP in voice	mean diff.	prob.
Shanghai	12	0.096	0.121	-0.025	0.0318
Ningbo	6	0.033	0.067	-0.034	0.0240
Changyinsha	6	0.070	0.094	-0.025	0.0286
Wenzhou	6	0.272	0.318	-0.047	0.0055
Pooled data	30	0.113	0.144	-0.031	0.0001

These results confirm that the states of the glottis are different between "voiced" and voiceless consonants. Together with the results of the spectral measurements described above, this establishes that there is a phonation difference between "voiced" and voiceless unaspirated consonants, not just a difference in the pitch of the following vowel, in all of these dialects. In other words, these two series of consonants are not produced alike in initial position in isolated words, as had been suggested by Cao (1982).

Results in bisyllabic words and sentences

The spectral measurements described above were also made on the selected syllables in the medial positions given in Table 6 above. Two speakers each of the Shanghai and Changyinsha dialects produced the six bisyllabic words and also provided some sentences, namely sentences (a) and (b) for the Changyinsha dialect and all three sentences for the Shanghai dialect. In addition, the AF/AP ratio was obtained for the bilabial-initial syllables in the bisyllabic words and the sentences from both the Shanghai and Changyinsha speakers. The results are given in Table 11. The spectral measurements show at best a weakly significant difference at the vowel onset between "voiced" and voiceless unaspirated consonants in this context. The Shanghai H2-H1 measure is closest to a significant level. No significant difference in the aerodynamic ratio was found for either dialect. Overall, there is no reason to think that there is a difference in the voice quality at the beginning of the vowels, as was observed in the isolated monosyllables. Our spectrograms confirm that the "voiced" consonants are indeed phonetically voiced in these contexts, but this voicing contrast is not strongly reflected in the measurements we have made.

Table 11. Spectral and aerodynamic measurements in medial positions in Shanghai and Changyinsha dialects.

dialect	no. of pairs	measurement	mean diff.	prob.
Shanghai	12	H2-H1	1.442	0.0462
	12	F1-H1	2.283	0.0829
	4	AF/AP	-0.096	0.3646
Changyinsha	10	H2-H1	1.970	0.1661
	10	F1-H1	1.240	0.5175
	4	AF/AP	-0.015	0.3599

Discussion

The analysis given above has clearly illustrated that the use of a phonation difference for the contrast of voiced and voiceless unaspirated consonants in isolation is shared by the four dialects of Wu we have examined. The mean energy difference in both the H2-H1 and the F1-H1 measurements is less after voiced consonants than after voiceless unaspirated consonants. In other words, the first harmonic at the onset of the vowel in the "voiced" cases generally has relatively higher energy than that in the voiceless cases. Furthermore, the AF/AP ratio is greater in the "voiced" cases than that in the voiceless cases. This indicates that the glottis is relatively more open at vowel onset in the "voiced" cases, and as a result, there is relatively greater airflow through the glottis in the syllables with the "voiced" initial. This agrees with Ren's unpublished data on glottal width in the three stop series in Shanghai obtained using glottal transillumination (Ren 1988). In Ren's data the glottal aperture is shown to be wider at the time of consonant release in the "voiced" stops than it is in the unaspirated voiceless ones.

Recall that Zhengzhang (1985) and Norman (1988) indicate that in Wu dialects of southern Zhejiang the "voiced" category of consonants is plain voiced. This is contrary to our finding with respect to the Wenzhou dialect, which is explicitly included by Norman in this group. We examined our data to see if a distinction between northern and southern dialects was to be found. One comparison used the Changyinsha dialect as the representative of northern Wu and Wenzhou dialect for southern Wu, since these are the most distinctive representatives of their respective groups. Analysis of variance was employed to look for a main effect of dialect in the spectral measurements. For the H2-H1 measure a significant effect of dialect was observed ($F(1, 68) = 7.747$, $p =$

.007). As can be seen from Table 7 the magnitude of the voiced/voiceless difference is considerably greater in Changyinsha than in Wenzhou. A highly significant difference was also found between these dialects in the AF/AP measure (F (1,20) = 116.857, *p* = .0001). Values for both consonant types are much higher in Wenzhou than in Changyinsha, as can be seen in Table 10. The reason for this result is unclear; it may only reflect a difference of individual loudness in speaking, or be related to the fact that the Wenzhou speakers happened to be male, whereas the Changyinsha speakers were both females. For this reason, the AF/AP measure cannot be used to establish a difference between the dialects. However, our spectral results do seem to reflect some difference between typical northern and southern dialects in the realization of the contrast we are examining. We tested whether this difference might be related to an earlier onset of voicing in the southern representative by measuring VOT in the bilabial word pairs given in Table 4. VOT was essentially zero for both classes of consonants in both dialects and no significant difference (F (1,44) = 0.0083, p = 0.9278) was found between the dialects. Hence the spectral differences are assumed to reflect different degrees of breathiness at vowel onset, not a voicing difference in the consonant closure.

In a second analysis we combined all of the Changyinsha and Shanghai speaker's data, as the representatives of northern Wu, and compared that with the combined data from Wenzhou and Ningbo, the representatives of southern Wu. The H2-H1 measurement is still significantly different across the two groups of dialects (F (1, 175) = 6,657, *p.* = 0.0107). This is because the mean values for **both** consonant classes are lower for the two northern dialects than the two southern ones. However, the magnitude of the mean **difference** between the "voiced" and voiceless classes is quite similar in the two groups of dialects. Consequently, this result does not enable us to comment on the possibility of a general difference between northern and southern dialects. In fact, it might suggest that the Shanghai and Ningbo dialects are not truly representative of the northern and southern dialect groups to which they are usually assigned, since adding them to the totals confuses rather than sharpens the picture.

Another disputed question concerns whether the phonation type difference described above is an inherent property of the consonants themselves, or of the following vowels, or "pervades the entire syllable." We have shown above that the entire syllable is not affected phonetically. But taking Chao's view a little further, Cao (1987a) suggested that the phonation type difference is associated with the following vowels. On the other hand, Ren (1987) supposed that it "is not an inherent phonation type in the vowels themselves, but rather the result of interaction between segments." Most phonologists would continue to treat it as a property of the consonants alone.

The individual segment in speech cannot stand in complete separation from the context, and there generally exist some transitional features between segments. These transitional features are generally accounted for by considering the properties of the abutting segments involved. No special property inherent in the transition itself needs to be invoked. So, for example, in the spectrograms in Figure 1 an upward movement of the second formant of the /a/ vowel following release of the bilabial closure can be observed. The fact that this transition originates from a low onset can be predicted from the fact that the consonant closure is bilabial, whereas the extent of the upward movement can be predicted from the fact that the vowel concerned is /a/. We would suggest that the breathy property of the "voiced" stops in Wu dialects is somewhat similar to the transitions dependent on place of articulation, in that it produces a transitional portion in the following vowel overlaid on the spectral characteristics of that vowel. Phonologically, it is a property of the consonant alone although the phonetic manifestation is largely detectable at the onset of the following vowel.

As we noted, air pressure was measured during the consonantal closure and air flow was taken immediately after release. The AF/AP ratio reflects different states of the glottis during the consonantal portion of the syllables with "voiced" and voiceless unaspirated initials. On the other hand, the difference in energy distribution of harmonics between "voiced" and voiceless cases indicates that the breathiness does affect the following vowel portion.

We now return to one of the major issues of interest in this study, the possible relevance of these data to the interpretation of phonetic basis of the historical split into Yin and Yang tones. Our study has confirmed the observations of others to the effect that the "voiced" consonants in Wu dialects are only truly voiced in unstressed intervocalic positions. Interestingly, it is precisely in these positions that the tonal contrasts between Yin and Yang tones are entirely or largely neutralized in many of the Wu dialects. This is also the position in which a contrast in terms of the factor we have labeled breathiness is not found. The situation is summarized in Table 12. It provides a possible indication that the breathiness contrast may be a more powerful influence on pitch than plain voicing is. Since the tone split is pervasive across Chinese dialects, such a strong influence might be more plausibly assumed to be its cause.

Table 12. Relations among the breathiness, tonal and voicing differences

stress type	voicing	tonal	breathiness
stressed	-	+	+
unstressed	+	-	-

Conclusion

In summary, the present investigation leads to the following conclusions. The contrast of "voiced" consonants and voiceless consonants in Wu dialects of Chinese is more related to a difference in glottal adjustment which causes a more breathy phonation to follow the "voiced" consonants than to the actual presence or absence of voicing. The use of a special glottal adjustment for "voiced" consonants is typical of Wu dialects from different parts of the region in which they are spoken, not just of the northern subgroup. Phonologically this property is inherent in the consonants but its clearest phonetic manifestation is in the onset to the vowel. It seems likely that this particular type of phonatory setting is a more powerful cause of pitch lowering than normal voicing, and that there is strong reason to accept the view that Middle Chinese "voiced" stops were phonetically characterized by some breathiness.

Acknowledgments

The authors gratefully acknowledge the assistance of all the subjects who participated in this study. We are also grateful for help and advice provided by many members of the UCLA Phonetics Laboratory group, most especially Peter Ladefoged. This work was funded in part by the Ministry of Education of the Peoples' Republic of China, and the National Institutes of Health of the United States.

Footnotes

1. King et al (1987) note an absence of voicing during stop closures in the "Shanghai" speaker they studied. However, it is clear from the number and shape of the tones they report that their subject was actually speaking Mandarin, not Shanghai dialect.

2, 3. The two characters do not represent the historically "voiced" initial words cited, since these words are only used in the spoken language but have died out from the current written system of Chinese. The characters given here just represent the synonyms of the intended words.

References

Bickley, Corine. 1982. Acoustic analysis and perception of breathy vowels. *Working*

Papers, Speech Communication Group, MIT 1:71-81.

Cao, Jianfen. 1982. Changyinshahua guquanzhuomu de fayin tedian (The characteristics of the ancient initial voiced consonants in Changyinsha dialect). *Zhongguo yuwen* 167: 273-278.

———— 1987a. Lun qingzhuo yu daiyin he butaiyin de kuanxi (The relationship between phonological category and phonetic nature of "voiced" and voiceless consonants). *Zhongguo yuwen* 197: 101-109.

———— 1987b. The ancient initial "voiced" consonants in modern Wu dialects. *Proceedings of the 11th International Congress of Phonetic Sciences, Volume* 4: 169-172. Academy of Sciences of the Estonian S.S.R., Tallinn.

Chao, Yuan Ren. 1970. The Changzhou Dialect. *Journal of the American Oriental Society* 90: 45-56. Reprint in *Aspects of Chinese Sociolinguistics,* Stanford University Press, Stanford (1976) pp.48-71.

———— 1967. Contrastive aspects of the Wu dialects. *Language* 43: 92-101.

———— 1936. Types of plosives in Chinese. *Proceedings of the 2nd International Congress of Phonetic Sciences.* Cambridge University Press, Cambridge: 106-110.

———— 1930. Tingxie daoyingwen (A Dictation to an English backwards recording). *ShiyuShuo Jikan.* 2(2).

———— 1928. *Studies in the modern Wu dialects.* Qinghua University, Beijing.

Dart, Sarah N. 1987. An aerodynamic study of Korean stops: measurement and modeling. *Journal of the Acoustical Society of America* 81: 138-147.

Fant, Gunnar. 1983. Preliminaries to analysis of the human voice source. *Quarterly Progress Report , Speech Transmission Laboratory, Royal Institute of Technology, Stockholm* 1982/4: 1-27.

Forrest, R.A.D. 1948. *The Chinese Language.* Faber & Faber, London.

Hombert, Jean-Marie, John J. Ohala & William Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55: 37-58.

Karlgren, Bernhard. 1915-1926. Études sur la phonologie Chinoise. *Archives d'études orientales,* Vol.15 (in 4 parts). Leiden: E.J.Brill; Uppsala: K.W. Appelberg.

———— 1940. Grammata serica, script and phonetics of Chinese and Sino-Japanese. *Bulletin of the Museum of Far Eastern Antiquities* 12: 1-471.

King, Ling, Hary Ramming, Lieselotte Schiefer & Hans G. Tillmann. 1987. Initial F_0 contours in Shanghai CV syllables — an interactive function of tone, vowel height and place and manner of stop articulation. *Proceedings of the 11th International Congress of Phonetic Sciences., Volume* 1: 154-157. Academy of Sciences of the Estonian S.S.R., Tallinn.

Kirk, Paul L., Peter Ladefoged, & Jenny Ladefoged. 1984. Using a spectrograph for measures of phonation types in a natural language. *UCLA Working Papers in*

Phonetics 59: 102-113.

Li Fanggui. 1971. Shangguyin yanjiu (A study in the speech sounds of Old Chinese). *Qinghua Journal of Chinese Studies, New Series* 9:1-61.

Li Rong. 1986. Wenlinguhua "ɦædæ" daoguolai ting haishi "ɦædæ" (The same auditory impression: listening to the recording of the word "ɦædæ" forwards and backwards in Wenling dialect). *Fangyan* 1986: 106.

Luo Changpei. 1956. *Hanyu Yinyunxue Daolun* (An introduction to Chinese Phonology). Zhonghua shuju, Beijing.

Luo Changpei & Wang Jun. 1981. *Putong Yuyinxue Gangyao* (Elements of General Phonetics), new ed. Commercial Press, Beijing.

Maddieson, Ian & Peter Ladefoged. 1985. "Tense" and "lax" in four minority languages of China. *Journal of Phonetics* 13: 433-454.

Norman, Jerry. 1988. *Chinese*. Cambridge University Press, Cambridge.

Ramsey, S. Robert. 1987. *The Languages of China.* Princeton University Press, Princeton.

Ren Nianqi. 1987. An acoustic study of Shanghai stops. Unpublished manuscript. University of Connecticut, Storrs.

——— 1988. A fiberoptic and transillumination study of Shanghai stops. Paper presented at International Conference on Wu Dialects, Hongkong, December 1988.

Sherard, Michael. 1972. *Shanghai phonology*. Ph.D. dissertation, Cornell University, Ithaca.

Shi Feng. 1983. Suzhouhua zhuoseyin de shengxue texin (The acoustic characteristics of the voiced stops in Suzhou Dialect). *Yuyian Yianjiu* 1983: 49-83.

Shi Wentao. 1979. Ningbo fangyan benzikao (Etymological notes on the Ningbo dialect) *Fangyan* 1979: 161-170.

Wang Li. 1956. *Hanyu Yinyunxue* (Chinese Phonology). Zhonghua Shuju, Beijing.

Yan Yiming. 1988. Wuyuqu renkou de zaitongji (Re-statistics on the population of Wu dialects). *Zhongguo Yuwen* 205: 287-288.

Yuan Jiahua. 1960. *Hanyu Fangyan Gaiyao* (An Outline of Chinese Dialects). Wenzi Gaige Chubanshe, Beijing.

Zee, Eric and Maddieson, Ian. 1980. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *Glossa* 14: 45-88.

Zhengzhang Shangfang. 1980. Wenzhou fanyan er-weici de yuyin bianhua (The suffix *er* and sandhi in the Wenzhou dialect) *Fangyan* 1980:245-262.

——— 1985. Pucheng fanyan de nanbei qufen (North and south divisions of the Pucheng dialect). *Fangyan* 1985: 39-45.

——— 1988. Guanyu Wuyu de renkou (The popuation of Wu dialects). *Zhongguo Yuwen* 205: 289-291.

A note on "Information conveyed by vowels"

Peter Ladefoged
Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, CA 90024-1543
[To be published in *The Journal of the Acoustical Society of America*]

In an early experiment using synthetic speech it was shown that raising or lowering the formants in an introductory sentence affected the identification of the vowel in a following test word. This experiment has now been replicated using natural speech produced by a phonetician using two different overall settings of the vocal tract. 75% of the subjects heard a given stimulus as "bit" when preceded by one sentence, and as "bet" when preceded by another.

A number of papers in a focus session at the 113th meeting of the Acoustical Society of America (May 1987, Indianapolis) cited a 30 year old paper (Ladefoged and Broadbent 1957) that showed the potential relative nature of vowel quality. This paper reported research using an early speech synthesizer (Lawrence 1955), which was used to produce the sentence "Please say what this word is" and four test stimuli somewhat similar to the words "bit, bet, bat, but." It was shown that a given test word might be heard as "bit" by 97% of the subjects when preceded by a sentence in which the overall pattern of F1 was comparatively high, and as "bet" by 95% of the subjects when the introductory sentence had a low F1 pattern.

In the discussion following the presentation of the papers in the focus session some doubts were expressed about this experiment. It was not completely clear if this effect was largely due to the poor intelligibility of the synthetic speech. Subjects might have been straining to understand what was being said and so have been forced into unusual modes of speech perception. The situation was not helped by the vague remarks of the present author, nor by the fact that the only published account of an attempted replication was an abstract in *The Journal of the Acoustical Society of America.* This abstract was entitled 'Information conveyed by vowels: a negative finding' (Dechovitz 1977a), and contained the comment "In the present study no evidence for such a perceptual mechanism [that given in the experiment cited above] was obtained using natural speech." Confusion concerning the validity of the original experiment was not fully cleared up by the fact that this orally presented paper subsequently appeared in an unpublished lab report (Dechovitz 1977b) with a new title: 'Information conveyed by vowels: a confirmation.' In order to throw further light on this problem it was decided to replicate the experiment in an abbreviated form using natural speech.

The same introductory sentence was spoken by the author in two different ways. In one he used a very spread lip position, which substantially raised the higher formants, together with a generally raised tongue position, which lowered F1. In the other he made the recording with a more rounded lip position (which lowered the higher formants) and an overall lower tongue position (which raised F1), while also speaking at a slightly slower rate. This recording was then played back at 120% of its original speed, raising the higher formants again, so that they were comparable with those in the first utterance, and also raising F1, so that it was appreciably higher than in the first utterance. Only two test stimuli were recorded, one being the speaker's normal "but", and the other being a word with a vowel intermediate in quality between his normal "bit" and "bet". The recordings were made and edited using an 8 bit digital speech processing system with a sample rate of 22 kHz. The overall signal to noise ratio was about 38 dB (limited by the tape reproducing system rather than the digitizer) and the frequency response was flat (within 2 dB) up to 8,000 Hz (limited by the sample rate and the subsequent playback speed up).

The subjects were students in an introductory linguistics course. Only those who were native speakers of English, and who had had all their schooling in California were scored. Subjects were given duplicate written and oral instructions in which they were told that they would hear a voice saying "Please say what this word is" followed by a word that might be either "bit" or "bet" or "bat" or "but". Their task was to check the appropriate word on their answer sheets; if in doubt they were to guess. The experiment was conducted in an ordinary university classroom under good but far from perfect acoustic conditions.

The experiment consisted of three trials each separated by a break of about 15 seconds in which the experimenter urged subjects to do their best. The real purpose of this break was to destroy the auditory memory of the previous stimulus. The first trial consisted of the first sentence followed by the "but" stimulus. This trial was not scored; it was used simply to get the subjects accustomed to the experimental procedure. In the second trial the second sentence was followed by the ambiguous "bit-bet" stimulus; and in the third trial the first sentence was followed by the same ambiguous stimulus.

The results confirmed the validity of the original experiment. Of the 59 subjects, 44 (= 75%) heard the same stimulus as "bit" on the second (first scored) trial, and as "bet" on the third (the other scored) trial. When interpreting these results, we must recall that, as Nearey made evident in his paper in the focus session, there are many factors in the perception of vowels, and that in some circumstances some of them have greater weight than others. In this experiment subjects were influenced by the relative nature of vowel quality; in others, as the other papers show, the quality of the vowel depends more strongly on aspects of the stimulus itself. The moral of all this work seems to be that we should avoid statements of the form 'vowels are differentiated acoustistically in such and such way,' and say instead "vowels can be (or may be) differentiated acoustically ...' Humans have scores of ways of perceiving speech; and any one experiment can reveal only a few of them.

References

Dechovitz, David (1977a). "Information conveyed by vowels: a negative finding." *J. Acoust. Soc. Amer.* 61.1, S39.

Dechovitz, David (1977b). "Information conveyed by vowels: a confirmation." *Haskins Laboratories Status Reports.* SR-51—52, 213-219.

Ladefoged, Peter and Broadbent, Donald (1957). "Information conveyed by vowels." *J. Acoust. Soc. Amer.* 29, 98-104.

Lawrence, Walter (1955). "The synthesis of speech from signals which have a low information rate," in *Communication Theory*, W. Jackson, ed. (London, Butterworth Scientific Publications).

The symbols for clicks

OSWIN KÖHLER
Institut für Afrikanistik der Universität zu Köln, Meister-Ekkehart-Strasse 7, 5 Köln 41, F.D.R.

PETER LADEFOGED
Linguistics Department, UCLA, Los Angeles, CA 90024-1543, U.S.A.

JAN SNYMAN
African Languages, UNISA, P.O. Box 392, Pretoria, 0001, South Africa

ANTHONY TRAILL
Linguistics Department, University of Witwatersrand, 1 Jan Smuts Avenue, Johannesburg, South Africa

RAINER VOSSEN
Universität Bayreuth, Postfach 30 08, D-8580 Bayreuth, F.D.R.

Click sounds occur in languages that are in regular use (that is, excluding ceremonial languages and other special speech forms) only in languages of the Khoisan family and those Bantu languages which have been influenced by them. It therefore seems appropriate that the IPA symbols for clicks should be determined largely by the usage of linguists working on these languages, and with due consideration of the orthographies that are in use locally. At present this is not the case. Among Khoisan linguists, the only IPA click symbol that is in use is that for the bilabial click [ʘ]. The three other symbols on the current IPA chart, [ʇ] for the dental click, [ʗ] for the alveolar, and [ʖ] for the alveolar lateral click, together with an additional symbol for the palatal click [ǂ], (which was proposed by Beach (1938), but never adopted by the IPA), are used virtually only by non-specialists in these languages. The only major publications on Khoisan languages in which they can be found are Beach (1938), and one paper by Doke (1925), who also used these symbols in his work on Zulu (Doke 1926).

We propose that the IPA symbols should be replaced by those shown in Table 1, thus bringing the IPA into line with the major users of these symbols.

TABLE 1: Proposed changes in symbols for clicks.

Current IPA	ʘ	ʇ	ʗ	ʖ	
Proposed	ʘ	|	!	‖	ǂ

These symbols were devised by Lepsius (1855), and have been in use in Southern Africa since that time. They have been used in a very large number of technical works on the languages of the area, including not only well known early books such as Krönlein (1889), Meinhof (1909) and Bleek and Lloyd (1911), but also in our own contemporary work as cited in Table 2. In addition they are used in books and literacy materials in one of the main languages of the area, Nama/Damara. Writers on Bantu languages such as Zulu and Xhosa occasionally use the IPA symbols; they were extensively used by Doke and others half a century ago, but nowadays works on Bantu languages including those on phonetic topics, tend to use the orthographic symbols, or the Lepsius symbols.

TABLE 2 : Symbols for clicks in different Khoisan Languages. Only the alveolar click is shown; the others would be similar. K = Köhler (1981); L & T = Ladefoged & Traill (1984); S = Snyman (1975); T = Traill (1985); V = Vossen (1986a,b).

Accompaniment	K	L & T	S	T	V
Voiced	!̰a	g!a	g!a	!ga	!̰a
Voiceless unaspirated	!a	k!a	!a	!a	!a
Aspirated	!ha	k!ha	!ha		!ha
Voiced nasalized	!̄na	ŋ!a	ŋ!a	!na	!̄na
Voiceless nasalized		ŋ̥!a		!n̥a	!̄a
Prenasalized with voiceless release	!̄a				
Preglottalized & nasalized		ʔ!ŋa		ʔ!na	
Voiced uvular stop & prenasalized		ɴ!ɢa		ɴ!ɢa	
Voiceless uvular stop	!qa	q!a		!qa	!qa
Aspirated uvular stop		q!ha		!qha	
Voiceless velar fricative or affricate	!xa	k!xa	!xa	!xa	!xa
Ejective uvular stop		k!q’a		!qʔa	
Ejective velar (or uvular) affricate (or stop)	!x’a	!q’a	!xʔa	!kxʔa	!xˀa
Delayed aspiration		!ha	!¯ha	!ha	!ˀha
Glottal stop	!’a	!ʔa	!ʔa	!ʔa	!ˀa
Prevoiced plus aspirated		g!ha	ɕ|ha	g!ha	
Prevoiced plus velar fricative		g!xa	g!ɣa	g!xa	!̰xa
Prevoiced plus velar (or uvular) ejective		g!q’a	ɕ!xʔa	g!kxʔa	
Prenasalized delayed aspiration			ŋ!ha		
Prenasalized voiced fricative					!̄xa

In both Khoisan and Bantu languages these clicks can occur with several different accompaniments such as nasality and different kinds of voicing. Unfortunately there is less uniformity among linguists active in the area concerning the symbolization of these accompaniments. Some of the different systems for Khoisan languages are illustrated in Table 2. We hope that the linguists involved will be able to get together shortly and come to some agreement. But we also hope that the IPA will take action and change the basic symbols without waiting for this to happen.

References

BEACH, D.M. (1938). *The Phonetics of the Hottentot Language.* Cambridge: Heffer.

BLEEK, W.H.I. and LLOYD, L. (1911). *Specimens of Bushman Folklore.* London: Allen.

DOKE, C.M. (1925). "The phonetics of ʇhɑ: Bushman." *Bantu Studies* **2.3**, 129-165.

DOKE, C.M. (1926). "The phonetics of the Zulu Language." *Bantu Studies* Special number. [reprinted, 1969, Nendeln/Liechtenstein: Kraus Reprint.]

KÖHLER, O. (1981). Les langues khoisan. In *Les langues dans le monde ancien et moderne. (sous la direction de Jean Perrot)* Paris: Centre National de la Recherche Scientifique. 3, 455-615.

KRÖNLEIN, J. G. (1889). *Wortschatz der Khoi-Khoin.* Berlin.

LEPSIUS, C.R. (1855). *Standard Alphabet for Reducing Unwritten Languages and Foreign Graphics Systems to a European Orthography in European Letters.* London: Williams & Norgate. [2nd Ed, 1863, reprinted with an Introduction by J. Alan Kemp. Amsterdam: John Benjamins B.V. (1981).]

LADEFOGED, P. and TRAILL, A. (1984). "Linguistic phonetic descriptions of clicks." *Language* **60:1**,1-20.

MEINHOF, C. (1909). *Lehrbuch der Nama-Sprache.* Berlin.

SNYMAN, J.W. (1975). *Zu|hõasi Fonologie & Woordeboek.* Rotterdam: A.A. Balkema.

TRAILL, A. (1985). *Phonetic and Phonological Studies of !Xóõ Bushman.* (Quellen zur Khoisan-Forschung, **3**.) Hamburg: Helmut Buske.

VOSSEN, R. (1986a). "Zur Phonologie der ||Ani-Sprache." In: R. Vossen and K. Keuthmann (eds.), *Contemporary Studies on Khoisan,* **2**, 321-345. (Quellen zur Khoisan-Forschung, **5**.) Hamburg: Helmut Buske.

VOSSEN, R. (1986b). "Some observations on nominal gender in Naro." in F. Rottland (ed.), Festchrift sum 60. Geburtstag von Carl F. Hoffmann, 373 - 390. (Bayreuther Beiträge zur Sprachwissenschaft, 7.) Hamburg: Helmut Buske.

Vocal — a Macintosh computer program

Peter Ladefoged
Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, CA 90024-1543

The program **Vocal** was originally devised in order to help us assess the acoustic results of different articulatory gestures. We want to know the acoustic structure of the sounds that the brain tells the vocal organs to make. Of course, we do not know how the brain works, but we can make plausible guesses concerning the way in which the vocal organs are controlled. Ultimately we would like a system that starts from discrete units that specify articulatory movements, and joins them into continuous speech. The present Vocal system does not contain such a module, but it is built in such a way that it could be easily added. Modeling articulatory-acoustic relations remains the prime function of the system; but there are now additional components that enable us to estimate, very roughly, possible articulatory causes of acoustic data used as input, and also to do rudimentary synthesis of vowels and phrases directly from formant data.

There were a number of design considerations; as this is an event-driven program, they are best discussed in terms of the menus involved.

1. The vocal tract should be specified in a way that is appropriate for any individual. See **File** menu.
2. It should be possible to create and edit sequences of vocal tract states representative of articulatory movements. See **Edit** menu.
3. The articulators should be controlled in a way that might reflect the way speech is organized. See **Articulators** and **Options** menus.
4. It should be possible to visualize movements. See **Display** menu.
5. The acoustic data should be calculated appropriately; and it should be possible to estimate possible articulatory shapes from acoustic data. See **Acoustic** menu.
6. Each of the above considerations should be realized in code modules with minimal inter-dependence with a user (and programmer) friendly interface. See **Help** menu. [But note that, alas, the code is not as modular as it should be.]

The program is written in Lightspeed Pascal, fairly well commented in the parts that have been used for teaching purposes, but otherwise somewhat hacked. The program itself is in the public domain; and the source code can be obtained free. It runs on a MacPlus (or better), with no additional peripherals needed, thus providing a means of studying articulatory-acoustic relations and also offering some rudimentary synthesis for the cost of a Mac (about $1200, so don't expect too much).

The current (24 November 1988) menus are explained on the following pages.

File menu

New, Open, Close, Save, Save as.., Revert, Append

These commands follow the standard Macintosh practices for manipulating files. In this program the primary files are data representing the parameters for controlling vocal tract shapes. The data in any one file may represent up to 100 tract shapes.

Create individual tract makes it possible to enter data for the neutral tract of any subject. When doing this:

The upper incisors are regarded as fixed reference points.
The user may specify 16 points on the upper or posterior surface, and 16 points on the tongue.
The relative positions of the lower teeth, the lower lip, and the glottis can also be entered.
The uvula can be placed at any of the appropriate points on the upper surface.
The epiglottal pocket can be placed at any of the appropriate points on the surface of the tongue.

Save individual tracts and **Open individual tracts** operate on data files created by the previous command.

Open formants file opens a file of formant frequencies that had been created by previous use of this program.

Print does not work yet. (But, as in all Macintosh programs, a MacPaint file corresponding to whatever is on the screen can usually be obtained by holding down Command and Shift and typing 3; and the screen can be printed directly by holding down Command and Shift and typing 4.)

Quit does.

Edit Menu

Cut, Copy, Paste

These are the standard Macintosh items.
In this program they allow a vocal tract shape (or, to be more exact, the set of parameter values corresponding to a vocal tract shape) to be reordered within a set of tract shapes defining a sequence of articulatory positions.

Interpolate does a linear interpolation between any two tract shapes within the set.

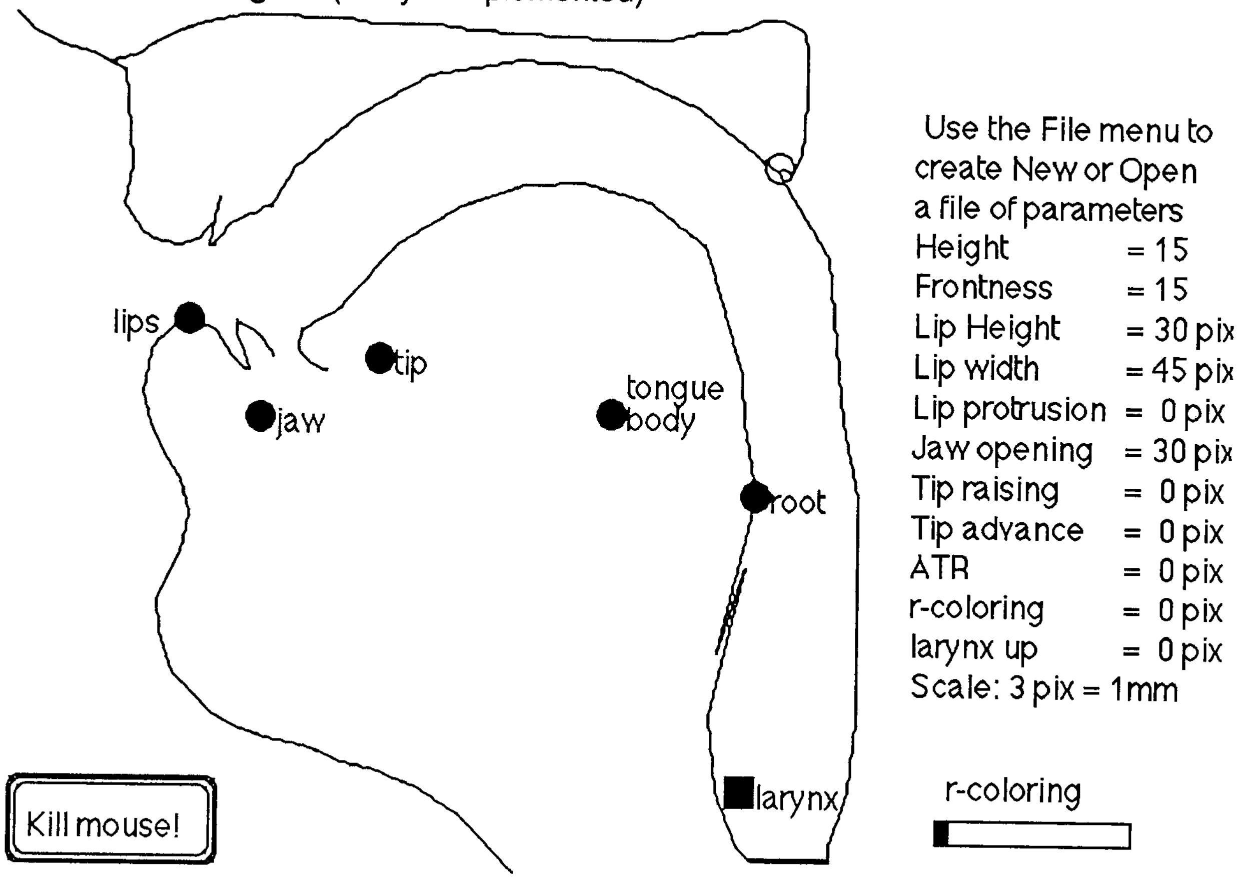
Smooth does a three point smoothing of the parameters in any sequence of tract shapes so as to avoid too abrupt changes.

Articulatory Menu

Activate mouse sets up points on the articulators (see diagram) that can be dragged into new positions. When a satisfactory shape has been achieved, clicking in "Kill mouse!" returns to the main menu.

Numerical values of the parameters listed below can also be entered.

High-Low	**OR**	**Front raising**	of the body
Front-Back	**OR**	**Back raising**	of the tongue
Lip-Height			
Lip width			
Protrusion		(rounding)	
Jaw opening			
Tip raising			
Tip advancing			
ATR		(Advanced Tongue Root)	
Rhotacization	(R-coloring)		
Larynx raising			
Velum lowering	(Not yet implemented)		



Use the File menu to create New or Open a file of parameters
Height = 15
Frontness = 15
Lip Height = 30 pix
Lip width = 45 pix
Lip protrusion = 0 pix
Jaw opening = 30 pix
Tip raising = 0 pix
Tip advance = 0 pix
ATR = 0 pix
r-coloring = 0 pix
larynx up = 0 pix
Scale: 3 pix = 1mm

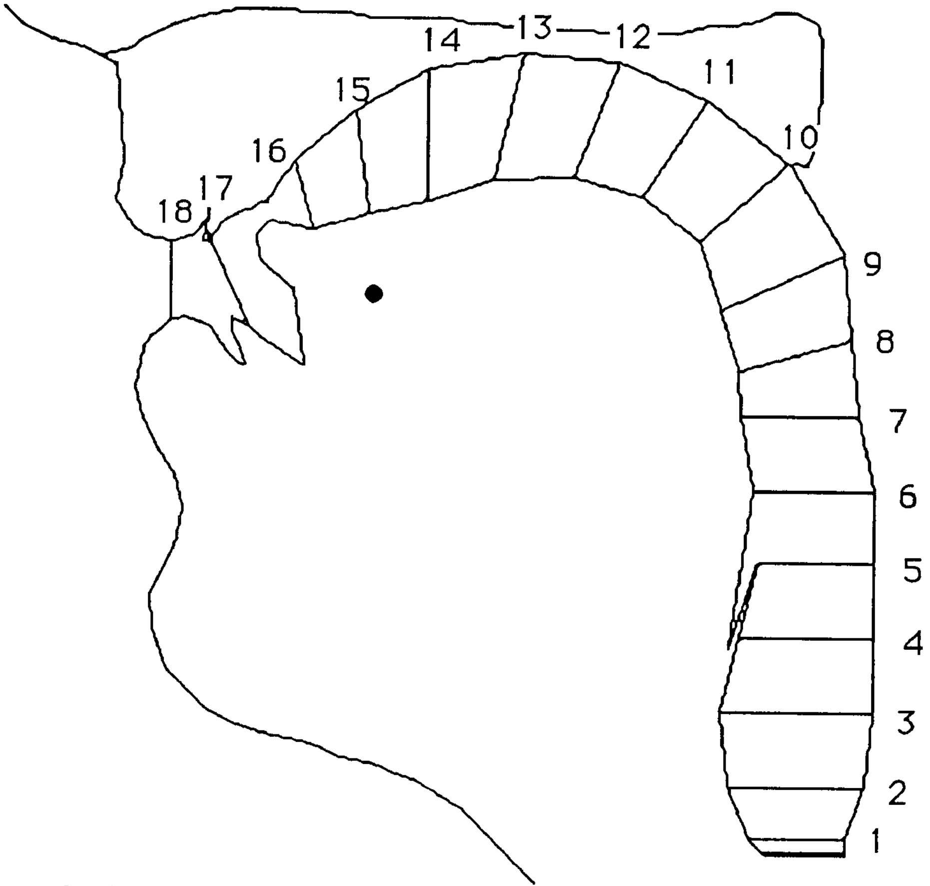
Articulations (Not yet implemented, but will (soon?) enable one to select one out of a set postures, that can then be made into a gesture using Interpolate and Smooth.)

A note on values of parameters

The values for most parameters are given in pixels (points on the Macintosh screen), with 3 pixels being equivalent to 1 mm. This allows users to specify movements with the greatest possible accuracy. Height and Frontness (the usual phonetician's terms for the position of the body of the tongue) are specified on an arbitrary scale of 0 - 30.* It is also possible to specify tongue movements in terms of the factors of Front Raising and Back Raising, which were found by Harshman, Ladefoged and Goldstein (1977) to account for tongue shapes in American English vowels with a high degree of accuracy. This possibility is noted in the section on the Options menu.

A note on movements of the articulators

There are a number of further points that should be noted in connection with the movements of the articulators. Points (1-15) on the body of the tongue move along the lines shown in the diagram below. Tip of the tongue raising is taken to be a rotatory movement around a point below line 15 (shown on the diagram below, but not normally visible). Tip advancing (or retracting) is a horizontal movement that moves the tip of the tongue by the specified amount; but it also spreads the movement over the body of the tongue, causing points on lines 9-16 to move forward (or backward). Upward (or downward) movement of the larynx affects principally the lowest section of the tract, but each of the lines below the epiglottis is also moved a proportionate amount.



There are further comments on the interaction between different articulators in the discussion of the Options menu.

Display Menu

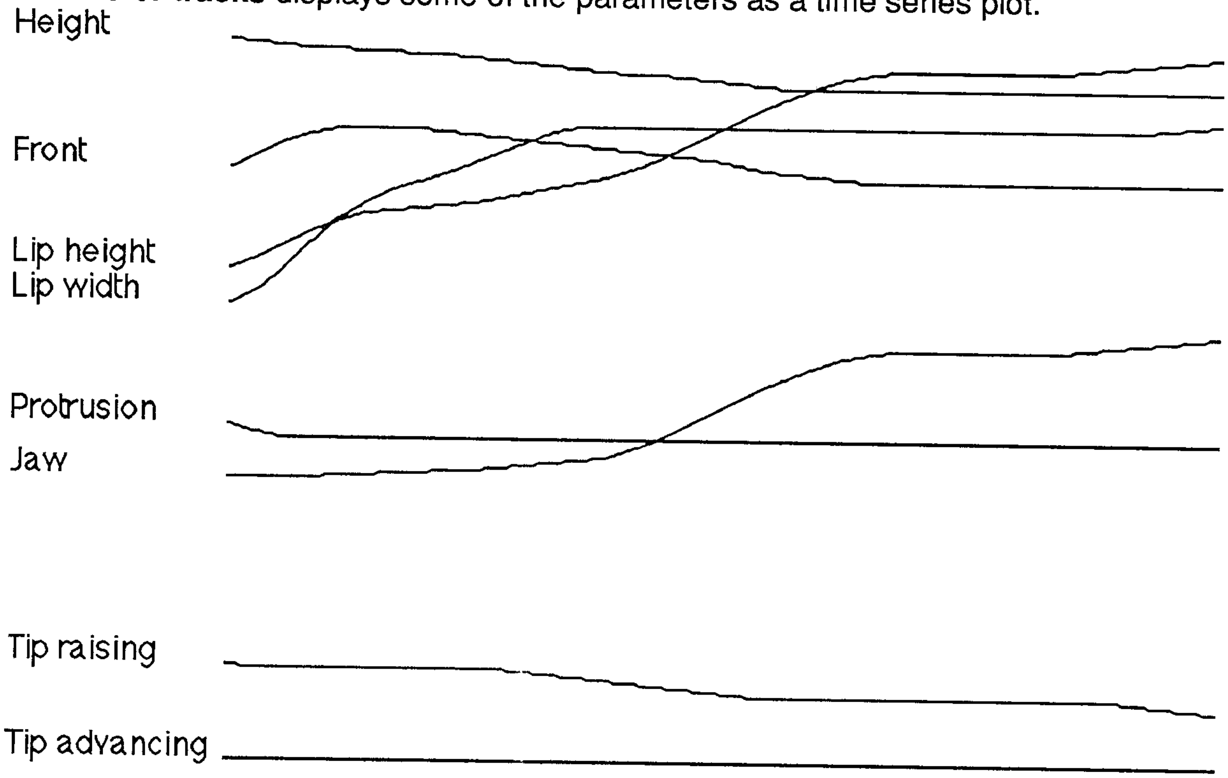
Single tract displays the requested tract.

Multiple tracts allows a number of tracts to be displayed (with any particular shape being emphasized by entering its tract number twice).

Parameter table displays the values of the parameters. This is mainly for checking purposes, as the parameters cannot be edited while viewing them in this form. (To do this, get out of the program and use a text editor.)

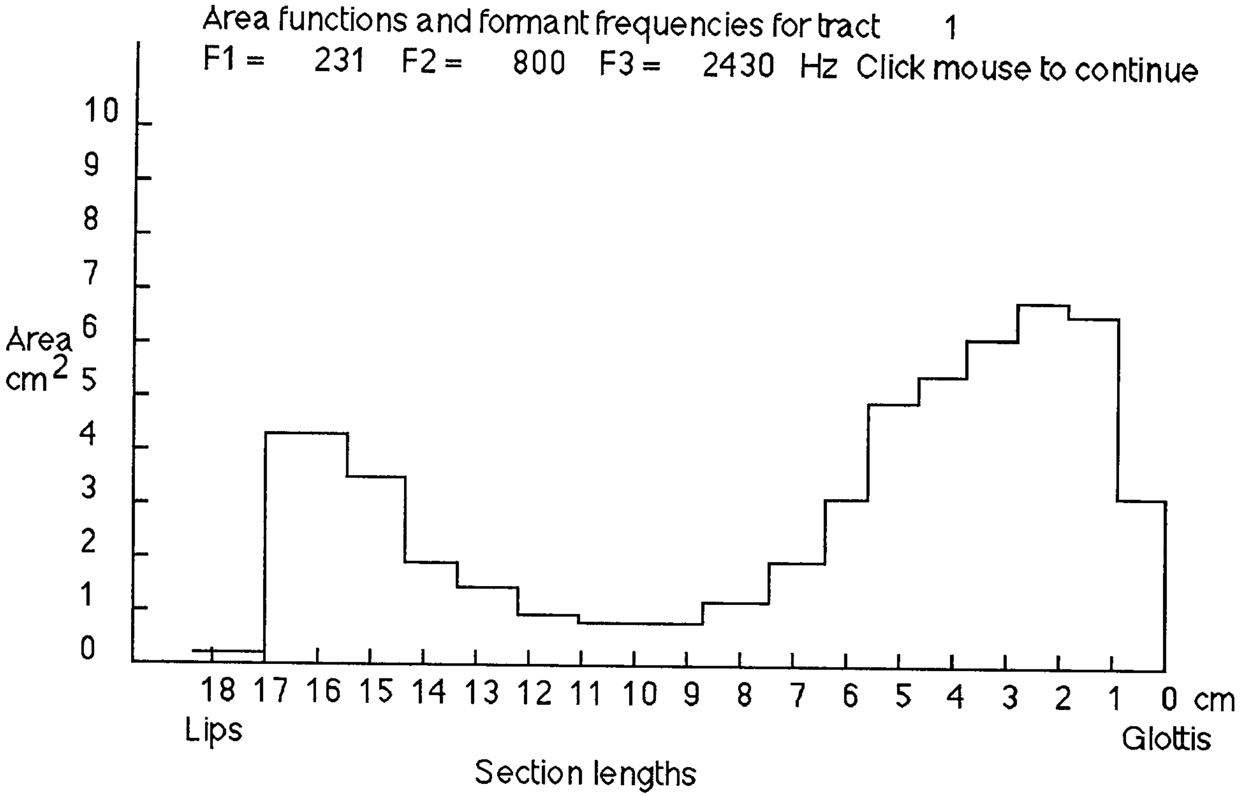
#	high	back	liphi	width	protr	jaw	tipup	tip+	AT	rhota	larnx	velum
1	24	11	8	30	5	35	0	0	0	0	0	0
2	23	16	12	35	2	35	-1	0	1	0	0	0
3	22	21	17	43	0	35	-1	0	2	0	0	0
4	21	24	22	53	0	35	-1	0	2	0	0	0
5	21	26	26	62	0	36	-1	0	1	0	0	0
6	20	26	29	68	0	36	-1	0	0	0	0	0
7	20	26	30	73	0	37	-1	0	0	0	0	0
8	19	26	31	76	0	37	-1	0	0	0	0	0
9	19	25	32	79	0	38	-1	0	0	0	0	0
10	18	24	33	83	0	38	-1	0	0	0	0	0
11	17	23	35	87	0	39	-1	0	0	0	0	0
12	16	22	37	91	0	40	-2	0	0	0	0	0

Parameter tracks displays some of the parameters as a time series plot.



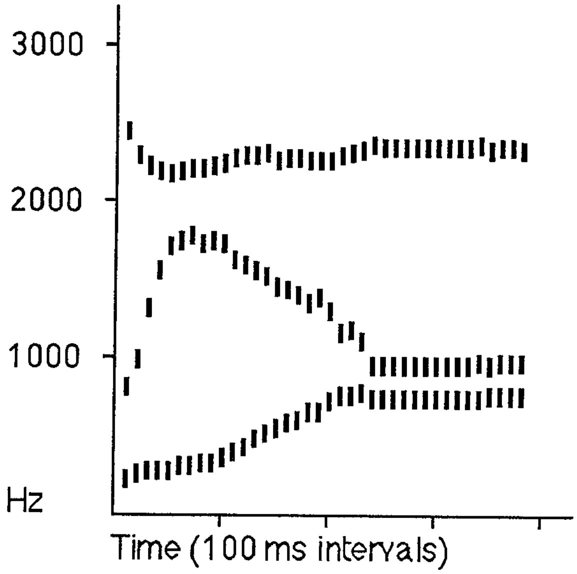
(Note: The illustration above shows all 38 rows of values in the file **'We are' tracts**, whereas the table shows only the first 12 values.)

Area function shows the lengths and areas of the set of tubes that was most recently determined when calculating the formant frequencies.



Area functions and formant frequencies for tract 1
F1 = 231 F2 = 800 F3 = 2430 Hz Click mouse to continue

Spectrogram shows the set of formant frequencies that have been calculated in the form of a pseudo-spectrogram.

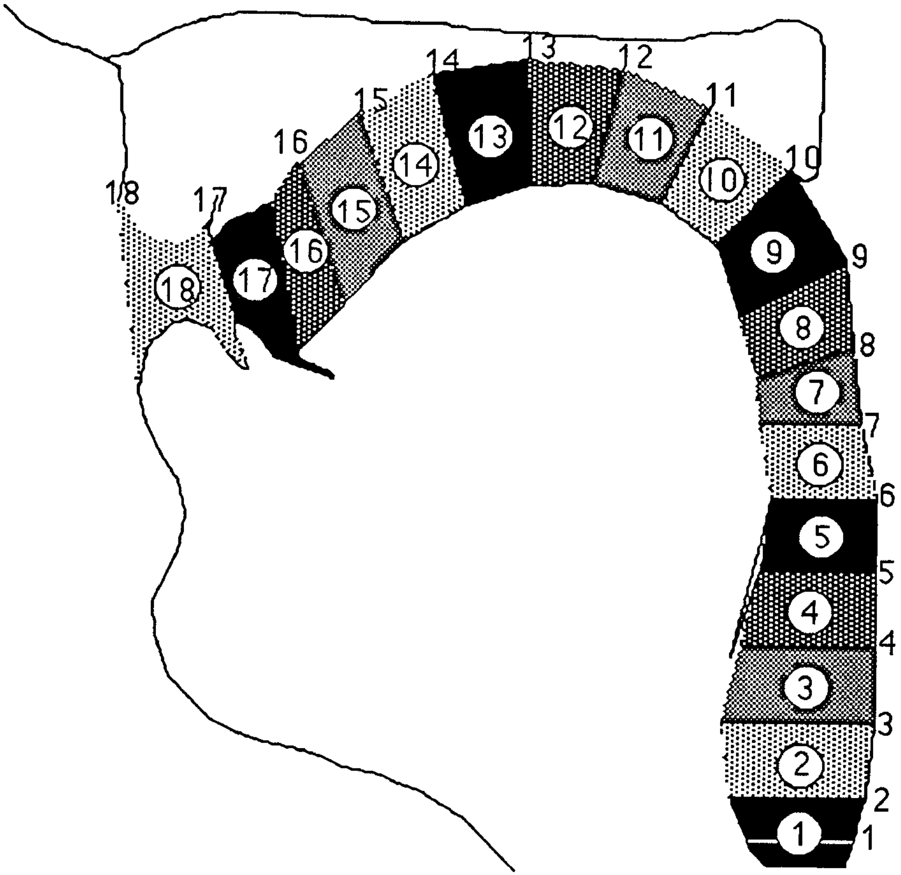


The illustration above shows the result of calculating the formant frequencies for the 38 vocal tracts in the file **'We are' tracts**, assuming that one occurs every 10 ms. The file of formant frequencies is stored in **'We are' formants**, so that it can be recovered for future plotting, or for statistical manipulations.

Formants: Single vocal tract
Formants: Whole file

These commands calculate the formant frequencies corresponding to one or more vocal tracts, using the following steps:

(1) The vocal tract is split into 18 sections. The first section is defined as being from the glottis to line 2 on the tongue. The boundaries between sections in the neighborhood of the tip of the tongue are adjusted so as to best account for the actual position of the tip of the tongue. The remaining divisions are on the lines used for moving the articulators. The diagram below shows the relation between the lines used for calculating the movements of the articulators, and the sections used in calculating formant frequencies.

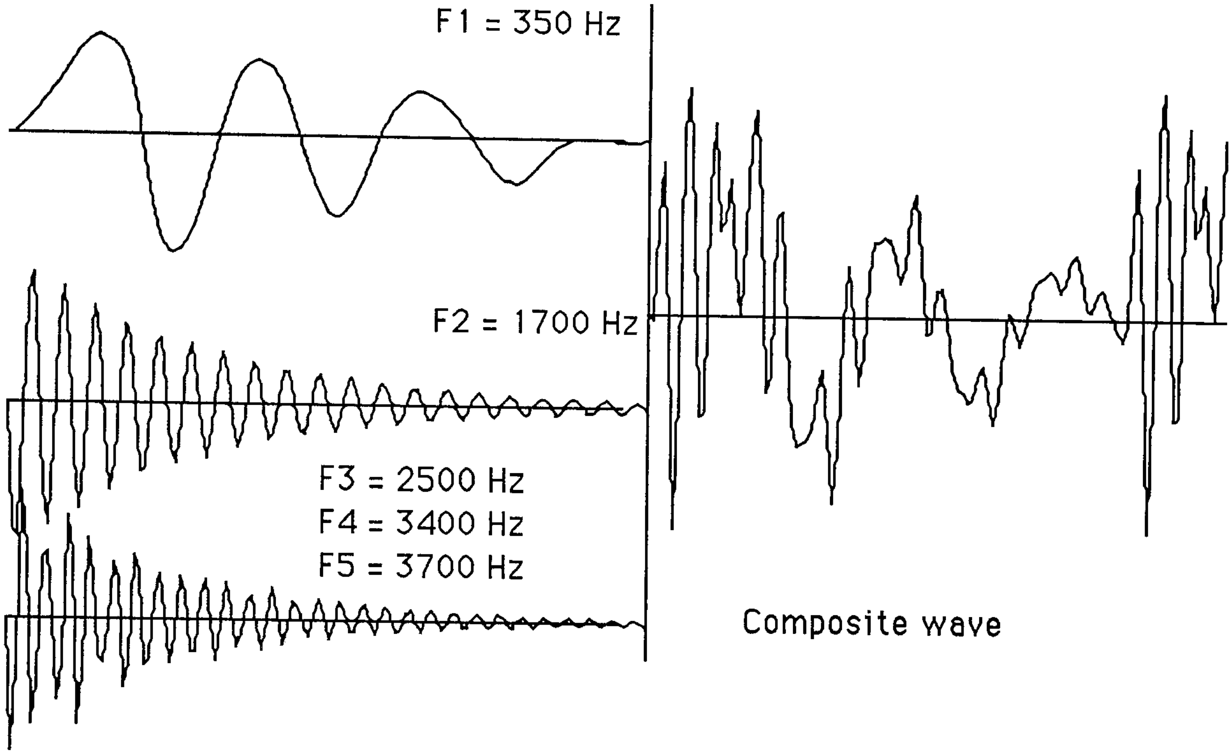


(2) The length and mean cross-section of each of the18 sections are calculated. The mean cross-sections are converted into areas by reference to a look up table based on the measurements of Ladefoged, Anthony & Riley (1971), modified to bring them more into line with the mean measurements reported by Wood (1982).

(3) The formant frequencies are calculated using an algorithm described by Liljenkrants and Fant (1975).**

Tract shape from formants. Input formant frequencies can be used as the basis for calculating one of the possible tract shapes that could have produced them. This is a very approximate calculation, based on the algorithm given by Ladefoged, Harshman, Goldstein and Rice (1978).

Play current vowel synthesizes a waveform corresponding to the set of formant frequencies for the displayed vocal tract (if they have been calculated), and then reproduces it over the Macintosh or (preferably) an external loudspeaker. The vowel is about 300 ms long, and has a slight rise-fall intonation pattern. A display of the first and second formants, the higher formants, and the composite wave is also provided.



In the illustration above the labels have been added to the screen dump, using a MacPaint type program. The waveform is synthesized by adding pre-calculated waves (stored in the file Wave Data). A damped wave is stored for each possible formant at 5% intervals from 230 - 3300 Hz. These waves are sets of 125 points (= 93 Hz at the Macintosh sample rate of 11,240 Hz), making this the longest period (lowest frequency) that can be synthesized. Up to 850 Hz the waves are modified versions of real speech waves taken from the vowel sequence [i-a], low pass filtered at 1000 Hz, and smoothed over the last 25 points. (A special program was written to cut out and modify suitable waves.) From 850 - 2200 Hz they are exponentially damped waves with a bandwidth that increases from 30 to 60 Hz. From 2200 - 3300 Hz they are similar waves, but with fixed higher formants of 3400 Hz and 3700 Hz added in. The program is written in a way such that individual formant amplitudes could be varied, but at the moment they are not.

Repeat plays the previously calculated wave.

Play whole file does (if you have calculated the formants for the whole file, or read them in from a previous calculation). At the moment, if the file is longer than 34 tracts, only the waveform corresponding to the first 34 tracts can be played. The built in slight rise-fall intonation will be added.

Formant vowel synthesis asks for the values of three formants, and then synthesizes and plays this vowel (without reference to the tract shape displayed), by adding together previously calculated damped waves and superimposing a rise-fall intonation pattern as described above.

Formant phrase synthesis plays the wave corresponding to a file of formants.

Options Menu

Display sections shows the lines along which 15 points on the body of the tongue are moved. (See illustration accompanying "A note on movements of the articulators".)

Show parameter values provides the numerical values of the parameters alongside the displayed tract. (See illustration in "Articulatory menu" section.)

Confirm display changes allows the user to determine whether to have to confirm every change (or to get rid of that annoying little box querying you every time).

Lips and tongue moved by jaw The jaw height determines, in part, the position of the tongue and the lip height (as might be required, for example, in the approach towards an [s]).

Lips and jaw moved by tongue height The tongue height completely determines the lip and jaw height (as may occur in some vowel sequences).

Protrusion controls lip width The degree of protrusion determines the lip width, allowing these two possibly separate gestures to be regarded as a single coordinative structure.

Factor based makes the movement of the tongue controlled by the factors determined by Harshman, Ladefoged and Goldstein (1977). Input values should be 1000 times the values given in the original publication, making them integers in the range -2000 to +2000, so as to keep them distinct from Front-back and High-low specifications

Help menu

The Help menu give on-line brief instructions on how to use various aspects of the program.

Footnotes

* There is a problem in using traditional tongue Height and Backness features to specify vocal tract shapes. How do we interpret a two number specification of this kind as a specification of the position of 16 points on the tongue surface? No one has ever shown how these traditional parameters should be used to specify the shape of the tongue as a whole. However, it is well known (Ladefoged 1982) that they can be related to acoustic parameters in a direct way. Accordingly, in our model specifications of the Height feature are re-interpreted so as to give an appropriate value for F1 (expressed in terms of Bark frequencies), and specifications of Backness are similarly re-interpreted as values of (F2-F1). From these values of F1 and F2 a

value of F3 is calculated, using the equation in Ladefoged and Harshman (1979). These three formant frequencies are then used to determine the degrees of Front-Raising and Back-Raising of the tongue, as described by Ladefoged, Harshman, Goldstein, & Rice (1978). The factors Front-Raising and Back-Raising are then used as described by Harshman et al. (1977) to generate the positions of the 16 points on the tongue.

This is a complicated way to get a specification of tongue shapes from values of the traditional terms; but, as far as we know, there is no better way to use values of the features Height and Backness to determine the locations of *all 16* points on the tongue. It is also interesting in that is a direct representation of the view proposed elsewhere (Ladefoged 1982) that the phonological feature Height should be regarded as having auditory correlates that determine the natural classes of sounds.

** Any offers of a piece of code that would enable us to do better synthesis would be gratefully received. We would love a procedure of the form shown below (stated in Pascal, so that we can give decent names, but C or even FORTRAN would be welcome):

```
Procedure FormantCalculation (sectionLengths, sectionAreas:
array [1..18] of real; formantFrequencies, formantAmplitudes: array
[1..5] of real);
```

References

Harshman, Richard, Ladefoged, Peter, & Goldstein, Louis. (1977) "Factor analysis of tongue shapes." Journal of the Acoustical Society of America **62** (3), 693-707.

Ladefoged, Peter (1982). A Course in Phonetics. Harcourt Brace Jovanovich: New York.

Ladefoged, Peter & Harshman, Richard (1979) “Formant frequencies and movements of the tongue.” In Frontiers of speech communication research (B. Lindblom & S. Öhman, editors) pp. 25-34. London: Academic Press.

Ladefoged, Peter, Harshman, Richard, Goldstein, Louis, & Rice, Lloyd. (1978) "Generating vocal tract shapes from formant frequencies." Journal of the Acoustical Society of America **64** (4), 1027-1035.

Ladefoged, Peter, Anthony, J.F.K. & Riley, C. (1971) "Direct measurement of the vocal tract." UCLA Working Papers in Phonetics **19**, 4 -13. Los Angeles: Phonetics Laboratory.

Liljenkrants, Johan & Fant, Gunnar (1975) "Computer program for VT-resonance frequency calculations." Speech Transmission Laboratory, Quarterly Progress and Status Report **4**, 15-20.

Wood, Sydney (1982) "X-ray and model studies of vowel articulation." Working Papers, Lund University, Department of Linguistics, 23: 1-191.

The UCLA Hypercard language data base

Jenny and Peter Ladefoged

As part of our NSF sponsored research, we are trying to make our data on the sounds of the world's languages more readily available. We have found that the Macintosh Hypercard system provides a very suitable method of achieving this aim. At present we have a data base illustrating interesting phonetic contrasts in more than 30 languages (see Figure 1). We anticipate that this collection will grow to many times this size in the near future.

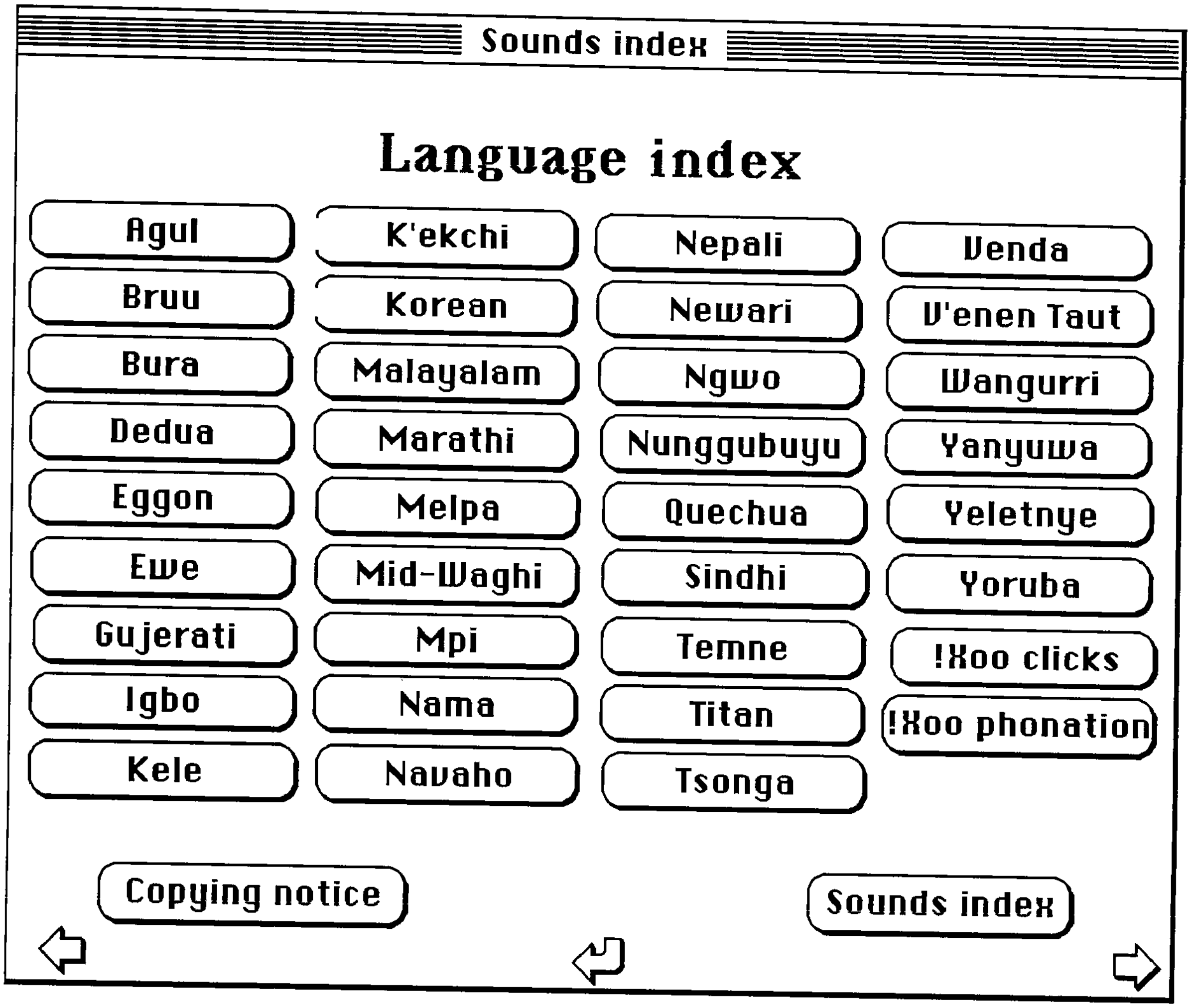
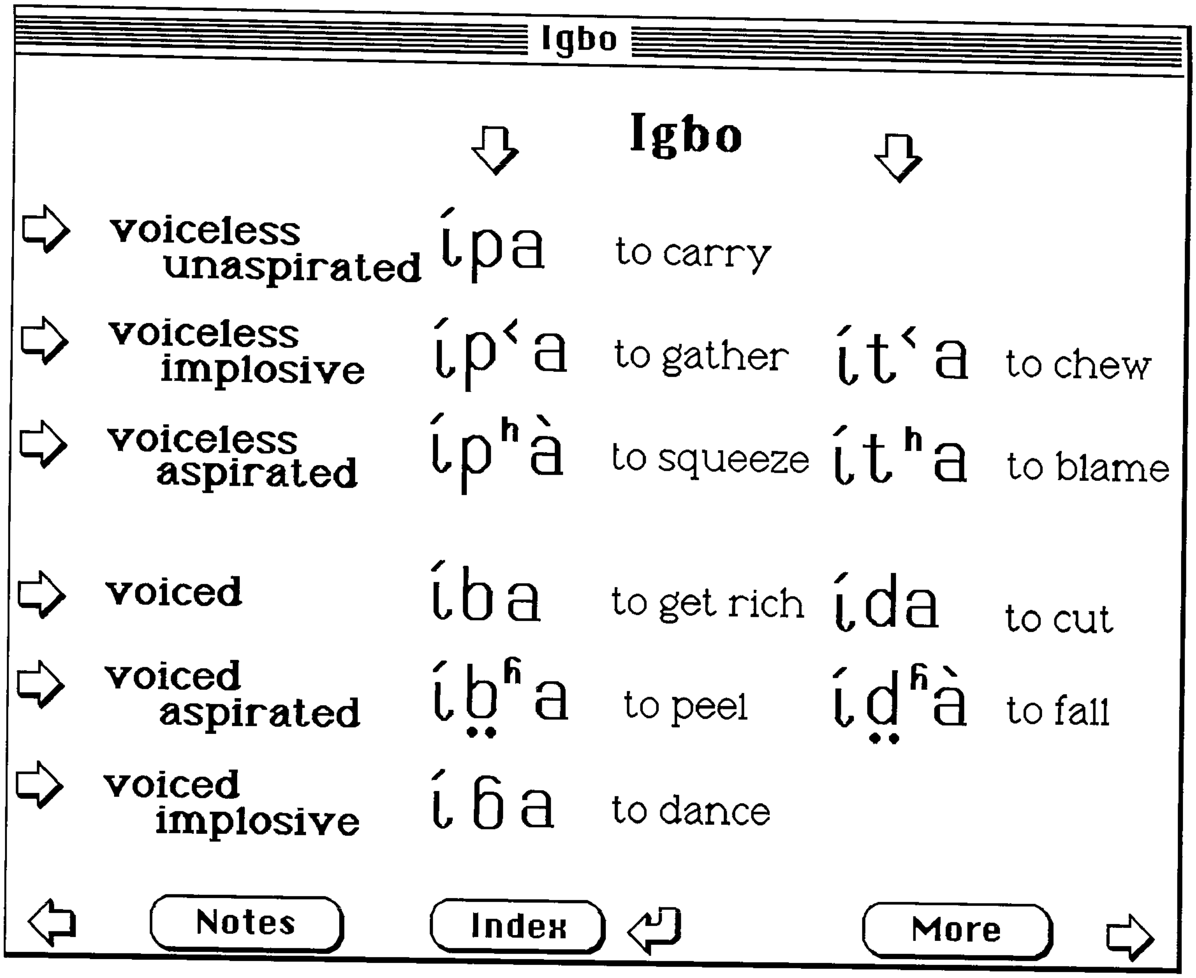


Figure 1. The computer screen listing the languages currently available in the data base.

The user can hear any of the languages shown on the computer screen reproduced in Figure 1 by using the mouse to click on the name. Clicking on Igbo, for example, brings up the computer screen shown in Figure 2. In general the data base contains only the contrasts of maximum interest in each language. To save computer disk space we have not digitized a complete set of phonological contrasts in each language. Clicking on any individual word on the screen plays a digitized recording of that word. Sets of words can be played by clicking on the arrows above or alongside. In every case the word or phrase has been spoken by a native speaker; most of the recordings are taken from our own fieldwork. The original recordings are digitized at a sample rate of 22 kHz, using 8 bit samples. The intensity level is carefully controlled so that the full range of 48 dB is maintained. In nearly all cases the digitized sounds are virtually indistinguishable from the original recordings.

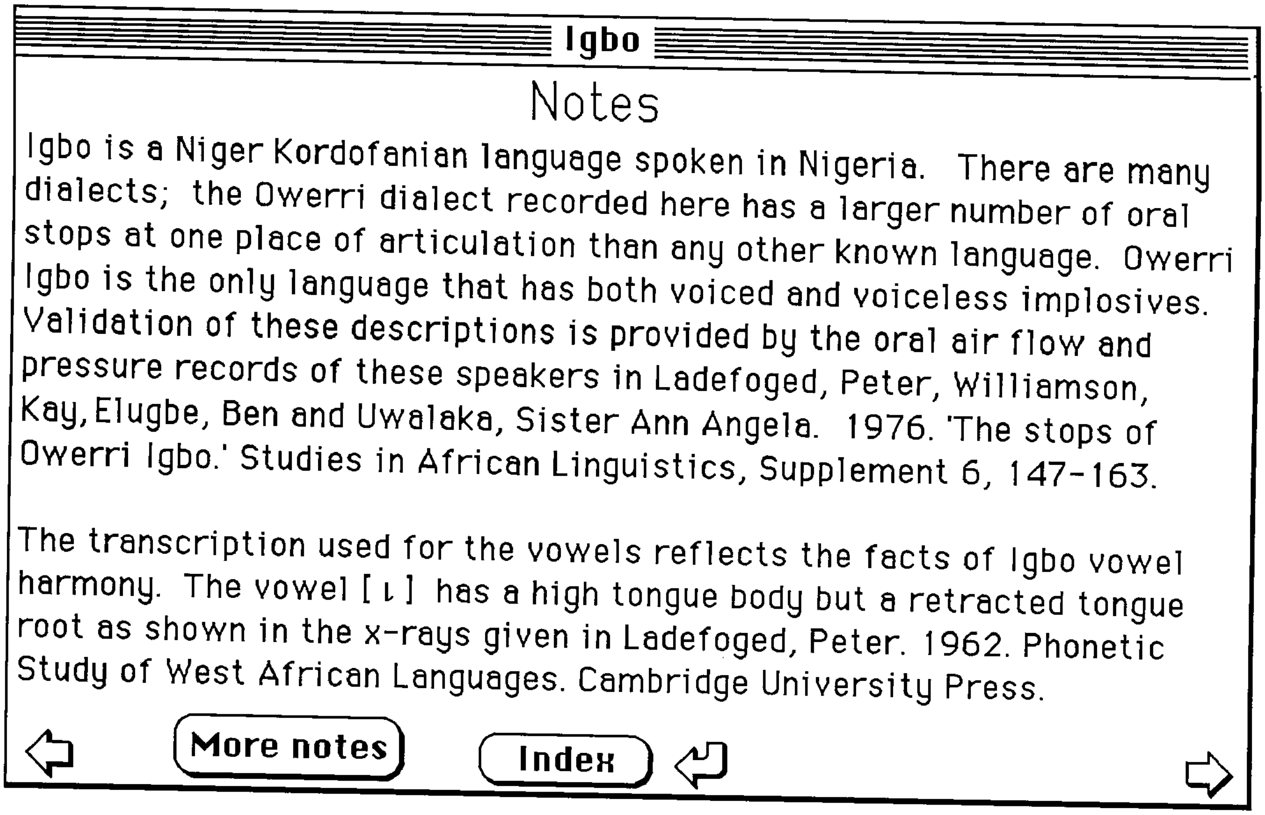


		Igbo		
voiceless unaspirated	ɩ́pa	to carry		
voiceless implosive	ɩ́p<a	to gather	ɩ́t<a	to chew
voiceless aspirated	ɩ́pʰà	to squeeze	ɩ́tʰa	to blame
voiced	ɩ́ba	to get rich	ɩ́da	to cut
voiced aspirated	ɩ́b̤ʱa	to peel	ɩ́d̤ʱà	to fall
voiced implosive	ɩ́ɓa	to dance		

Figure 2. One of the computer screens illustrating some of the sounds of Igbo.

For some languages we have also stored further information. Clicking on "Notes" in the Igbo display brings up the screen shown in Figure 3; and further screens, such as that shown in Figure 4 are also available. The recordings of these words can be heard by clicking directly on the display of the instrumental records. (In this case the recordings are of different tokens, as the flow mask used in making the aerodynamic records impedes high quality recording.)

We have found that this data base is already proving a valuable teaching resource. The possibility of being able to hear words in any order, without having to cue up a tape recorder, makes this system a very useful way of learning to appreciate hard to hear contrasts. The minimal equipment needed for using the data base is a MacPlus with an external drive. The addition of an amplifier and loudspeaker is strongly recommended, as the small Macintosh internal speaker is not capable of providing the reasonably high quality that is present in the recordings. Extended use of this data base is not really practical without a hard disk. At the moment there are about 12 meg of data stored. Anyone sending us sufficient floppy disks (or equivalent cash) is welcome to a copy of all or any part.



Notes

Igbo is a Niger Kordofanian language spoken in Nigeria. There are many dialects; the Owerri dialect recorded here has a larger number of oral stops at one place of articulation than any other known language. Owerri Igbo is the only language that has both voiced and voiceless implosives. Validation of these descriptions is provided by the oral air flow and pressure records of these speakers in Ladefoged, Peter, Williamson, Kay, Elugbe, Ben and Uwalaka, Sister Ann Angela. 1976. 'The stops of Owerri Igbo.' Studies in African Linguistics, Supplement 6, 147-163.

The transcription used for the vowels reflects the facts of Igbo vowel harmony. The vowel [ɩ] has a high tongue body but a retracted tongue root as shown in the x-rays given in Ladefoged, Peter. 1962. Phonetic Study of West African Languages. Cambridge University Press.

Figure 3. An example of the short background information given on one of the languages in the data base.

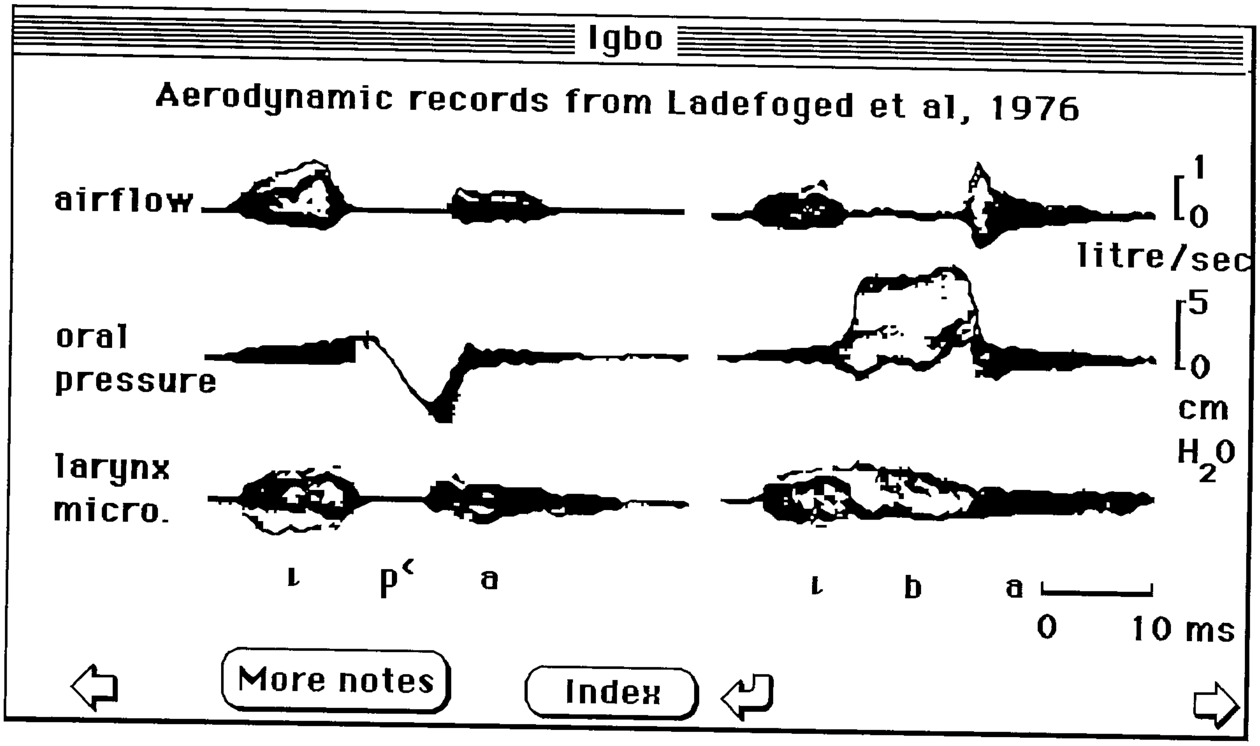


Figure 4. An example of instrumental data that is available for a few languages.

A computerised phonetic dictionary

Thanks to a contract from IBM, we have recently completed a revision of an extensive phonetic dictionary. This dictionary had its origins in work by Dr. June Shoup, at the university of Michigan. The current version contains about 250,000 entries, each showing the word in orthography, the part of speech, and an IPA transcription of the most favored pronunciation in California, followed by other possible standard Californian pronunciations. The large number of entries results from the fact that the dictionary contains many derived forms, such as plurals and different verb forms, many of which would not normally be considered as separate lexical items.

The transcriptions were made by a group of students under the direction of Cheryl Chan; all the members of the group were native speakers of Californian English. The transcription of each word was based on judgments of the standard pronunciation in California. These transcriptions were compared with those given by J.S. Kenyon and T.A Knott, *A Pronouncing Dictionary of American English* G. & C. Merriam: Springfield (1953), and with the pronunciations given in *The Random House Dictionary of the English Language* (2nd edition, unabridged) Random House: New York (1987), and other standard pronunciations were incorporated. Some error checking was done by means of a computer program (written by Juan Alvarado) which detected various type of erors, including unmarked or invalid stress numbers, double primary stress on non-compound words, improperly coded parts of speech, improper formatting, erroneous characters in the orthography, missing transcriptions, etc. We are, however, well aware that there are still a number of mistakes.

The dictionary is stored as a set of MS DOS files, and is currently resident on an IBM AT computer. A program has been written (also by Juan Alvarado) to allow any person familiar with IPA symbols to search for entries of interest based on any of the following criteria: orthography, transcription, stress pattern, and part of speech. There is a manual explaining the use of the program. Copies of the dictionary, associated programs, and the manual are available on $5^1/_4$" floppy disks. The current version takes about 30 floppy disks and is available for non-profit research use only at a cost of $75, or $50 if the floppies are returned after having been copied.

P.L.

UCLA Software

UCLA SoundWave

This is a customized version of the commercial program SoundWave™, making it possible for the user to display a waveform on different time scales. It is also possible to use the mouse to measure durations of portions of the waveform, and to have these measurements written out into a file. This is by far the quickest way of measuring durations known to us. We also attempted to modify the analysis routines, but eventually decided that they were unsuited for phonetic work, since they use integer arithmetic. Impulse Inc. (the publishers of SoundWave) have said that they have no objection to our making copies of this customized version available.

MacSynth

MacSynth is a program for synthesizing speech either from numeric values (which can be in a file) or from free-hand drawn values (input with a mouse) of formant frequencies and amplitudes, and amplitudes of the fricative components. It is a very limited synthesizer, but has its uses for teaching purposes.

Vocal

This program computes the formant frequencies which correspond to a given vocal tract shape, and vice versa. The formant frequencies are calculated using a slightly modified version of the algorithm described by Liljenkrants and Fant (1975). The displayed vocal tract may be manipulated using the mouse. A fuller account of this program is given elsewhere in this volume.

IPAPlus

This is a font for the MacIntosh developed at UCLA which has symbols for the majority of IPA symbols and some supplementary characters and diacritics. This font is available from the Phonetics Lab, but an aesthetically more satisfactory version with a more rational keyboard layout is available from Linguists Software, 925 Hindley Lane, Edmonds WA 98020 (Phone: 206 775-1130).

PlotFormants

This program reads in sets of formant values in Hz and plots the corresponding vowels in a two-dimensional display, using mel or Bark scales. The program can calculate and display F2' and show the distribution of a cluster of vowels by an elipse centered on the mean position and with radii of one standard deviation along the axes of the principal components of the distribution. It has recently been modified to have a more traditional Macintosh interface, but the new version is not yet fully debugged.