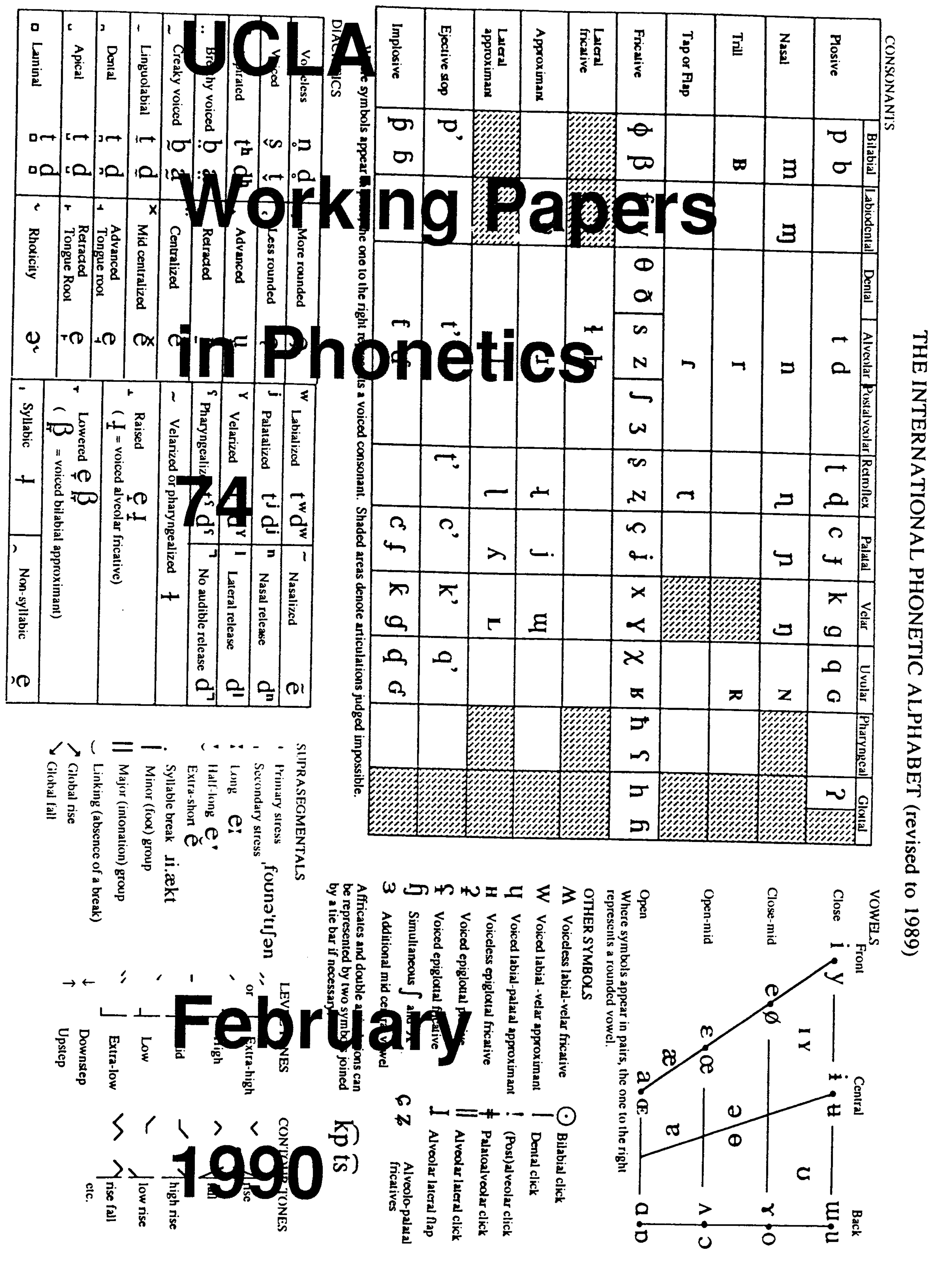
UCLA

Working Papers in Phonetics

74

February 1990



The UCLA Phonetics Laboratory Group

Beatriz Amos
Vernon Austel
Sue Banner-Inouye
Barbara Blankenship
John-DongwookChoi
Sarah Dart
Sandy Disner
David Feldman
Vicki Fromkin
Robert Hagiwara
Bruce Hayes
Susan Hess
Caroline Henton
Pat Keating
Paul Kirk
Marianna Knox-Seith
Jenny Ladefoged
Peter Ladefoged
Mona Lindau
Ian Maddieson
Kristin Precoda
Maria Psihountas
Bonny Sands
Stephan Schuetze-Coburn
Andrea Schwartz
Steve Solomon
Donca Steriade
Henry Teheranizadeh
Camille Teran
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:

NSF grant BNS 87-20098
USHHS grant NS-22726
USHHS grant 1 RO1 DC00642
USHHS grant 1 T32 DC00029
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 74

February 1990

Table of Contents

Kabardian vowels revisited

John-Dongwook Choi

Introduction

The vowels of Kabardian, a Circassian language spoken in the northwest Caucasus, have been the object of much discussion among linguists. While there is general consensus that Kabardian has a 'vertical' vowel system, i.e. a system in which the contrastive function is maintained solely along the height dimension, there is considerably less agreement as to how many vowels exist along this vertical axis. The debate centers on whether there are two vowels (high and mid) or three vowels (high, mid, low). Early analyses of the language by Jakovlev (1923) and Trubetzkoy (1925) argue for a three-vowel system. For Jakovlev, the distinction between the higher two vowels is qualitative while the difference between the lower pair of vowels is one of duration as well as quality, i.e. /ɨ, ə, aː/ (transcription mine). Trubetzkoy agrees that there are three vowels, but states that the length difference between the lower two vowels is not distinctive, arguing that the length of [aː] is 'concomitant' to the degree of aperture.

These analyses were challenged later by Kuipers (1960) who argues for a vowel-less analysis of the language. I will not recapitulate Kuipers' analysis here and only point out that this analysis has not generally been accepted. More recently, however, linguists such as Halle (1970) and Colarusso (1975; 1988), have adopted certain aspects of Kuipers' analysis and argued that the low phoneme can be eliminated, reducing the vowel inventory to two, i.e. /ɨ, ə/ (transcription mine). Basically, the low vowel is analyzed as a vowel+glide sequence in the same way as is possible for the phonetic long high and long mid vowels [i:, e:, u:, o:] which occur in the language, and which are considered to be /ɨj, əj, ɨw, əw/, respectively. Evidence that [i:, e:, u:, o:] are phonologically vowel+glide sequences comes from alternations of the type shown in the following forms (Catford 1984: 33):

/bzɨy/	[bzi:]	'light-ray'	/bzɨy+f'ɨ/	[bzijɨf']	'good light-ray'
/bzɨw/	[bzu:]	'bird'	/bzɨw+f'ɨ/	[bzuwɨf']	'good bird'
/ʒəy/	[ʒe:]	'sleep!'	/ʒəy+n/	[ʒejɨn]	'to sleep'
/psəw/	[pso:]	'live!'	/psəw+n/	[psowɨn]	'to live'

In word final position, the glides delete and the vowels are compensatorily lengthened. Phonetically, these long vowels are sometimes slightly diphthongized, particularly [o:] as [ou:].

Kuipers eliminates the low vowel by extending this coalescence process to [aː], deriving it from /əh/ (cf. Jakovlev for whom /aː/ was phonemic). [aː] is thus analysed as a long allophonic variant of the mid vowel /ə/ conditioned by a following /h/. The arguments for this reduction are two-fold, based first on distributional evidence and secondly on morphophonemic evidence. The distributional argument centers around the fact that [aː] is the only vowel that occurs word initially in Kabardian. This is supplemented by the observation that phonetically, initial [aː] is preceded by a murmured [ɦ] for some speakers. These points are taken as evidence that the word initial low vowel is phonemically /hə/, and that in non-word initial position, this sequence undergoes metathesis followed by the same process that derives the long nonlow allophones from phonemic vowel+glide sequences. Kuipers argues further that this accounts for the plural morpheme [aː] which he analyses as underlyingly /hə/.

This paper presents data from a spectrographic study of Kabardian vowels that are relevant to this debate. I will first outline the procedures taken in the study and present the data. I will then re-examine the arguments for a two-vowel vs three-vowel analysis in light of this data, as well as compare the results of this study with those of a similar study conducted by Catford (1984). Finally, I will present an alternative analysis arguing that Kabardian is, in fact, a three-vowel language in which duration is used redundantly to increase the perceptual salience of the otherwise minimally contrastive vowels.

Phonetic study

Data

The data used in the current study consist of connected speech. Three native speakers of the literary Terek dialect were asked to read a short text from Shagirov's article on Kabardian in *Jazyki Narodov SSSR* (Shagirov 1967: 182). This text contains 118 vowels. The speakers (henceforth S_1, S_2 and S_3) were all educated adult males. The text does not represent a complete, systematic sample of all consonantal environments with controlled prosodic parameters that a carefully constructed, exhaustive word list, for example, would have provided. However, there are enough tokens in the relevant contexts to form a preliminary assessment of allophonic vowel distribution as well as examine duration.

A phonemic inventory of Kabardian consonants is presented in Table 1.

Table 1. Phonemic consonant chart.

		labial	alveolar	alveo-palatal	palato-alveolar	palatalised velar	velar	rounded velar	uvular	rounded uvular	pharyngeal	laryngeal	rounded laryngeal
stops	voiceless	p	t			k´		kʷ	q	qʷ		ʔ	ʔʷ
	voiced	b	d			g´		gʷ					
	glottalic	p’	t’			k’´			q’	q’ʷ			
affricates	voiceless		ts										
	voiced		dz										
	glottalic		ts’										
fricatives	voiceless	f	s ɬ	ç	ʃ		x	xʷ	χ	χʷ	ħ		
	voiced	v	z ɮ	ʑ	ʒ		ɣ		ʁ	ʁʷ			
	glottalic	f’	ɬ’	ç’									
	nasals	m	n										
	trills		r										
	glides	w				j							

Table 2. Duration-related environments (post-vocalic)

Figures refer to the total number of tokens measured for duration in the respective environment. This table does not include the vowels [i:, e:, u:, o:].

	high	mid	low
voiced obstruents	27	33	21
voiceless obstruents	42	48	15
nasals/approximants	18	75	27

Table 3. Quality-related environments (pre-vocalic)

Figures refer to the total number of tokens measured for formant frequencies in the respective environment. This table does not include the vowels [i:, e:, u:, o:].

	high	mid	low
alv-pal, pal-alv, pal-vel	24	15	21
velar	3	15	3
uvular	3	15	3
rounded	21	27	9
pharyngeal	0	9	0
elsewhere	.36	75	33

Table 2 lists the environments that are known to affect vowel duration along with the total number of tokens measured for duration in the respective environments; e.g. there are 27 high vowels measured for duration in the environment preceding a voiced obstruent. Prosodically, the only context taken into consideration in calculating the proportion of vowels represented in duration-related environments was the presence or absence of a pause. The effect of pause was not consistent across the speakers, however, and is addressed further in the discussion of duration.

Table 3 provides a breakdown of the text according to the consonantal environments that precede the vowels. The particular differentiation of the contexts listed follows from descriptions in the literature which state that vowel quality is determined by the quality of adjacent consonants (Kuipers 1960). All the consonantal environments that are described as triggering allophony are represented in the text before all vowels. The only exception to this involves the pharyngeal; the text provides only one token of the mid vowel in this environment. There are no instances of the high and low vowels following a pharyngeal consonant.

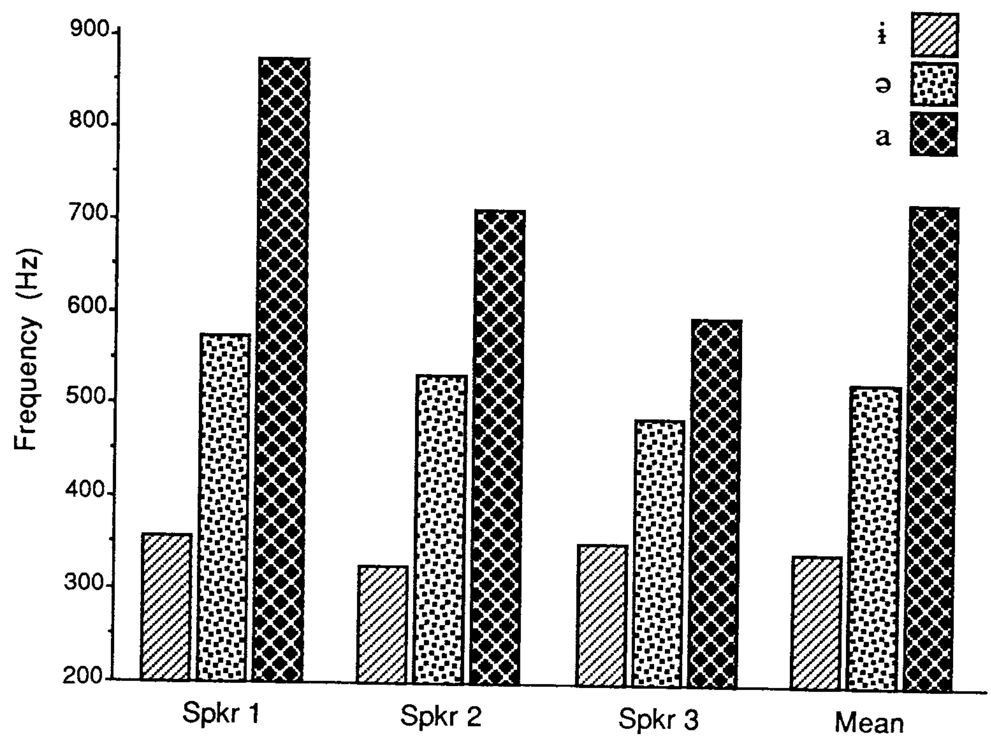
Procedures

The data were 12 bit sampled at a rate of 10 kHz on a digital spectrograph. Wide band spectrograms up to 4 kHz were examined, in tandem with waveform displays, to identify the vowels and measure their duration. Segmentation was conducted following the procedures outlined in Peterson & Lehiste's (1960) study of duration in English when possible. The first three formant frequencies for each vowel were measured from wide band FFT power spectra with a filter bandwidth of 300 Hz. At this resolution, the FFT transform length was 4.883 msec. Cursors were used to define a 25 msec frame in the middle of the steady state of each vowel. A series of FFT spectra were then taken within this frame, and the mean of this series was displayed. Given that some of the vowels were as short as 30 milliseconds in duration, the window could not be enlarged without deleting tokens from the database. Most of the long vowels were monophthongal; in the cases where the long vowel was slightly diphthongized, there was usually a distinct steady-state vowel portion which was taken as the locus for FFT analysis. There were a few tokens where no steady state was available, in which case the analysis frame was positioned in the middle of the diphthong.

Table 4. Mean formant frequency values
Individual and overall mean formant frequency values collapsing across differing consonantal environments.

	Vowel	F1	F2	F3
Speaker #1	ɨ	341	1545	2506
	ə	519	1460	2455
	a	702	1381	2483
Speaker #2	ɨ	329	1474	2609
	ə	500	1410	2522
	a	696	1330	2512
Speaker #3	ɨ	342	1399	2377
	ə	460	1300	2250
	a	585	1278	2206
Overall	ɨ	337	1473	2497
	ə	493	1390	2409
	a	661	1330	2400

Figure 1. Mean F1 values



Results

F1 Variation across /ɨ/, /ə/ and /a/

The mean formant frequency values for the allophones of /ɨ/, /ə/ and /a/, excluding the long vowels [i:, e:, u:, o:], are presented in Table 4. These figures represent the vowels collapsed across differing consonantal environments. The F1 values in Table 4 are represented in bargraph form in Figure 1. As the histogram shows, these data provide a clear argument for maintaining three distinct vowel heights. All three speakers exhibit a difference in F1 between the high and mid vowels (mean difference of 156 Hz) that is comparable to that between the mid and low vowels (mean difference of 168 Hz). If we assume, therefore, that /ɨ/ and /ə/ are distinct and that this distinction is one of vowel height, then /ə/ and /a/ show as great a distinction in phonetic vowel height.

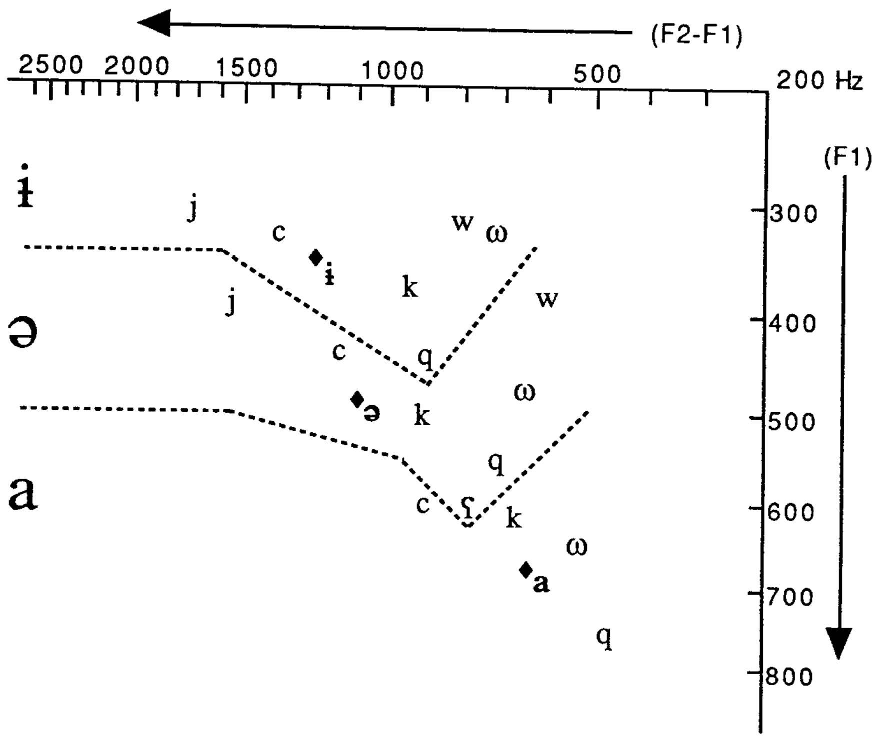
Vowel Allophony

Considerable variation in the phonemic vowels' values for F2 is exhibited phonetically. This variation is contextually determined by the quality of the preceding consonant as shown in Figure 2 and accords with Kuipers (1960) treatment of the assimilation process as a progressive one. Attempts to plot the vowels in similar fashion based on the following consonantal environment results in a much more chaotic distribution. Each point in the vowel space corresponds to the mean formant frequency values for the relevant vowel in the consonantal environment specified. The long vowels [i:, e:, u:, o:] are denoted with j and w; vowels in the environment of consonants with a palatal component are denoted with c, those in the environment of plain velars with k, those in the environment of plain uvulars with q and those in the environment of rounded consonants with omega. The mid vowel in the environment of a pharyngeal consonant is denoted by ħ. Labials, alveolars and plain laryngeals do not influence vowel quality. Vowels following these neutral consonants are represented by the diamonds and provide reference points from which to discuss the other variants. The dashed lines show how the underlying vertical contrasts are maintained in the vowel space.

Consonantal influence on following vowels is quite systematic. Vowels in the environment following alveo-palatals, palato-alveolars and palatalized velars show a lowering of F1 and an increase in F2 vis-a-vis their neutral counterparts. The increase along the F2 parameter, however, is not commensurate to that induced by a following [j]. The [i:] and [e:] vowels occuring before /j/ are approximately 350 Hz - 500 Hz greater on average in their values for F2. Rounded consonants trigger a decrease in F2 comparable to

Figure 2. Allophonic distribution

Each symbol represents the mean formant frequency values for the vowel when preceded by consonants of the symbolized type, averaged over all three speakers. The consonant types are discussed in the text.



that preceding the glide /w/. The vowels following these rounded consonants also seem to exhibit slightly higher F1 values as well. In the case of the mid vowel /ə/, however, /əw/ and /Cʷə/ seem to comprise two distinct allophones given the large difference in F1. After uvulars, the vowels show a sharp increase in F1 as well as a considerable decrease in F2. Velars also trigger a decrease in F2 for the high and mid vowels, but not for the low vowel. This is not too surprising though if we consider that the low vowel in a neutral environment is already considerably backed. The asymmetry in the influence of the velars on F1 is similarly unsurprising; the tongue body must rise to contact the velum, causing a decrease in the first formant of a low vowel. Lastly, the mid vowel following /ħ/ exhibits a very sharp increase in F1 accompanied by a slight decrease in F2, bringing [əħ] into the space of the lower vowel. The text did not contain any tokens of the high or low vowel in this environment.

The quality of the vowels produced by the three speakers in this study, as well as the distribution of the various allophones, matches the descriptions in the literature. What is of particular interest is that the low vowel exhibits allophonic variation similar to that seen in the high and mid vowel, the only difference being that [a] does not combine with the glides. This asymmetry is used by Kuipers (1960) to motivate the /əh/ analysis of the phonetic low vowels.

Duration

A mean duration value for each short vowel was determined for each speaker and is shown in Table 5. In calculating these mean value, vowels occuring in prepausal position were disregarded. The frequency of pause intervals varied across the three speakers and across the different vowel heights as is shown in Table 5. Since pauses condition greater vowel length, as is illustrated in Table 6 by comparing the duration of prepausal and nonprepausal mid vowels, the mean duration values would have been skewed if the prepausal vowel tokens were included.

As Figure 3 shows, the low vowels are longer than the mid vowels. This does not, however, necessarily constitute evidence for a distinctive length contrast in Kabardian. More generally, the data show that duration increases as aperture increases for the three vowels /ɨ, ə, a/. A correlation between vowel height and duration has been observed in many languages. In her discussion on quantity, Lehiste (1970) lists English, German, Danish, Swedish, Thai, Lappish, and Spanish among the languages which fit the generalization that a high vowel is shorter than a low vowel, all other facts being equal. We can refer to this relationship as intrinsic vowel duration. What is of interest in the case of Kabardian, however, is whether the difference in duration constitutes evidence for a distinctive length contrast or not. The question might be reworded in another way: i.e., how large a difference is there be between long and short vowels in cases where the difference is linguistically significant, and how does this compare with intrinsic durational differences?

This question is answered by comparing the duration differences reported in languages said to have a phonemic length contrast to duration differences reported for intrinsic duration. To illustrate, mean duration ratios (V:/V) were calculated for the long and short vowels in Swedish (Elert 1964) and German (Zwirner 1937) and plotted in Figure 4. For contrast, the mean duration ratio (lower V/higher V) for Swedish [e] and [a] (Elert 1964) and English [ɪ] and [ɛ] (Peterson & Lehiste 1960) were also calculated and plotted. All figures come from connected speech and/or words in carrier sentences, and not word lists.

Figure 3. Mean duration (in milliseconds)
Mean values of all nonprepausal tokens.

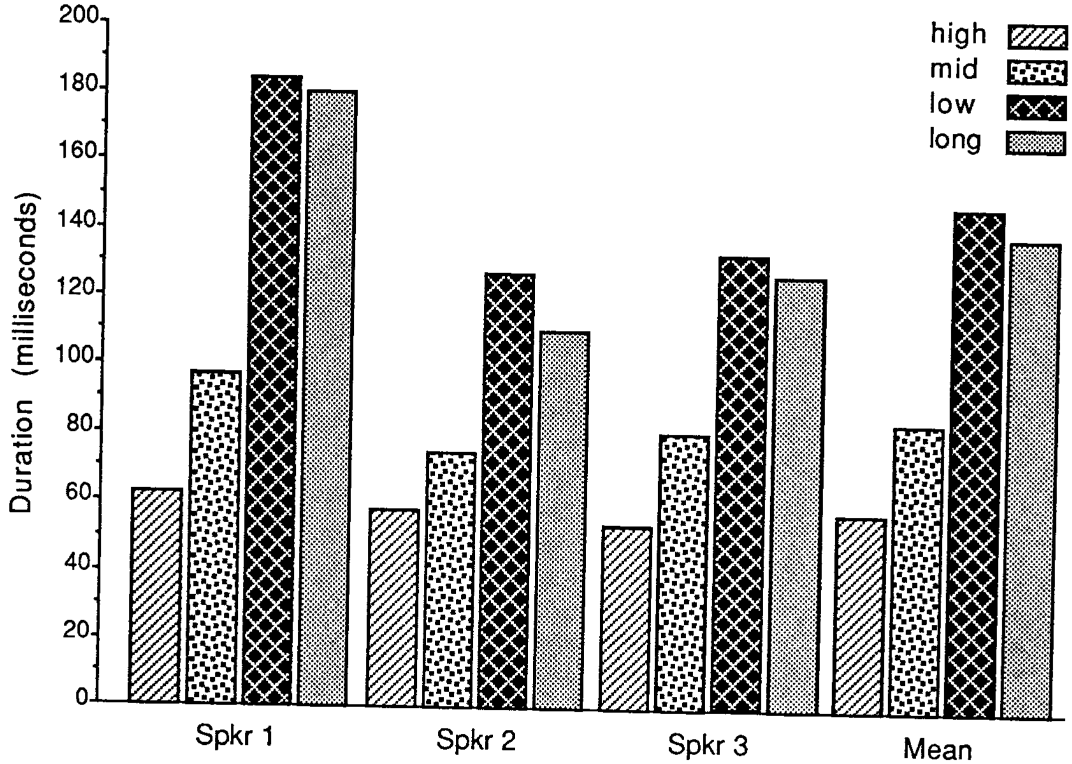


Table 5. Variation in the distribution of prepausal vowels
The first figure represents the number of vowels in prepausal position; the other figure represents the percentage that number represents of the total number of tokens.

	high (29 tokens)	mid (52 tokens)	low (21 tokens)
Speaker #1	2 / 04%	15 / 29%	2 / 10%
Speaker #2	3 / 07%	10 / 19%	0 / 00%
Speaker #3	2 / 04%	17 / 33%	1 / 05%

Table 6. Mean duration of the mid vowel in prepausal and nonprepausal positions (in milliseconds)

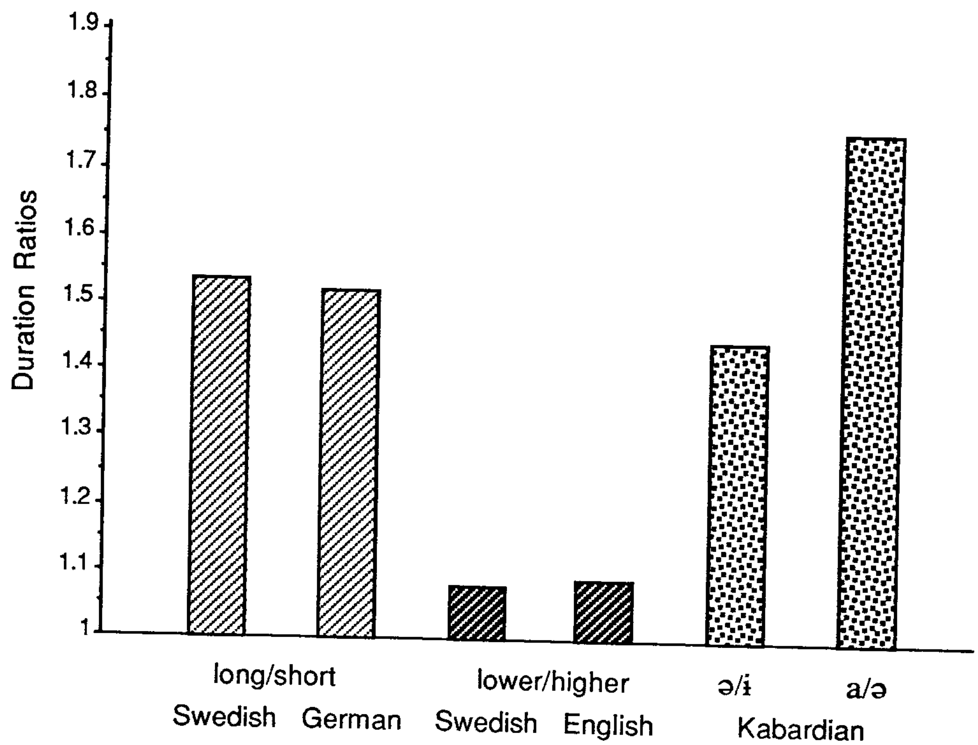
	prepausal	nonprepausal	difference
Speaker #1	133	97	36
Speaker #2	98	75	23
Speaker #3	107	70	37

Table 7. Kabardian mean duration ratios

The ratios are to be interpreted so that based on the overall mean values, the mid vowel is 1.45 times greater in length than the high vowel and the low vowel is 1.76 times greater in length than the mid vowel

Speaker	high	mid	low	(mid/high)	(low/mid)
1	63	97	184	1.54	1.90
2	58	75	127	1.30	1.70
3	54	81	133	1.50	1.64
Overall	58	84	148	1.45	1.76

Figure 4. Phonemic vs intrinsic duration



As evident from Figure 4, phonemic short-long ratios contrast sharply with those reported for intrinsic durational difference due to vowel height. The ratio for Swedish long/short vowels is 1.53, for German long/short vowels 1.52, Swedish [a]/[e] 1.08 and English [ɛ]/[ɪ] 1.10. By plotting the duration ratios for Kabardian [ə]/[ɨ] and [a]/[ə] on this histogram, it becomes evident that the differences in length between the vowels in Kabardian are more similar to the ratios reported for phonemic short-long distinctions. Note also that the low vowels are almost identical in length to the derived nonlow long vowels (Figure 3). It is not, however, possible to conclude that vowel length has phonemic status in Kabardian because this would involve recognizing that both [a] and [ɨ] are distinguished by length from [ə], each in opposite directions. This would be equivalent to concluding that the phonological contrasts in Kabardian vowels are best represented as /ɨ, ə·, a:/. Under this interpretation of the data, then, we are left with a three-way length distinction.

If we consider the length differences to be comparatively extreme examples of length variation due to vowel height, then an analysis is available that is based on the theory of redundant features proposed in Stevens, Keyser and Kawasaki (1986). The Kabardian vowel system differs from the vast majority of other three-vowel systems (e.g. Alabama, Mura, Gugu-Yalanji, Tagalog, *inter alia* (Maddieson 1984)) in that the contrast is maintained uniquely along a single dimension. The universal tendency toward triangular three-vowel systems with front-back contrasts as well as height distinctions is explained in terms of a general principle of maximal contrast (Jakobson 1941), more recently discussed and expanded by Lindblom (1986) into what is known as dispersion theory. If languages do indeed strive for maximal dispersion (hence maximal perceptual saliency) within the vowel space, how then does one account for a linear system of the type found in Kabardian?

A possible explanation is that Kabardian uses length redundantly to increase the perceptual salience of its linearly arranged vowels. This accords with Stevens, et al. (1986: 428) who argue that redundant features are "used to enhance the perceptual cues for distinctive features in a given language." In the case of Kabardian, the underlying height contrasts are strengthened by the introduction of length that co-occurs with the already present F1 distinctions. Such enhancement can be thought of as increasing the perceptual saliency, and therefore the functional efficiency, of a linear three-vowel system to the levels found in the maximally dispersed triangular systems. In this way, this view allows for the existence of linear three vowel systems in Lindblom's dispersion model.

This view also offers an explanation, albeit speculative, for what seems to be the three-way length opposition that exists phonetically. If length is implemented redundantly to maximize the perceptual salience of the height dimension, there are two possible strategies that a language could take. One is to increase and the other is to decrease the intrinsic duration of a vowel relative to other vowels. Kabardian may be a language that implements both strategies. The low vowel is lengthened and the high vowel is shortened, both in relation to the mid vowel. This would also explain the "ultra-short" (Kuipers 1960: 24) duration of the high vowel.

Finally, it should be noted that Catford (1984) reports mean duration measurements that differ significantly from the results of the present study, although they also indicate that duration is dependent on vowel height. Catford's data, which was based on a small corpus comprised of both a word list and a short text read by one speaker (Catford, personal communication), reveal mean duration values that make it even more evident that the length difference between the low and mid vowel is one of intrinsic duration rather than one of phonemic contrast: [ə] = 117 msec, [a] = 143 msec. This calculates to a low/mid vowel ratio of 1.22, a figure that is much lower than the 1.76 ratio reported in the present study. It is, at the same time, slightly higher than the sample intrinsic duration ratios presented for Swedish and German. Catford's data further supports the explanation given above for what seems to be a three way phonetic length opposition; i.e. the Kabardian vowel system strives for maximal contrast by increasing the intrinsic durational differences of its vowels. This explanation, if extended, predicts that vowel systems with this linear character should cross-linguistically exhibit higher durational differences between high, mid and low vowels than languages with nonlinear systems. It would be interesting to examine other languages with purported linear vowel systems to see if this hypothesis is empirically supported.

Phonological implications

Two conclusions have been made on the basis of the data collected from the spectrographic study:

1. There are three phonetic vowel heights.
2. There is a phonetic three-way length contrast which seem to constitute linguistically significant durational difference between the high, mid and low vowels.

These findings, upon immediate consideration, would seem to suggest a three-vowel analysis. If we adopt, moreover, the hypotheses discussed above regarding length, then we may conclude that the three-vowel analysis of the type proposed by Trubetzkoy (1925) is essentially correct.

There still remain the arguments for the two vowel analysis /ɨ, ə/ in which the low vowel [a:] is analyzed as the mid vowel /ə/ followed by the glide /h/ in a manner analogous to the derivation of the nonlow long vowels [i:, e:, u:, o:] from /ɨj, əj, ɨw, əw/, respectively. Recall that these arguments are based on the distribution of [aː] and the plural morpheme /hə/ which surfaces as [aː]. According to Catford (1984), Kuipers' arguments are founded on incomplete data. First, although it is true that [aː] is the only word initial vowel in Kabardian, Catford points out that literate Kabardians pronounce /ɨ/ and /ə/ in isolation when citing the corresponding letters in their cyrillic-based alphabet. In these cases, the vowels are also preceded by [ɦ], suggesting that murmured [ɦ] is "simply a Kabardian way of producing a 'soft attack' on a vowel" (Catford 1988: 5). Given this, Kuipers' first argument does not hold. As to the second argument, Catford states that Kuipers is simply mistaken about the plural morpheme. Kuipers' /hə/ is realised as /xə/ in most dialects of Kabardian. Moreover, although morphological tests show quite clearly that [iː, eː, uː, o] are in fact phonemically vowel+glide sequences, no such test is available for [aː]. There is diachronic evidence (Kumakhova and Kumakhov 1979: 82) that [aː] might have historically been conditioned by a following consonant, but this segment would have been a voiced uvular fricative, and not a laryngeal.

Given Catford's counterarguments, it seems that there is no phonological evidence for a two-vowel analysis. Moreover, the phonetic evidence also resists a two-vowel interpretation. Recall that the two-vowel analysis is contingent on deriving the low vowel from the mid vowel in the environment of /h/. If we assume, following Keating (1987), that /h/ lacks any supralaryngeal features with which to colour the vowels, there is no obvious reason for /h/ to trigger lowering. The spirit of the two-vowel analysis might be preserved if we invoked some version of underspecification theory and analyzed the distinction to be one of schwa quality associated to a single timing unit versus an unspecified sequence of two timing positions. However, this begs the question of why length should condition a feature [+low].

Conclusions

In summary, results of the current study of Kabardian vowels shows that there are three distinct phonetic categories along the F1 axis in the vowel space, supporting the claim that Kabardian is a three vowel system. This hypothesis is further supported by the patterning of vowel allophony; similar distribution of vowels occurs within each division of the F1 space, suggesting that the phonology recognizes three distinct vowels. Durationally, each vowel exhibited a length contrast with the other vowels such that the mid vowel was longer than the high vowel, and the low vowel was longer than the mid vowel. Two hypotheses are possible on the basis of this result. The first interprets the length differences in Kabardian in terms of a redundancy in which duration is utilized to attain maximal contrast in a linear vowel system. The second hypothesis goes beyond Kabardian and suggests that languages with linear vowel systems will exhibit greater intrinsic duration differences than languges with two dimensional vowel systems. The phonological interpretation of these results under either of these hypotheses suggest an analysis in which Kabardian has three underlying vowels /ɨ, ə, a/ which contrast solely in terms of their values for tongue body height. Length differences then serve to increase the perceptual salience of the underlying height contrasts.

Acknowledgements

I thank Peter Ladefoged, Ian Maddieson and the members of the UCLA phonetics lab for their helpful comments and suggestions. I would also like to thank Ian Catford for sharing his knowledge of Kabardian as well as for helping me find and record the native speakers who served as consultants in this study. Portions of this paper were presented at the 118th meeting of the Acoustical Society of America in St. Louis, Missouri. This research was supported by NSF Grant BNS 87-20098.

References

Catford, J. C. (1984). Instrumental data and linguistic phonetics. In J.W. Higgs and R. Thelwall (eds.), *Occasional Papers in Linguistics and Language Learning* 9. Coleraine: The New University of Ulster. 23-48.

Catford, J. C. (1988). Vowel systems of Caucasian languages. MS.

Colarusso, J. (1989). *The Phonology of the Northwest Caucasian Languages*. New York; Garland Publishing, Inc.

Colarusso, J. (1988). Kabardian. MS.

Elert, C. C. (1965). *Phonologic Studies of Quantity in Swedish*. Uppsala: Språkförlaget Skriptor.

Jakobson, R. (1941). Kindersprache, aphasie und allgemeine lautgesetze. *Språkvetenskapliga Sällskapets i Uppsala Förhandlingar 1940-1942*. Uppsala: Almqvist and Wiksell. (Reprinted in R. Jakobson. (1962). *Selected Writings I*. The Hague: Mouton. 328-401.)

Jakovlev, N. F. (1923). *Tablitsy Fonetiki Kabardinskogo Jazyka*. Moscow.

Keating, P. (1987). Underspecification in phonetics. *UCLA Working Papers in Phonetics* 67: 30-50.

Kuipers, A. H. (1960). *Phoneme and Morpheme in Kabardian*. The Hague: Mouton.

Kumakhova, Z. J. and M. A. Kumakhov. (1979). *Funktsionalnaja Stilistika Adygskix Jazykov*. Moscow: Nauka.

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

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala and J. Jaeger (eds.), *Experimental Phonology*. Orlando: Academic Press. 13-44.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Peterson, G. E. and I. Lehiste. (1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693-703.

Shagirov, A. K. (1967). Kabardinskij jazyk. In E. A. Bokarev, et al (eds.) *Jazyki Narodov SSSR, Vol. 4, Iberijsko-Kavkazskie Jazyki*. Moscow: Nauka. 165-183.

Stevens, K., S. J. Keyser and H. Kawasaki. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell and D. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc. 426-449.

Trubetzkoy, N. S. (1925). Review of Jakovlev 'Tablitsy fonetiki kabardinskogo jazyka', *Bulletin de la Société de Linguistique de Paris* 36. Paris. 277.

Zwirner, E. and K. Zwirner. (1936). *Grundfragen der Phonometrie*. Berlin: Metten & Co.

Shona Velarization: Complex Consonants or Complex Onsets?

Ian Maddieson

Paper presented at the Linguistic Society of America Annual Meeting
Washington, D.C. December 28-30 1989

One of the advantages of an articulator-based phonological feature theory is its ability to distinguish between combinations of articulations which can be produced simultaneously and those combinations which cannot be co-produced (Halle 1983, Clements 1985, Sagey 1986, Maddieson & Ladefoged 1989, Gorecka 1989). Basically, more than one articulation of the same degree of stricture can be produced only if distinct major articulators are used. For example, interdental and alveolar fricatives cannot be simultaneously produced, since this would require the tongue tip to be in two different places at once. Grouping articulations by the major articulator involved—in this case the Coronal articulator—provides a straightforward account of the combinatory constraints. This insight, as Ladefoged and Halle (1988) point out, is not captured in standard phonetic theory which simply provides a list of places of articulation.

At the same time as blocking impermissible articulatory combinations, articulator-based feature theory implicitly predicts the occurrence of those segment types which do use pairs of distinct articulators. However, some of the expected segments seem not to occur. In particular, Maddieson and Ladefoged (1989) claimed that segments with simultaneous dorsal and coronal closures are not attested in any language unless they are clicks, i.e. sounds using the velaric ingressive mechanism. From some accounts, it might seem that Shona provides evidence for the occurrence of pulmonic coronal-dorsals and so would invalidate that claim.

Most discussions of Shona phonetics continue to rely on studies by Clement Doke published in 1931 (Doke 1931a, 1931b).[1] Doke described a widespread process of velarization in Shona. From a diachronic perspective, this process may be summarized as in (1).

(1) Shona Velarization.

```
        σ   σ          σ
        /\  |         /|\
        CV1.V2   —>   CKV2
```

where C is a non-velar obstruent (including a nasal), V_1 is a high vowel (/i/ or /u/), and K represents a velar obstruent with voicing determined by the voicing of C.

Due to other historical processes and to particular morphological patterns there are relatively few cases where the V_1 generating velarization is a front vowel. If V_1 is rounded and C is coronal (or sometimes a bilabial nasal), then the velar element retains rounding. Some illustrative examples are given in table 1. The sequences affected by these changes include many in non-alternating stems (as in 1A) but there are also productive alternations, particularly in the formation of passive verbs (1B), and when noun class prefixes /mu-/ and /tu-/ occur with nouns whose stems begin with a nonhigh vowel (1C). Similar changes occur with noun concord elements for the classes with these prefixes. The details vary somewhat between the dialects; we will focus on Zezuru, which is the most widely spoken and is increasingly serving as a standard (Mkanganwi 1975).

==

Table 1. Velarization examples

A. within stems e.g.

orthography	Doke	gloss	
imbwa	imbgá	"dog"	
uzutwe	uzútkwe	"(mushroom sp.)"	
kudya	kúdʒgá	"to eat"	(< *ku-di-a)

B. formation of passive; verbs with a stem-final true consonant

active verb (orthography)	gloss	passive (Doke)	gloss
-ɓata	"to hold"	-ɓatkwa	"to be held"
-tanda	"to chase"	-tandga	"to be chased"
-tasva [2]	"to ride"	-taʂkwa	"to be ridden"

cf. verbs with no stem-final vowel or with stem-final glide, e.g. ɓa "to steal", ɓiwa "to be stolen", ɗa "to love" -ɗiwa "to be loved", -gaya "to grind" -gayiwa "to be ground".

C. noun class prefixes /mu-/ (cl. 1, cl. 3, cl. 18), /tu-/ (cl 12, dim. pl.) with nouns whose stems begin with a nonhigh vowel, e.g.

		orthography	Doke	gloss
cl 1	mu+ana	*mwana*	mŋana	"child"
cl. 3	mu+edzi	*mwedzi*	mŋedzi	"moon"
cl. 12	tu+ana	*twana*	tkwana	"little children"

cf. the syllabic form of these prefixes in mu-royi "witch" cl. 1, mu-ti "tree, bush, herb" cl. 3, tu-futa "a little fat" cl. 12, etc.

==

In describing velarization of Zezuru consonants, Doke uses digraphs, trigraphs and tetragraphs. For velarized bilabials these include **pk**, **bg**, **mŋ**, etc. and for velarized coronals, **tkw**, **tskw**, **dʒg**, etc. It is very clear that Doke regarded each of these functionally as single consonants and not as consonant clusters, despite these multi-letter combinations for their representation. Each is separately entered on the chart of Zezuru consonants (Doke 1931b: 231),

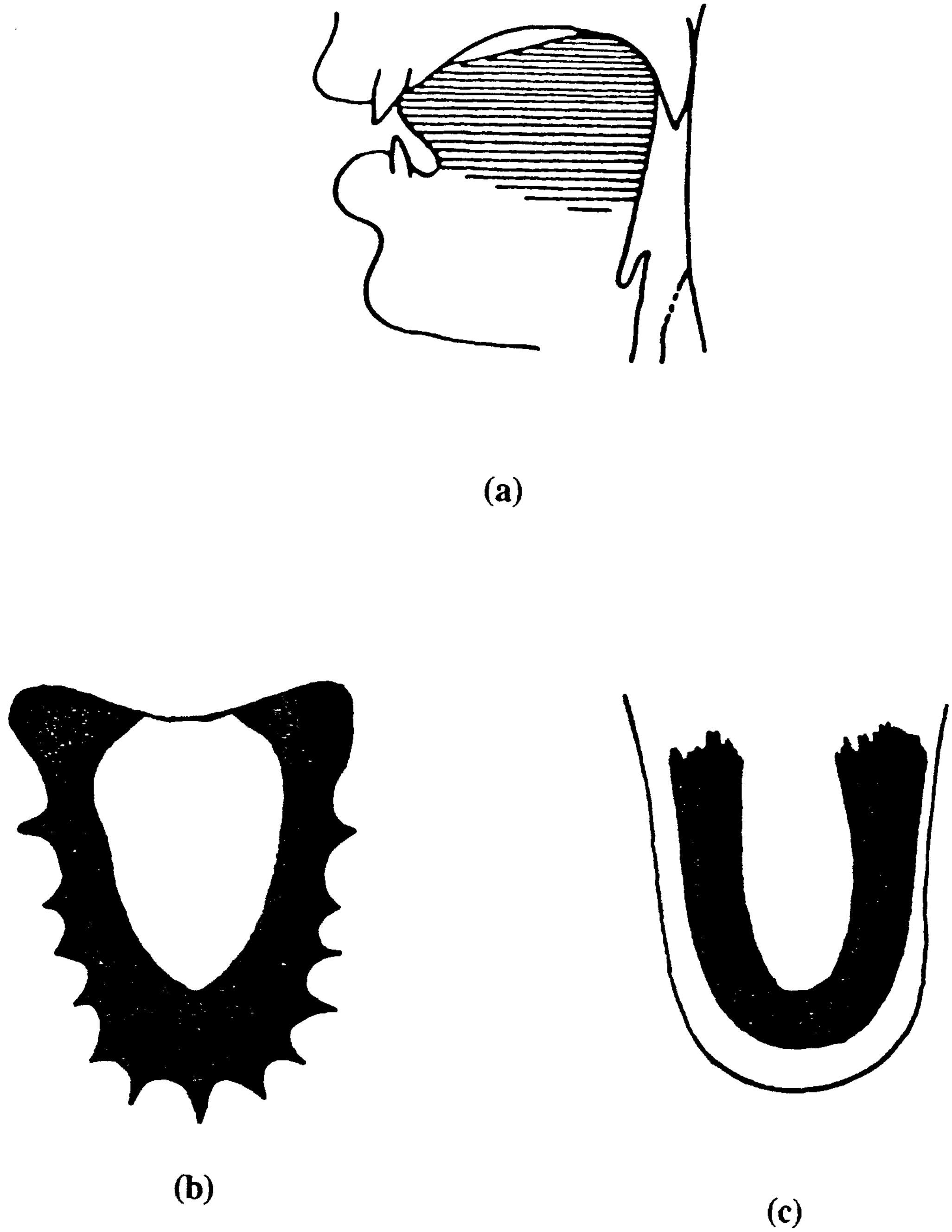


Figure 1. (a) Diagram of the articulatory position of **tkw** in Zezuru, following Doke (1931b). The position shown is inferred *inter alia* from the the contact area shown shaded on the palatogram (b), and the linguagram (c) showing the contact area on the front of the (protruded) tongue. Note that the false palate used does not extend back to the velar region, but the lateral contact toward the back of the palatal region implies the presence of central contact further back.

and Shona consonants are described as consisting of a set of plain consonants and a set of velarized consonants, in a way that is similar to the division between plain consonants and click consonants in Zulu (p. 24). The symbol representing the velar element always follows that representing the labial or coronal element, but Doke is not explicit about the relation in time of the velar and nonvelar articulatory gestures involved. At some points he seems to indicate that the articulations are sequential, as for example in describing velarization as an "abnormal raising of the back of the tongue" which "may take the place of the semi-vowel" (i.e. /w/, which follows another consonant) (p. 109), or when discussing the appearance of velar stop bursts in kymograph records (e.g. p. 113). Elsewhere there are indications that he considered them to be simultaneous, or at least substantially overlapping. For example, he compares Shona velarized consonants with the doubly-articulated labial-velar stops of Yoruba (p. 49). He also provides diagrams showing articulatory positions of a set of Zezuru consonants (p. 269-273). These are deduced from the contact areas shown in palatograms of a single speaker, and, presumably, from his general impressions of the mode of production observed during his fieldwork. The diagram representing production of **tkw** is reproduced as figure 1, together with the palatogram and linguagram on which it is based. It shows the tongue making two complete closures in the oral cavity, one a broad alveolar contact, the other velar. Static palatograms do not distinguish a sequence of movements, but only show the total area of contact, so it is possible that this diagram was intended to represent a kind of sum of successive articulations. However, this seems unlikely. In the set of diagrams representing articulatory positions in the Karanga dialect, the one representing [nʝ] is footnoted "The tongue-position diagram really shows successive movements, **ʝ** succeeding **n**." No comparable note is attached to the Zezuru **tkw** diagram. Similar double contact positions are shown for Zezuru velarized alveolar nasal, and for voiceless and voiced velarized palato-alveolar affricates (**nŋ**, **tʃk**, **dʒg**). These articulatory diagrams are reproduced, together with the palatograms and linguagrams on which they are based, as figure 2. Again, no footnote indicates that these represent successive movements. Most probably, Doke intended to indicate that in his opinion these velarized consonants do have two simultaneous closures. If his observation is correct, then pulmonic coronal-dorsals do occur.

In order to investigate this issue, data was collected from three speakers of the Zezuru dialect of Shona at the University of Zimbabwe. High quality audio recordings of a substantial number of words and phrases were made. In addition, the subjects provided intraoral air pressure data on a smaller list of words, which were also tape-recorded. Intraoral pressure was recorded through a thin flexible tube in the mouth. This tube was positioned between the cheek and the upper molar teeth and connected to a short length of aluminum tubing held in position by a clip fastened over the second molar, on the left or right according to subject preference. The open end of the aluminum tube is led inside the enclosure formed by the teeth and angled upward toward the roof of the palatal arch.[3] Air pressure is thus sensed in a palatal area behind any labial, alveolar or palato-alveolar closure and in front of a velar closure. Thus, on Doke's figure of the **tkw**

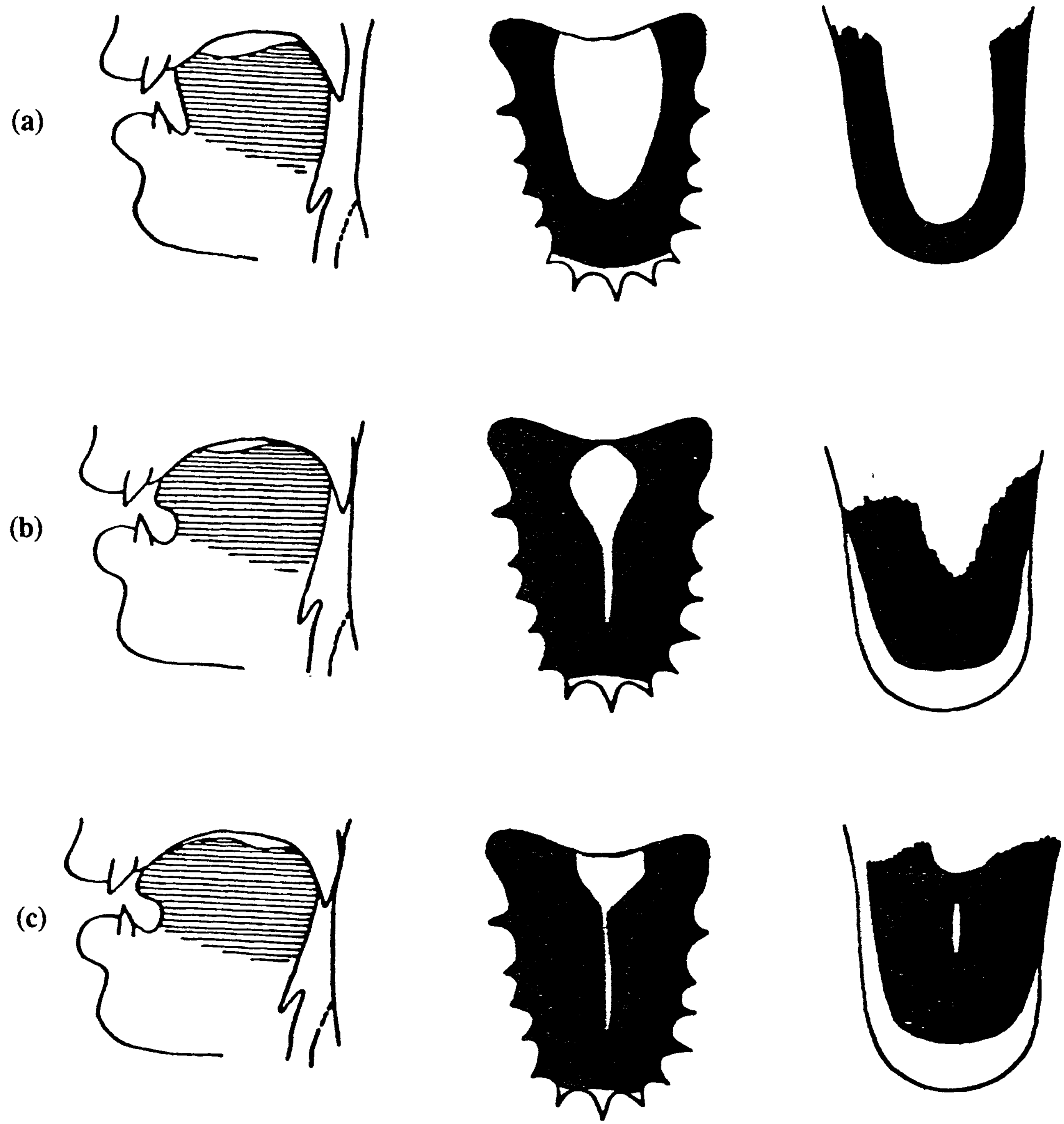


Figure 2. Presumed articulatory positions, palatograms and linguagrams of Zezuru (a) **nŋw**, (b) **tʃk** and (c) **dʒg**, following Doke (1931b).

articulation, the open end of the tube would be positioned in the space between the two closures shown.

With air-pressure sensed in this location we can expect to see the following patterns. During a vowel or other sound with no oral obstruction, the air pressure will be equal to atmospheric pressure and the pressure trace will be flat. During a pulmonic alveolar or labial stop, air from the lungs will be impounded behind the oral closure until the closure is released. This produces a characteristic air-pressure curve with a relatively steep initial rise, usually followed by a second phase with a slower rise if the closure is prolonged, and a sharp pressure drop on release (Müller & Brown 1980). However, during a velar stop we will observe no increase in pressure in the location being monitored, since the closure is formed behind this location. The pressure trace will remain flat. For coronal or labial fricatives an oral pressure increase will be detected due to the airflow resistance caused by the constriction, but the shape of the curve is more nearly symmetrical in its rise and decline than that seen for a stop. Coronal and labial nasals show a small increase in oral pressure with generally a plateau pattern.

The intraoral pressure data from two speakers and the audio data from all three speakers has been analyzed. (Pressure data from the third speaker is erratic since apparently the tongue contacted the open end of the tube.) Both data types provide information on the velarized consonants but the relationship in time of the two components can best be examined by viewing synchronized acoustic and pressure data from the simultaneous recordings. Figure 3 is a display of this kind from two male speakers. The speech waveform is on the top and the pressure record on the bottom.[4] This figure is of the passive verb form /ra-ka-taşkwa/ "it (cl. 5) was ridden". The pressure trace remains flat during the first two syllables for speaker 1. Neither the approximant /r/ nor the velar stop /k/ result in any increase of oral pressure where it is being recorded. Speaker 2 differs in having a tapped /r/ initially, which produces a brief pressure increase. For both speakers, pressure rises as the closure for the alveolar stop /t/ beginning the third syllable is formed. It then sharply decreases at the release. Speaker 2 has a shorter closure and the pressure rise is not as readily divided into two parts as it is for speaker 1. The considerable amplitude of the /t/ burst and duration of aspiration can also be observed in these tokens. It should be noted that this /t/ initiates a stressed syllable, since Shona has the common Bantu penultimate stress rule. These syllables provide data on timing of acoustic events and pressure curves for consonants with no velarization.

In velarized alveolar stops, as in the phrase /tkwana tkwaŋgu/ "my little children" (< tu+ana tu+aŋgu) in figure 4, the pressure rises for the alveolar closures, and the time course and magnitude are very similar to that of the intervocalic /t/ by itself. If a velar closure co-occurred with the alveolar closure the pressure rise would have been limited by the velar constriction. The pressure falls at the time of alveolar release and a substantial release burst can be seen (and heard) followed by a period of aspiration, indicating that the air impounded in the mouth is directly linked

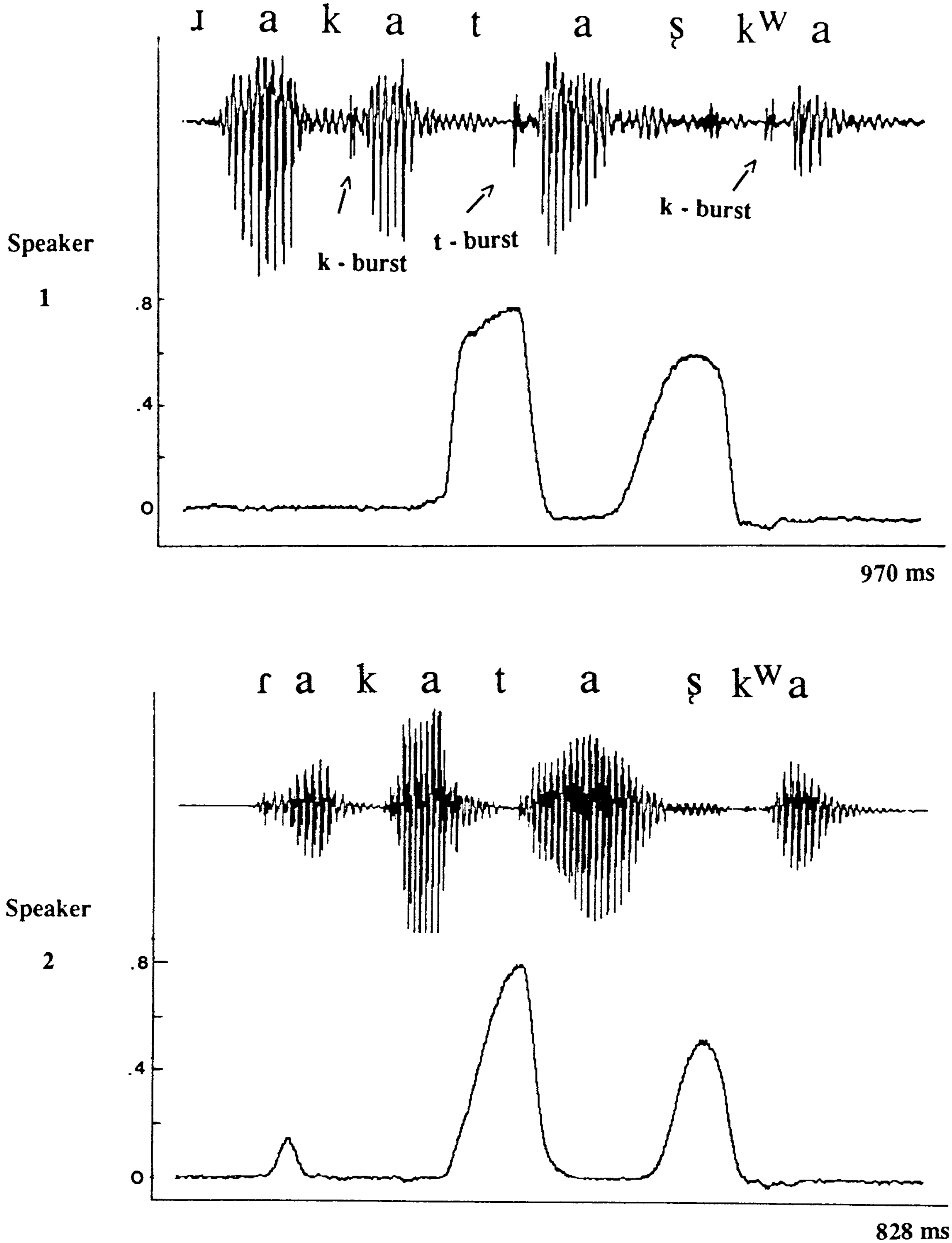


Figure 3. Waveform and intraoral pressure contour of the utterance [rakataşkwa] "it (cl. 5) was ridden" spoken by two male speakers. Relative pressure levels on figures 3-4, 6-7 are indicated on an arbitrary scale in which atmospheric pressure = 0 and the maximum pressure recorded during the entire session for a given speaker = 1.

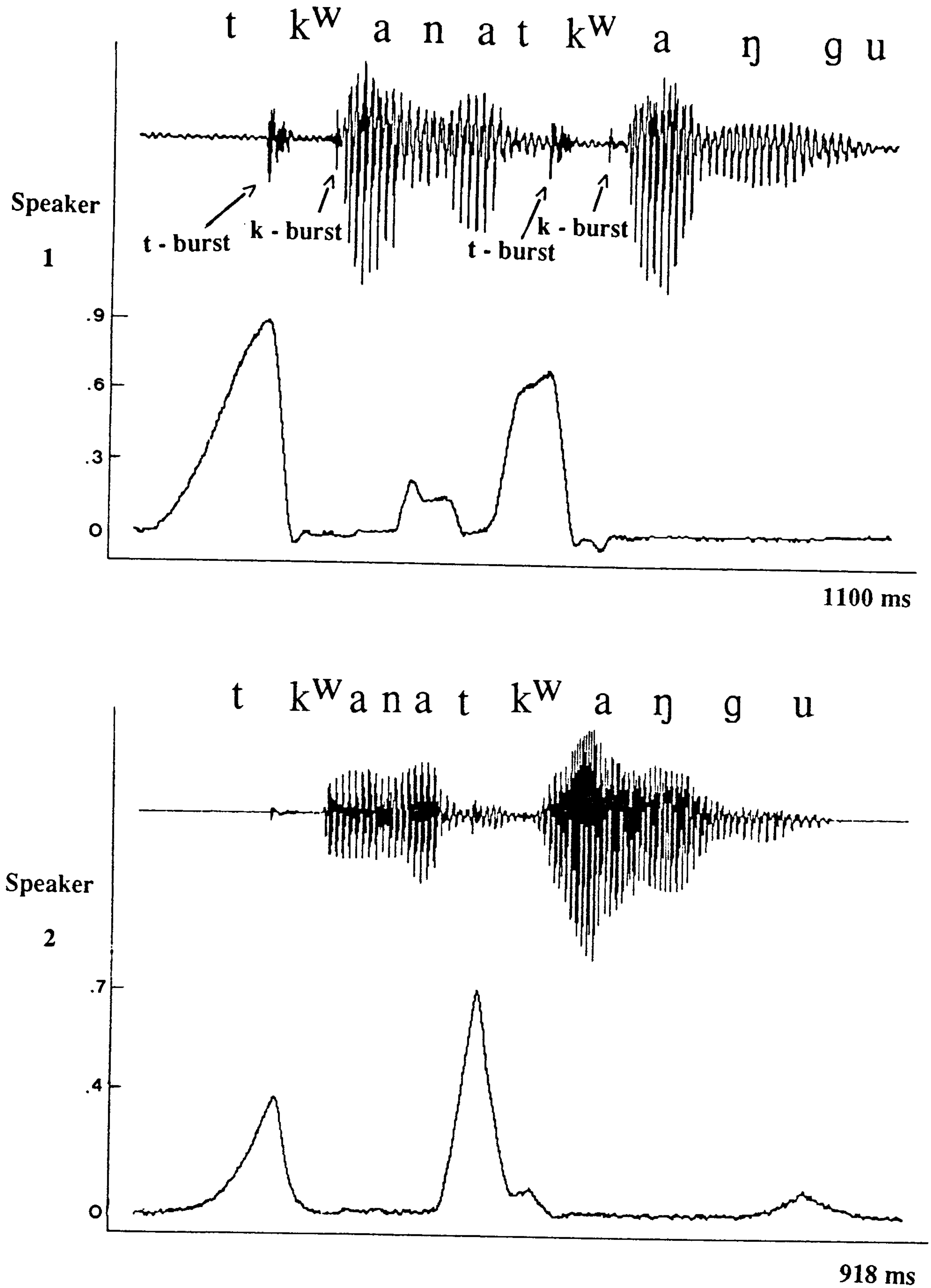


Figure 4. Waveform and intraoral pressure contour of the utterance [tkwana tkwaŋgu] "my little children" spoken by two male speakers.

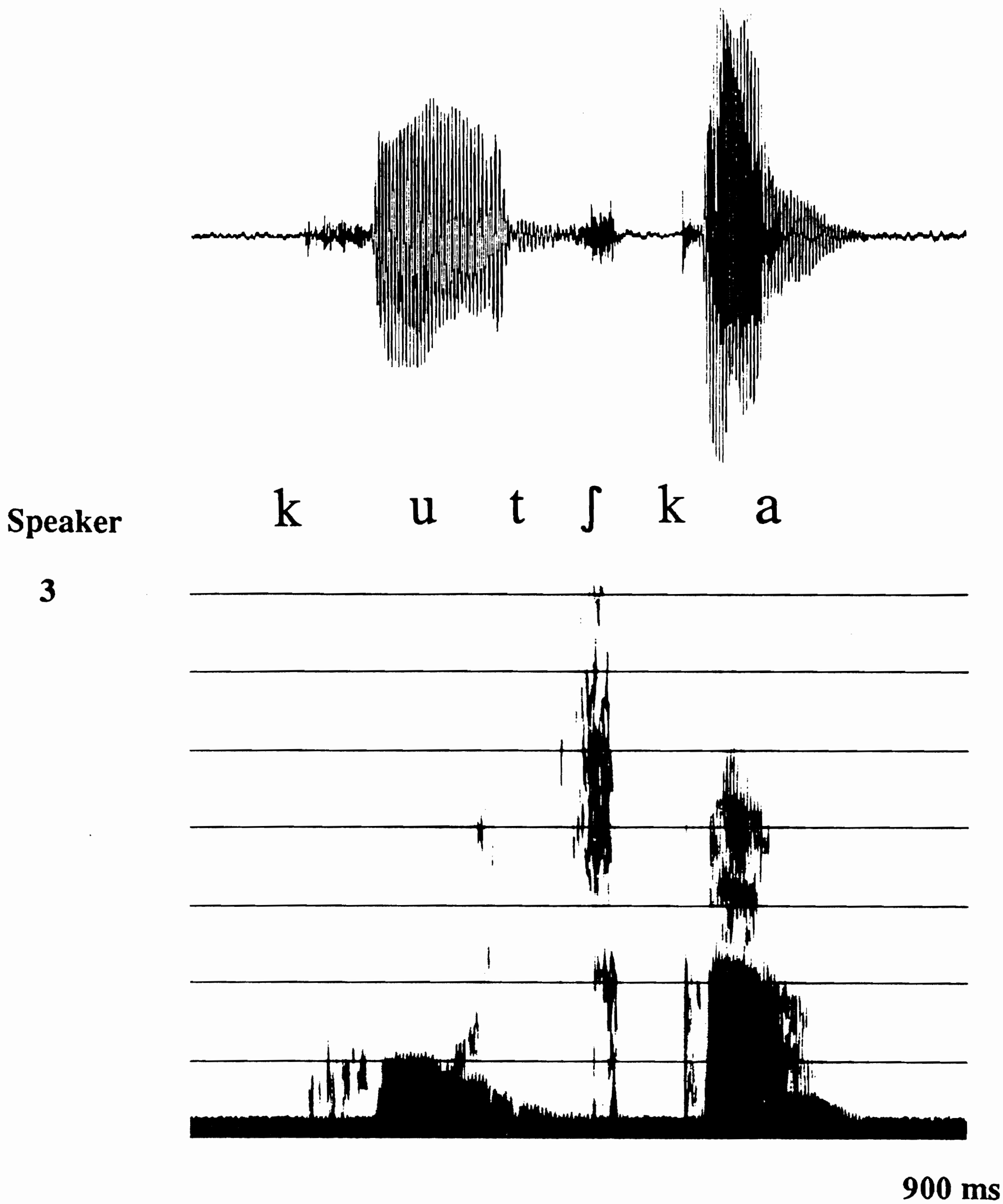


Figure 5. Waveform and wide-band spectrogram of [kutʃka] "to fear".

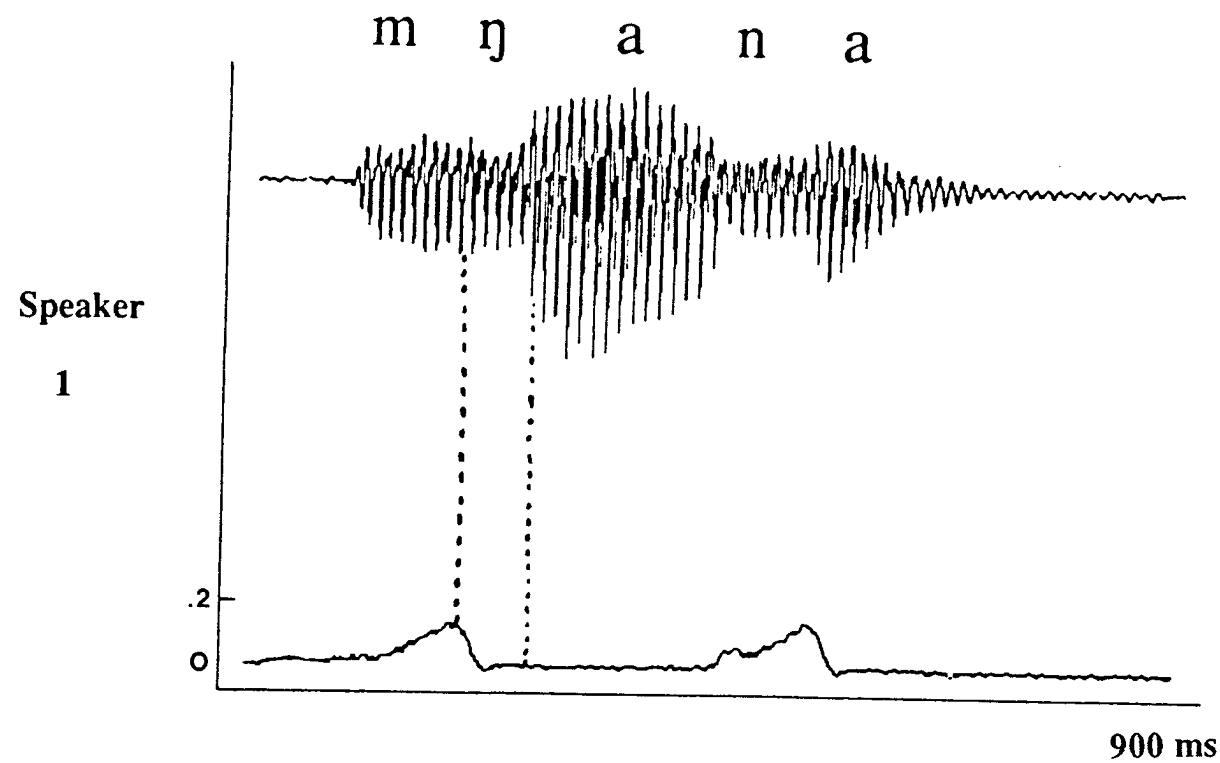


Figure 6. Sequential production of [mŋ]. Dotted lines indicate [ŋ] segment. Note relatively sharp decline in pressure on release of [m], similar to pattern seen in [n] segment later in the word.

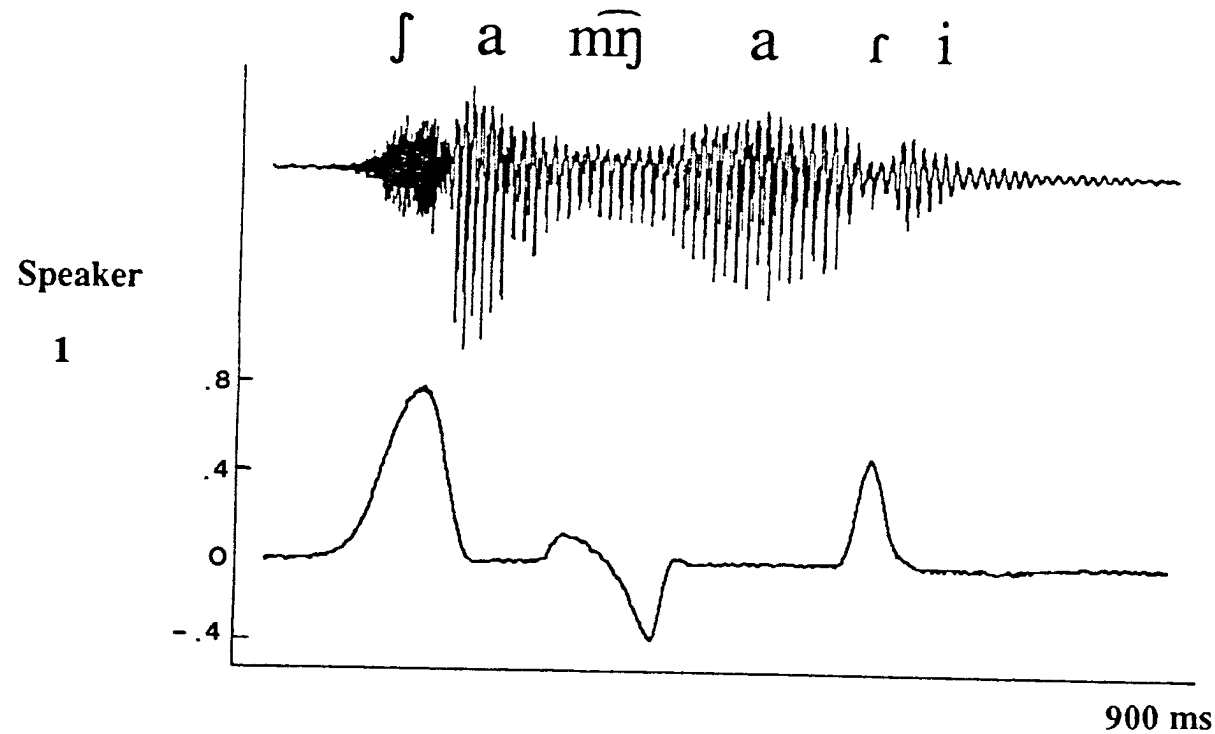


Figure 7. Overlapping production of [m͡ŋ]. Note slow initial decline of pressure and negative peak before release. Negative pressure in the cavity between the labial and velar closures results from lowering of the tongue body; before contact with the roof of the mouth is broken the size of the cavity increases and the air it contains is rarified.

to the air under pressure in the subglottal cavity that is the source of the power for aspiration. The characteristic sharp decline from the pressure peak and the strength of the release indicates that there is no overlapping velar closure in the mouth at this time. We conclude that the velar closure is only formed following this release. It receives its own release after being held for some 70-80 ms. Phonetically it is clear that we have a cluster of two stops, each with its own release, and they do not overlap in time at all. From close spectral and auditory examination it is also apparent that only the velar release is labialized.

The timing pattern of velarized fricatives is shown in figure 3, in the final syllable of the passive verb form /ra-ka-taʂkwa/. The fricative is not modified in any way by the overlaying of a velar articulation; compare the pressure trace for a non-velarized /ʃ/ in figure 7. Instead it is completed and then a velar closure is made and released. The same pattern occurs with affricates. They are produced as usual and then followed by a velar stop. This is illustrated by the spectrogram and waveform of the word *kutya* [kutʃka] "to fear" from speaker 3 (who is female) in figure 5, in which the acoustic pattern of closure-release-frication for the affricate is followed by closure and release for the velar stop.

Across the whole data set we also observe sequential articulations when the first element is a labial rather than a coronal, except for a pattern of optional variation in the case of bilabial nasals which are velarized. Here nasality spreads from the initial nasal to the velar element, and one may hear a sequence [mŋ] as in the token of [mŋana] (< mu, cl 1 + -ana "child") in figure 6. A shift from labial to velar occlusion can be seen in the waveform and pressure trace, and can be clearly observed in the spectrogram made of this utterance. However, productions in which the velar substantially overlaps with the bilabial closure may optionally occur and are particularly likely in word-medial positions. In this case, the air between the closures may even be rarefied somewhat as the tongue is lowered from the velar contact. A token of this type from the same speaker is shown in figure 7. Nasals of this type are somewhat longer (about 50 ms) than plain underlying nasals, even though their articulations are overlapped.

Some differences between speakers can be observed in the data. Speaker 3 has the longest velar closures and a tendency to shorter aspiration duration following the velar release. Speaker 2 has a greater tendency to produce stops with incomplete closure than the other two speakers, even single intervocalic stops. Speaker 1 shows the greatest tendency to retain lip-rounding in [mŋ] sequences from underlying /mu-/, that is, to pronounce [mŋʷ]. But the patterns exemplified in the figures shown are representative of the principal patterns found for all three speakers.

The phonetic facts then are clear. Shona does not have simultaneous coronal-dorsals; Doke's diagrams showing this are misleading. The velar element of velarized consonants is not co-produced with the coronal or labial element but follows without overlap. The one exception observed concerns velarized bilabial nasals, which may optionally involve overlapping labial and dorsal articulations. Of course, pulmonic labial-dorsal segments are well-known in other languages.

Given that Shona velarized consonants are phonetic sequences, are they nonetheless single segments phonologically, rather than obstruent sequences in the syllable onset? Several arguments have been marshaled in favor of saying that they are. One rests on syllable structure. In Shona, consonant sequences in loanwords are usually simplified or broken up by insertion of a vowel, even in the case of obstruent-liquid clusters. Examples are given in table 2. Universal syllable preference laws lead us to expect obstruent-liquid clusters to occur in a language when clusters of any less common type are found. Since on the loanword evidence, Shona disallows obstruent-liquid sequences, it can be argued that Shona velarized consonants must be analyzed as single segments, not as obstruent clusters. Sagey (1986) makes this point for Shona and also for similar phenomena in Kinyarwanda (Jouannet 1983).

Table 2. Shona Loanwords (examples from Doke 1931b, Fortune 1955, Biehler 1927)

loanword	gloss	source
vowel epenthesis		
ma-girazi	"eyeglasses"	< Eng. glasses
mu-puraŋga	"gum-tree"	< Port. prancha
basekoro	"bicycle"	< Eng. bicycle
burukwa	"trousers"	< Afrikaans broek
simplification		
mu-domeni	"agricultural extension worker"	< Eng. demonstrator
ci-ŋgezi	"English"	< Port. inglez
ci-tim(ir)a	"train"	< Eng. steamer
ci-kerema	"scoundrel"	< Afrikaans skelm

Despite the non-alternating cases (table 1 A), it seems likely that all the velarized consonants in Shona are to be derived from underlying consonant+high vowel or glide sequences. A Glide Formation rule turns high vowels into glides before another vowel, subject to various conditions that are not important for this discussion. In the case of the sequence /kuV/, e.g. concords of the cl. 17 locative noun prefix /ku-/ in a phrase like /ku-muʃa ku-aŋgu/ --> [kumuʃa kwaŋgu] "at my village", no further process needs to be posited. Glides following coronal or labial true consonants undergo the strengthening process that results in the velar obstruents found on the surface (O'Neil, 1935, and Fortune, 1955, state the process as just such a replacement of a glide with an obstruent). If the velarized consonants are in fact surface single complex segments they must also undergo a process that can be termed Incorporation, following Herbert's (1986) usage concerning prenasalized stops derived from underlying nasal+stop sequences. Incorporation of two segments into one is often motivated, at least in part, by the occurrence of "shared substructure" within the derived complex segment—essentially Obligatory Contour Principle effects. For example, the nasal and stop portions of prenasalized stops are homorganic, and this is reflected by the presence of a single place node in the tree which represents their structure. In some cases, the status of such elements as single segments can be supported by evidence that the timing slot vacated by the incorporation process is reassigned to a neighbouring segment. This is so in Luganda, where the vowel preceding a prenasalized stop is lengthened (Ashton et al, 1954). In Shona, neither place nor stricture features are shared between the original underlying initial consonant and the velar obstruent derived from the glide, which is a velar stop whether the initial consonant is a stop or a continuant, labial or coronal. As Shona is not a quantity language, it it not surprising that there is no compensatory lengthening of a vowel preceding a velarized consonant, but this does mean that no support for an incorporation process can be found from this line of argument.[5] The shared voicing specification of the parts of a velarized consonant will be discussed further below.

Although the degree of stricture of the initial consonant does not govern the output of the glide strengthening process, Sagey (1986) argues that this fact is no bar to single segment status for velarized consonants. In Sagey's account the velarized consonants are segments with two place features, with the non-dorsal component designated as the primary articulation by a "pointer". The structure of /tkw/, with the pointer singling out the coronal element, is shown in the simplified diagram in figure 8. The dorsal component is thus formally relegated to the role of a secondary articulation, as is the [labial] element which indicates that rounding is present in this structure. In Sagey's model, complex consonants of this type are blocked from having separate stricture specifications for the places involved. Stricture features apply to the primary place only—here indicating that the coronal element is a stop—and so the secondary articulation has no *phonologically relevant* stricture features. In this sense there is a shared substructure in the feature tree of the velarized consonants. Sagey argues that this property accounts for degree of stricture variation in the velar element, which in some Shona dialects (and optionally under certain conditions even in Zezuru) may involve velar frication or approximation rather than full occlusion. It also accounts for the lack of phonological contrast between [tw] and [tkw]. The fact that the dorsal element is realized as a velar closure, but the labial element is only an approximant is treated as a matter of language-specific phonetic implementation.

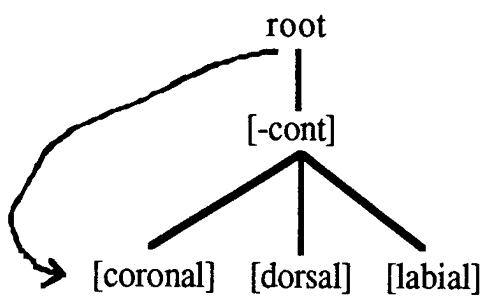


Figure 8. Simplified representation of /tkw/, after Sagey (1986).

The appropriateness of this account might be questioned on two grounds. Shona has a distinction between plain and labialized coronal fricatives (the latter transcribed here as [ş ẓ], orthographic *sv, zv*). Presumably, the labialized ones would be represented as in figure 9. Both the plain and labialized coronal fricatives can be velarized, but they remain distinct, e.g. the voiceless pair as [skw] and [şkw]. If both these velarized forms are represented as unordered coronal, dorsal and labial articulatory gestures within single segments, there is no way to represent the fact that in one case labialization only applies to the dorsal articulation, whereas in the other the coronal portion is also labialized. Two timing slots are required to represent the correct extent of labialization in these two different cases, and distinct phonetic material must be attached to them.

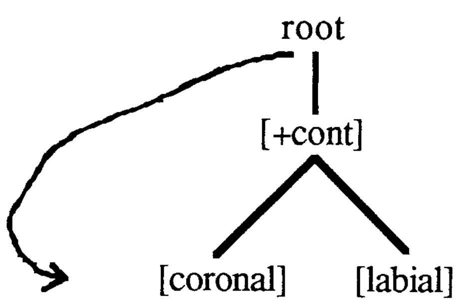


Figure 9. Simplified representation of labialized coronal fricative (after Sagey 1986).

The second issue is a more general one concerning the appropriate scope of rules of phonetic interpretation. There is a growing body of work on the question of relating phonological structures to measurable phonetic parameters, particularly in areas such as duration, pitch and nasality (e.g. Pierrehumbert & Beckman 1988, Cohn 1990). There is relatively little work on directly relating the basic features of consonant place and manner to phonetic output, apart from Browman and Goldstein's articulatory phonology model (Browman & Goldstein 1986, 1989), although many studies on phonetic aspects of coarticulation assume or imply some more general model of phonetic implementation (cf. Keating 1990). The common assumption across all this work is that phonetic implementation is concerned with specifying (a range of) real values of phonetic parameters for some terminal output of the phonology and specifying a path of interpolation to link these values together. Although this process includes providing details of timing, especially details of the intra- and inter-segmental timing relationships of various articulatory subsystems, the basic order of the elements is given by the phonological representation and represents part of the phonological knowledge of the language that a fluent speaker has. Realization of a putative complex segment as a sequence of temporally disjunct movements of distinct articulators goes beyond this notion of phonetic implementation. In Shona, the consonant-glide sequence that is the input to the velarization process is ordered, but the ordered relationship of the velar and nonvelar components of the resulting velarized consonants is erased if an incorporation rule is posited, rather than simply a glide strengthening rule. To recreate ordering as a matter of phonetic implementation is not only uneconomical but gives power to these rules that is like that of phonological rules, such as rules of epenthesis and metathesis.

To suggest that velarized consonants are segment sequences does not mean that there aren't syllable structure constraints in Shona, whose effects are *inter alia* reflected in the reshaping of loanwords. But these apply to structures before the velarization process has operated. As Myers writes "Syllables in Shona are very simple. They are all open, there are no long vowels or diphthongs, and the onsets consist either of a single consonant, or a consonant and a glide" (Myers 1987: 221). Consonant-liquid sequences are barred and hence are restructured in loans, but velarized consonants are underlying or derived consonant-glide sequences, the glides in many cases being transparently derived from underlying high vowels, especially /u/ —> /w/. The velarization process changes the glides into velar obstruents when they follow non-velar obstruents, but the sequential ordering and timing as a two-segment sequence is preserved.[6]

Shona is thus like many other languages in creating surface syllables that are disallowed at a level of "core syllabification" (Clements 1990)—compare Russian examples such as [rta, mxa] from lexical /rot+a mox+a/. The Shona example involves onset clusters of obstruents which, also as in Russian, must be all voiced or all voiceless. In Shona, voicing for the cluster is determined by the voicing of the initial consonant, the underlying obstruent. This can be predicted from universal considerations without any need to assume that it follows from single segment status. Since they do not contrast, it is reasonable to assume that glides and vowels are underlyingly unspecified for voicing. Universal default rules supply [- voice] to an obstruent unspecied for voicing and [+ voice] to a vowel or glide remaining as such on the surface. Universal sequential constraints block a tautosyllabic [- voice] consonant from occurring closer to the syllable nucleus than a [+ voice] consonant, so [+ voice] must spread to a following obstruent derived from a glide when the syllable-initial obstruent is underlyingly voiced. These universals predict the Shona voicing facts.

To the extent that stricture variation in the velar element of Shona obstruent sequences occurs, it can be seen as a repair strategy for the disfavored onset sequence created by velarization.

Parallel variation can be observed in languages which have derived disfavored consonant sequences by deleting an intervening vowel, such as in the Eggon examples in table 3 where a stop in C_2 position in the onset can be relaxed to a (lenis) fricative. Note that singleton stops in Zezuru are also subject to optional stricture variation, and laxing of both singleton stops and velar stops derived by velarization increases with speech rate. Such variation can be appropriately dealt with by phonetic implementation rules.

(4) Eggon obstrunet laxing in clusters

proto-form		Eggon variants	gloss
*o-bogo	—>	o.bgo —> obɣo	"arm"
*o-kopo	—>	o.pko —> okϕo	"ten"
*a-k͡piki	—>	a.k͡pʲki —> a.k͡pʲxi	"body"

Conclusion

In short, we conclude that Shona velarization does not result in creation of complex single consonants, and in particular does not generate either phonetic or phonological coronal-velar obstruents. Pulmonic coronal-velars remain unattested. We suggest that this may not be an accidental gap. Although the tongue dorsum and tongue blade are correctly regarded as independent primary articulators, there is a physical linkage between them — both being parts of the tongue. Perhaps this results in constraints on their simultaneous involvement in segments of a more restrictive nature from those applying between the lips and a part of the tongue. It seems probable that tongue articulators may only combine with each other as primary articulators in obstruent segments in which the tongue dorsum also acts as the initiator of the airstream, i.e. in clicks. This may be because the downward movement of the center of the tongue allows for greater independent movement of the two parts and generates stronger acoustic cues to the presence of two articulations.

Acknowledgments

This work was supported in part by NSF grant BNS 87-20098. The research reported here would not have been possible without the assistance of many people, in particular, Professors A.J.C. Pongweni, B. Annan and M.K. Mkanganwi, and Baxton Chikore, Trevor Ngorosha, Wedina Chindawe and Anne Jefferies at the University of Zimbabwe, Professor A. Traill of the University of the Witwatersrand, Professor P. Wenzel of the University of South Africa and the Board of the African Languages Association of Southern Africa.

Footnotes

1. Very little work on Central Shona phonetics has appeared since Doke's, although Pongweni (1971, 1981, 1983) has described various aspects of the phonetics of the Western Shona Karanga dialect in some detail. Karanga velarization is substantially different from Zezuru.

2. Orthographic 'sv' and 'zv' represent the so-called "whistling fricatives" of Shona, represented in phonetic transcription by [ʂ, ʐ]. These are alveolar fricatives with simultaneous approximation of the lips. This rounding differs from that of /w/ in having less horizontal approximation of the lip corners and a smaller degree of protrusion of the lips. Orthographic 'sw, zw' in Zezuru, of course, represent [skw, zgw].

3. The intraoral pressure was recorded using equipment designed at the University of the Witwatersrand and loaned by Professor A. Traill. The intraoral tube is led to a Sensym pressure transducer whose output is fed to a voltage-to-frequency converter. This converts the varying voltage output of the pressure transducer into a signal with varying frequency (actually a pulse train with a basic frequency of about 850 Hz). Increases above atmospheric pressure are indicated by an increase in the pitch of this base frequency and its harmonics, pressure decreases are indicated by a frequency decrease. This signal is recorded on one track of a dual-track tape recorder while a simultaneous audio recording of the speech is recorded on the other. The pressure system was calibrated with a U-tube manometer connected directly to the transducer input tube and the response shown to be linear; however, due to equipment limitations and lack of a controlled temperature environment, absolute pressure values cannot be calculated.

4. Figures 3, 4, 6 and 7 were made by simultaneously digitizing the speech and pressure signals at 40 Khz. The pressure signal was band-pass filtered to eliminate AC power noise and harmonics in the signal, and a peak-picking program was run on the resulting waveform. The output (the F_0 contour of the pressure signal) was smoothed using a 198 point moving window (approx 5 ms in length) to reduce the quantization effects introduced by the digitization. Only selected tokens were processed in this way but the entire set of pressure data was examined using the dual-channel capability of the Kay 5500 Digital Speech Processing workstation to display simultaneous spectrograms and pressure curves.

5. Incorporation would be well-motivated in the case of the overlapping production of [m͡ŋ] to distinguish this case from those where no co-production occurs.

6. It seems to follow that Shona would borrow English words such as "twin" or "dwelling" without simplifying the onset cluster (presumably as something like /mu-tkwini/, etc), but no relevant examples of assimilated loanwords are to be found in the sources consulted.

References

Ashton, E. O., E. M. K. Mulira, E. G.M. Ndawula & A. N. Tucker. *A Luganda Grammar.* Longmans, Green, London.

Biehler, E. 1927, reprinted with a preface by G. Fortune, 1950. *A Shona Dictionary with an Outline Shona Grammar.* The Jesuit Fathers, Chishawasha, & Longmans, Green, Cape Town.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6: 201-252.

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

Clements, G.N. 1990. The role of the sonority hierarchy in core syllabification. In J. Kingston & M. E. Beckman (eds.) *First Conference on Laboratory Phonology.* Cambridge University Press, Cambridge.

Cohn, Abigail. 1990. *Phonetic and Phonological Rules of Nasalization.* Ph. D. dissertation, University of California, Los Angeles.

Doke, Clement M. 1931a. *Report on the Unification of the Shona Dialects.* Stephen Austin & Sons, Hertford, for the Government of Southern Rhodesia.

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

Fortune, G. 1955. *An Analytical Grammar of Shona.* Longmans, Green, London.

Gorecka, Alicija. 1989. *The Phonology of Articulation.* Ph. D. dissertation. M.I.T., Cambridge, Mass.

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

Herbert, Robert K. 1986. *Language Universals, Markedness Theory and Natural Phonetic Processes.* Mouton de Gruyter, Berlin.

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

Keating, Particia A. 1990. The window model of coarticulation: articulatory evidence. In J. Kingston & M. E. Beckman (eds.) *First Conference on Laboratory Phonology.* Cambridge University Press, Cambridge.

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

Maddieson, Ian & Peter Ladefoged. 1989. Multiply-articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics* 72: 116-138.

Mkanganwi, K. G. 1975. A description of Shona spelling. *African Languages/ Langues Africaines* 1: 225-258.

Müller, E. M. & William. S. Brown. 1980. Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. Lass (ed.) *Speech and Language: Advances in Basic Research and Practice*. Academic Press, New York: 317-389.

Myers, Scott P. 1987. *Tone and the Structure of Words in Shona*. Ph. D. Dissertation, University of Massachusetts, Amherst.

Myers, Scott P. 1989. Structure preservation and the phonology of complex segments in Shona. *Paper presented at the LSA Annual Meeting, Washington, D.C., December 1989.*

O'Neil, J. 1935. *A Shona Grammar: Zezuru Dialect*. Longmans, Green, London.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese Tone Structure*. M.I.T. Press, Cambridge, Mass.

Pongweni, A. J. C. 1971. *The Phonetics and Phonology of the Karanga Dialect of Shona*. Ph. D. Thesis, University of London.

Pongweni, A. J. C. 1981. A spectral study of Karanga stops. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunicationsforschung* 34:

Pongweni, A. J. C. 1983. An acoustic study of the qualitative and pitch effect of breathy-voice on Shona vowels. *Journal of Phonetics* 11:129-138.

Sagey, E. 1986. *The Representation of Features and Relations in Non-linear Phonology*. Ph. D dissertation, M.I.T., Cambridge, Mass.

Coronal Places of Articulation*

Patricia A. Keating

1. Introduction

Coronal consonants are probably universal in the world's languages. Maddieson's (1984) statistical sample contains no languages without at least some coronal consonant, and only one language without coronal obstruents[1]. A number of typological observations support the special status of coronal consonants. Firstly, coronals include more contrasts of both place and manner than do other consonant classes. For example, with respect to manner, affricates and liquids are most often coronal. With respect to place, Maddieson's survey recognizes five primary places of articulation that are commonly classified as coronal (dental, alveolar, palatoalveolar, retroflex, palatal), and only five other primary places (bilabial, labiodental, velar, uvular, pharyngeal), so that coronals account for half of the primary places of articulation.

Secondly, coronals account for a high proportion of consonants in languages. For example, of the 20 consonants in Maddieson's modal inventory, ten are coronal. Also, Maddieson found that across the languages in the sample, the preferred inventory of stops and affricates contains three stop places of articulation (dental or alveolar, labial, and velar), plus one affricate place of articulation (palatoalveolar). Thus two of these four stop/affricate place categories are coronal. Further, if a language has four (rather than three) stop places, then again, two of the four are usually coronal[2].

---

*This paper is to appear as a chapter in *The Special Status of Coronals*, edited by Carole Paradis and Jean-Francois Prunet. Work was supported in part by NSF grant BNS 8418580.

[1] Maddieson does not use the term or category "coronal". The observations here are based on his findings but are couched in other terms. It should be noted that labial and velar categories are also almost universal.

Thus there are more coronal consonant types, and languages use them more. Put simply, coronals are special phonologically because there are so many of them. Presumably this sheer preponderance of coronal consonants is a factor in the status of coronals as the usual unmarked or unspecified place of articulation: if half of the consonants in a language are coronal, then any given consonant is more likely coronal than any other place class. In phonetic terms, coronals are special because they can be made in so many ways. The tongue blade seems to lend itself to a greater variety of articulations. What articulatory distinctions made with the tongue blade are linguistically relevant?

In this paper, the variety of possible coronal places of articulation is examined. We consider traditional place of articulation distinctions plus some "manner" distinctions that are generally used to make fine place distinctions. Some of the other manner distinctions found among coronals, such as lateralization, stridency, trill/tap, gradations of stricture, various release types, and certain secondary articulations, are not discussed here. The paper is organized as follows. In Section 2, some necessary terminology is reviewed, and anatomical definitions are discussed. In Section 3, various coronal places of articulation are described. Features that have been used to characterize these places are considered in Section 4. Section 5 provides a summary discussion.

2. Terminology

2.1. Tongue

Coronals can be defined as segments produced with the blade (including the tip) of the tongue. It was noted above that among the generally-recognized coronal places of articulation are dental, alveolar, palatoalveolar, retroflex, and palatal. (Palatoalveolar refers to the place of English [š] (IPA [ʃ]), while palatal refers to the place of the front glide [y] (IPA

---

[2] Maddieson describes the same data differently, because he classifies palatals and palatoalveolars as tongue-body articulations. However, both of these are now standardly considered coronal by phonologists.

[j]). American usage of "palatal" often encompasses both of these.) The IPA also includes alveolopalatals. Ladefoged and Maddieson (1986) add two less-common coronal places, linguolabial and interdental. These places of articulation lie from front to back in the mouth from the upper lip (linguolabials) to the hard palate (palatals), that is, virtually the entire span that can be touched or approached by the tip or blade of the tongue.

What part of the tongue counts as the blade? Different sources give different answers to this question. Catford (1977, 143) notes that there are two traditions: one from British phonetics, which he adopts, in which the blade is "the part that lies opposite the teeth and alveolar ridge when the tongue is at rest", i.e. just the tip plus 10-15 mm; and one from American speech science (see also Daniloff (1973, 173)) in which that part is called the tip, while the blade lies further back. Ladefoged (1982, 4) defines the tip and blade as "the most mobile parts" of the tongue, and Ladefoged (1989) defines the blade as the part not attached to the floor of the mouth, roughly corresponding to the part below the alveolar ridge. Ladefoged considers the blade to be a bit shorter than Catford suggests, no more than a centimeter long.

However, linguistically speaking the blade must be taken to extend somewhat further back than Catford or Ladefoged suggest. A sense of the extent of the blade in its linguistic uses can be gleaned from the following point. Alveolar stops and fricatives can be produced with the tongue tip down behind the lower teeth and some further back part of the tongue forming the constriction at the alveolar ridge. The phonological notion "coronal" surely depends on such articulations being made with the "blade" of the tongue, yet they are formed more than a centimeter behind the tip. Dart's (1988) linguograms agree with this observation. In my own case this suggests a blade length on the order of 15-20mm. The part of the tongue one centimeter behind the tip reaches only to the upper teeth.

To some degree such differences of definition may be a function of how extended the tongue is. The blade can be moved quasi-independently of the rest of the tongue (e.g. protruded, curled, wiggled). If the tongue is at rest in the mouth, this movable part will appear quite small, but if the tongue is extended out of the mouth or stretched in any other way, it will appear quite large because it is stretched. Thus if one considers the blade to be the part of the tongue that can be grasped in one's hand, and if one protrudes one's tongue to grasp it, then the blade will appear to be much longer than 10-15 mm. Perhaps in articulations with the tongue tip down, the tongue blade similarly stretches itself.

The definition of tongue tip also requires mention. Catford (1977) distinguishes between the very "apex" itself, and the "rim" around it. However, it seems just as valid to follow Ladefoged (1989) in considering the tip to include both of these at once, since in practice it is nearly impossible to use the very tip of the tongue without also involving a couple of adjacent millimeters.

Thus we will consider the blade of the tongue to be the movable part extending from one to two centimeters behind the tip, and we will consider the tip to include a small rim around the end of the tongue. Articulations with the tip are called apical; those with the blade are called laminal. Articulations made with both at once can be called apicolaminal. Traditionally, laminal refers only to the top surface of the blade, while sublaminal refers to the lower surface. These points are summarized in Figure 1.

Apical vs. laminal is sometimes equated with another descriptive parameter, the position of the tongue tip as "up" or "down". The tip is said to be up when it is raised above the lower teeth, so that the view of the tongue from outside the mouth is of the lower surface of the tongue. The tip is said to be down when it is behind or below the lower teeth, so that the view from outside is of the upper surface of the tongue. Individual speakers of English differ especially in whether /s/ is tip up or tip down.

2.2. Palate

As noted earlier, coronal articulations extend from the upper lip to the hard palate. Key divisions along the palate are represented in Figure 1. Behind the upper teeth is the alveolar ridge, a source of some confusion in articulatory descriptions. For phonetic purposes, the alveolar ridge is the

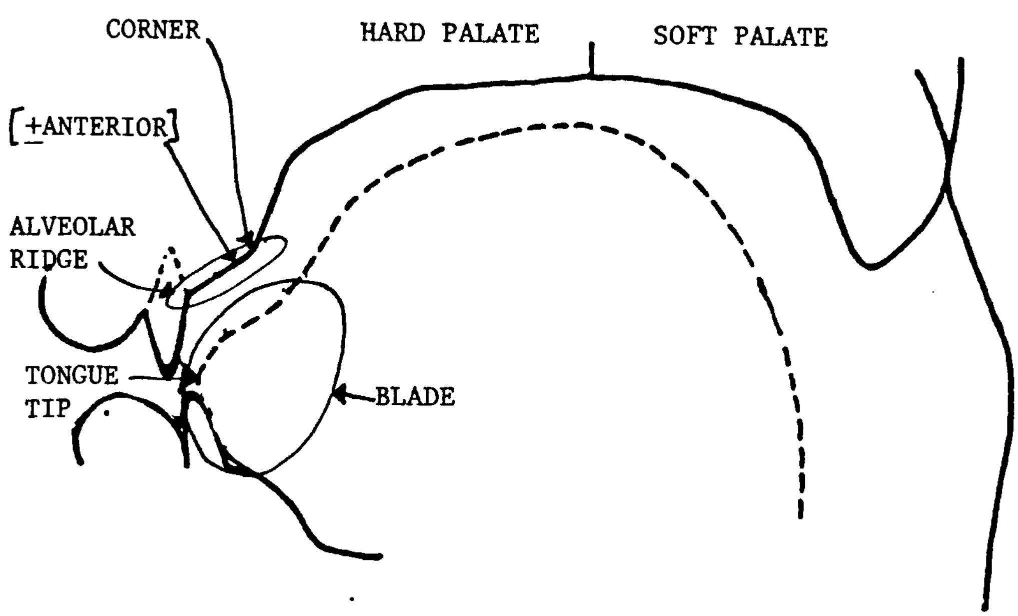


Figure 1. Overview of relevant anatomical distinctions: tongue tip and blade; alveolar ridge, corner, hard palate, soft palate; dividing point between [+anterior] and [-anterior].

entire area from the upper teeth back to the prominence at which the palate starts angling upward towards the roof of the mouth. This prominence is sometimes called the ridge, but can also be referred to as the "edge", "center", "corner", "turning point" or "protuberance" of the ridge. The alveolar ridge is this whole area, not just the prominence. Catford (1988, 86-7) has a helpful discussion of this point.

Given such definitions, we can now proceed to consider the variety of coronal places of articulation available to languages.

3. Descriptions of Coronals

In this section, the articulations of some of the coronal consonants are discussed. The observations are based on discussions in the literature and on study of published physiological data, especially X-ray tracings but also palatography.

3.1. Anterior Coronals

Coronals which are [+anterior] have their contact or constriction on the front part of the alveolar ridge, on the upper teeth, or, in the case of linguolabials, the upper lip. Linguolabials, interdentals, dentals, and alveolars are variably apical or laminal. Still, one might view linguolabials and interdentals as variants of a basic sound type, sharing an extension and protrusion of the blade, and differing largely in terms of apicality. Linguolabials would be primarily apical, in the sense that the tip is aimed at the upper lip, though it sometimes overshoots. Interdentals would be primarily laminal, in the sense that the blade contacts the teeth, but sometimes the tip does not quite protrude.

Figure 2 shows a dental and an alveolar. Dart (forthcoming) provides details about dental and alveolar articulations, particularly about cross-speaker variability in apicality. In both French and English, speakers vary in the place and the manner of their dental/alveolars. For example, Dart presents data that refute the claim by Ladefoged and Maddieson (1986, 78) that dental sibilants are always apical: 6 of the 14 dentals in her sibilant sample were laminal. See Ladefoged and Maddieson for further discussion of a variety of anterior coronals, especially strident vs. non-strident fricatives.

3.2. Palatoalveolars

Palatoalveolar constrictions (as for English [ʃ], Figure 3) are at or near the corner of the alveolar ridge. The tip may approach the ridge in front of the corner, while the blade approaches the corner; thus the blade runs parallel to the ridge. In these cases the articulation is both apical and laminal at once, and so the constriction is fairly long (and thus should be counted as primarily laminal rather than apical). However, for speakers with a prominent corner, coming nearly to a point, a laminal constriction can be quite short. Palatoalveolar articulation is most often laminal, sometimes apical. However, even the laminal articulation can have the tip "up", that is, raised above the lower teeth. Basically the tip lies behind the upper teeth, but far enough away from them that no dental constriction is formed. The tip is above the lower teeth so that a cavity can occur behind them, under the tongue. Catford (1977, 158) shows an articulation of this sort.

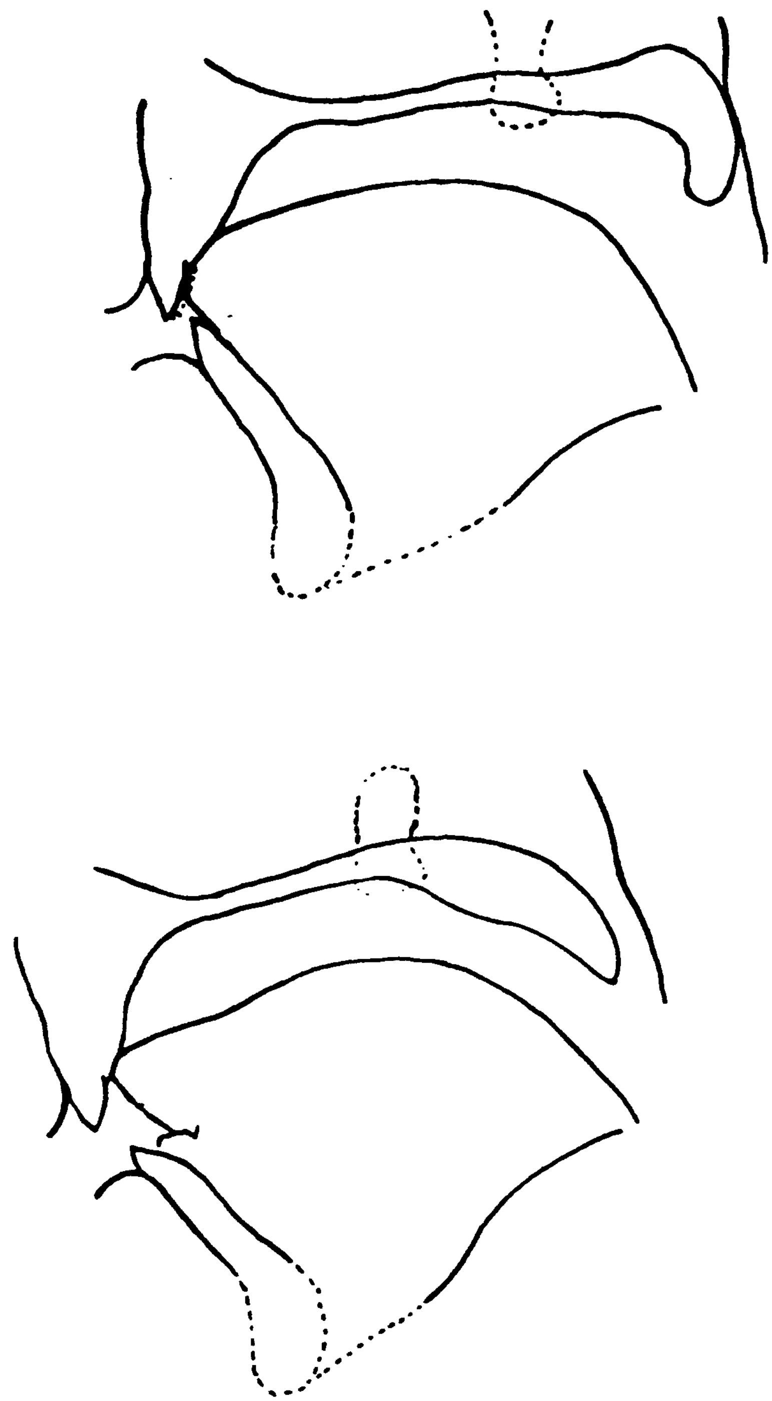


Figure 2. Dental (denti-alveolar) stop and alveolar nasal in French, after Simon (1967). Both are tip up, but the first is apico-laminal, while the second is apical.

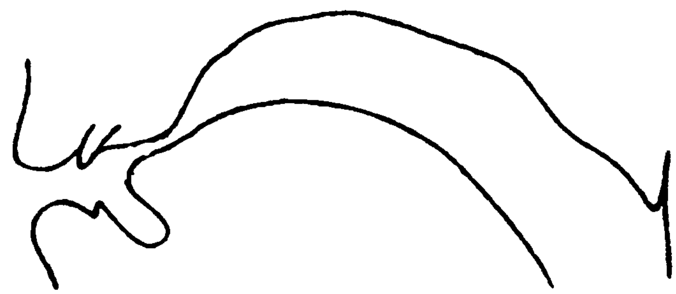


Figure 3. Palatoalveolar fricative in English, after Ladefoged and Maddieson (1986).

Palatoalveolars are also reported with a tip down articulation. However, it seems unlikely that this could ever mean that the tip contacts the lower teeth, since no cavity would be formed under the tongue. More likely, the "tip down" palatoalveolars have the tip just below the upper teeth, but free of the lower teeth.

Palatoalveolars also have a somewhat "domed" or convex tongue behind the constriction, which Ladefoged and Maddieson (1986) characterize as a slight degree of palatalization.

3.3. Retroflexes

Figure 4 shows two kinds of retroflexes. Many apical and sublaminal retroflexes (Figure 4a) involve curling back the tongue blade so that the tip or the underside of the blade forms a constriction along the palate. With just a slight curl, the very tip can touch the rear part of the alveolar ridge, in front of the corner. However, more commonly the constriction is behind the corner; the further back it is, then the more curled and stretched the tongue, the more the underside of the blade is used, and the longer the constriction. Ladefoged and Maddieson (1986) note that this description applies most clearly to stops; the retroflex fricatives in the languages of India are not as well documented, but seem not to involve the same kind of

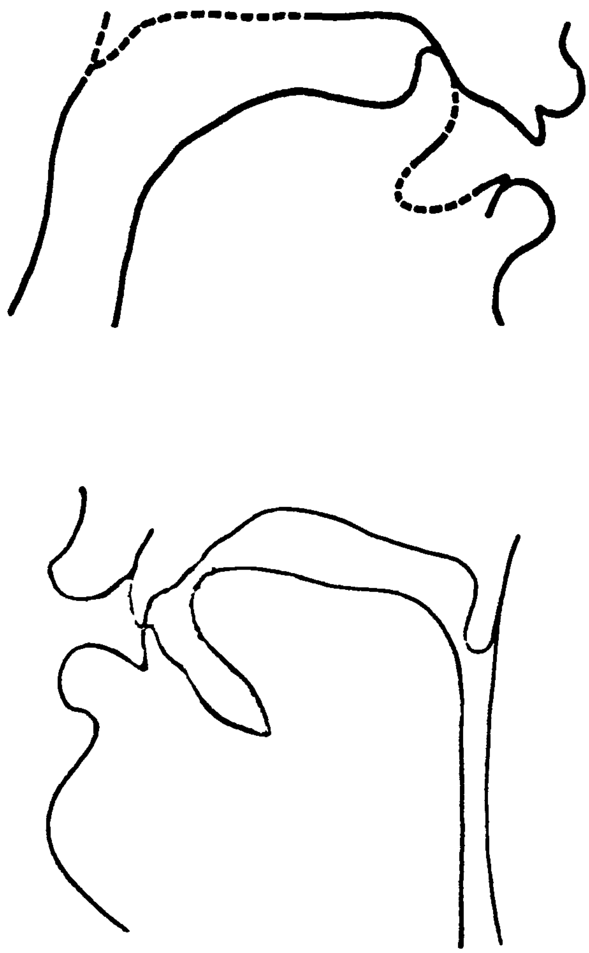


Figure 4. Retroflexes. 4a, sublaminal stop in Tamil, after Ladefoged and Maddieson (1986); 4b, flat apical fricative in Serbian, after Miletič (1960).

curling of the tongue. They have the same place of articulation on the palate as the stops, but the blade is not extended out from the body of the tongue. This makes it difficult to distinguish the tip from the rest of the blade in X-ray tracings. However, it should be noted that several tracings of Russian "/š/" and "/ž/" (e.g. Oliverius (1974), Dem'janenko (1966)) are clearly retroflexes of the expected type: apical with the tongue curled back.

A somewhat different kind of retroflex fricative (Figure 4b) is also described by Ladefoged and Maddieson (based on earlier work). These sounds are found in Mandarin Chinese and in Slavic languages, where they are often transcribed as palatoalveolars, though they sound more like other retroflexes[3]. Relative to palatoalveolars, or to the tongue at rest, the entire blade is moved up and back and is positioned just behind the corner of the alveolar ridge. The tip is up, and the tongue is flat from front to back, not domed. Ladefoged and Maddieson categorize them as (laminal) flat postalveolar sibilants, with a sublingual cavity. They describe the constriction as like that of [ʃ], but at the center of the alveolar ridge. They also note an articulatory difference between the versions found in Polish vs. Chinese: the former are rounded while the latter have a larger sublingual cavity. (These fricatives are both said to differ from the retroflex fricative of Tamil, which is further back, possibly apical, and has a larger sublingual cavity.)

Although Ladefoged and Maddieson characterize these retroflexes as laminal, data sources show more variability. The apicality of the contact for affricates is relatively easy to determine from available data. The retroflex affricates in the X-ray tracings of Ladefoged and Wu (1984) are either apical or laminal, though in either case with the tip up. Linguograms and palatograms, along with X-ray tracings, are available for the fricatives and corresponding affricates of Polish (Wierzchowska (1965; 1967; 1980)) and Serbian (Miletič (1960)). In these records, the stop portions of the affricates are clearly apical, possibly partly sublaminal. The fricatives also appear to be apical, in the sense that the linguograms show no narrowing anywhere along the blade. Since the palatograms show that there is indeed a constriction, it must be the tip forming it. The difference between Slavic and Dravidian fricatives, then, would appear to be in the location (backness) of the constriction, and thus in the size of the sublingual cavity.

---

[3] Contrasts between retroflexes and palatoalveolars are rare.

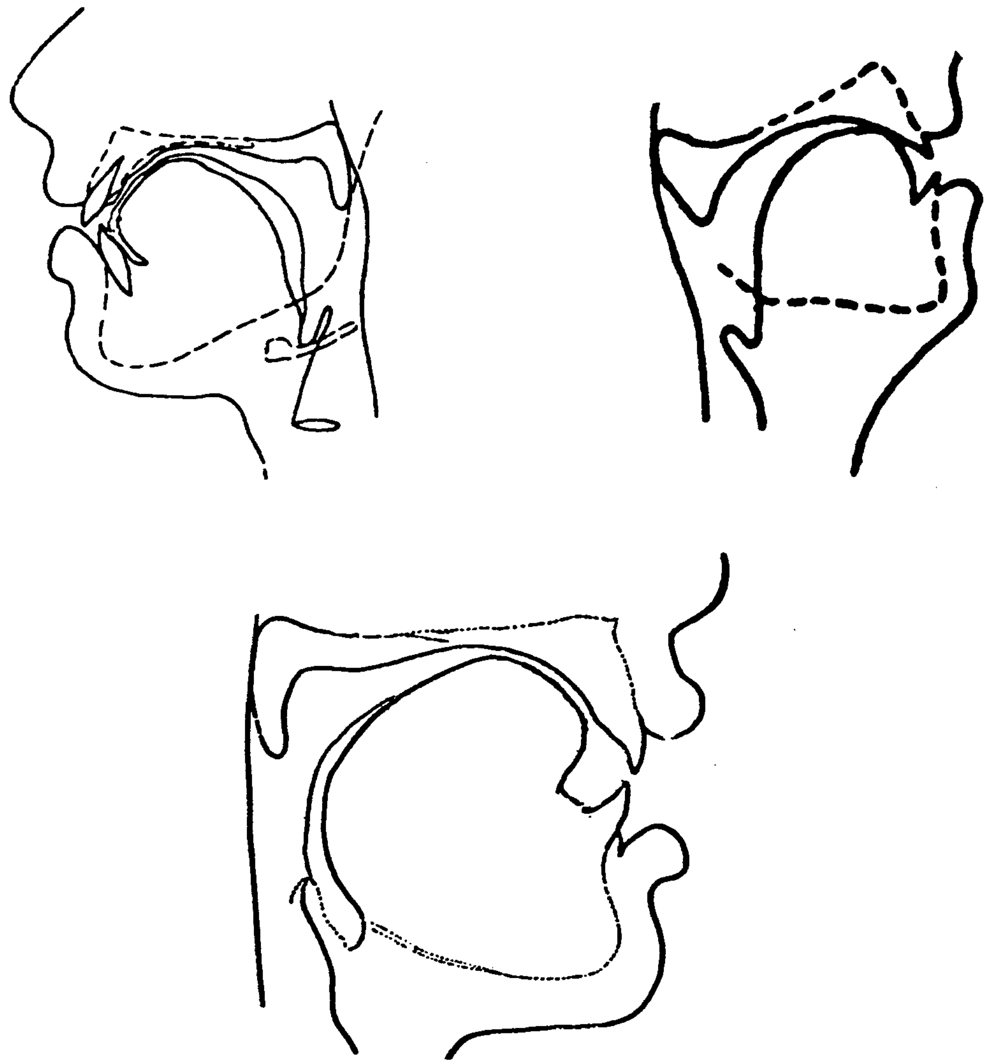


Figure 5. Alveolopalatals. 5a, Polish fricative, after Wierzchowska (1967; 1980) 5b, Polish affricate from same source; 5c, Mandarin fricative, after Ohnesorg and Śvarný (1955).

3.4. Alveolopalatals

Figure 5 shows three kinds of alveolopalatals. Alveolopalatals, or "prepalatals", probably occur most commonly as nasals and laterals, where they are generally confused with palatals. (For example, Maddieson (1984) collapses these categories.) They also occur as fricatives and affricates, for example in Polish and Mandarin [ɕ], where they sound like sharpened palatoalveolars or strident palatals. They most commonly involve the blade approaching the corner of the alveolar ridge. The tip is usually down, pointing to the lower teeth, but often does not touch them; however, tip up examples can also be found, e.g. Ladefoged and Wu (1984). In either case there may or may not be a cavity under the tongue. The front of the tongue is raised behind the constriction. Available X-rays of alveolopalatals (in

Mandarin and Polish) show quite a bit of variation, even within languages.

Figure 5a shows a fricative from Wierzchowska (1967; 1980), with a long constriction, the tip behind the lower teeth, and a small sublingual cavity. Figure 5b shows another tracing from the same author, this time of the stop component of an affricate just before the release. Here, the requirement of complete occlusion leads to a raising of the blade, with the tip also raised, resulting in a slightly larger sublingual cavity. Figure 5c shows a Mandarin alveolopalatal fricative with an even larger sublingual cavity, one as large as for other nonanterior coronals.

It should be noted that Halle's descriptions of supposed Polish alveolopalatal fricatives (Chomsky and Halle (1968), Halle (1988)) are also based on tracings from Wierzchowska. His descriptions are consistent with the figures he relied on (p.c.). However, judging from the transcription and example words given by the author, these figures seem to be of palatalized anteriors, not of alveolopalatals. Therefore the constriction is more forward, and shorter, than in the figures reproduced here.

3.5. Palatals

In SPE, palatals (such as [j] and [ç]) were considered to involve tongue body articulations, and so were [-coronal], but they were later reclassified as coronals on phonological grounds. (See Keating 1988b, among others, for summary.) Halle and Stevens (1979) proposed to redefine coronal to mean the blade *or front* of the tongue so as to include the palatals. However, this move seems unnecessary, as palatals generally do involve the blade proper, in addition to the front of the tongue.

Figure 6 shows a palatal stop. The key component of palatals seems to be that they are articulated near a large part of the hard palate, between the alveolar ridge and the roof of the mouth (Keating (1988a)). The tongue is both raised and fronted from its position for [i] vowels, so that parts of the blade and the front form a very long constriction. The tip, and the front part of the blade nearest the tip, are not involved and are usually low in the mouth, so that there is no sublingual cavity. Palatograms show that

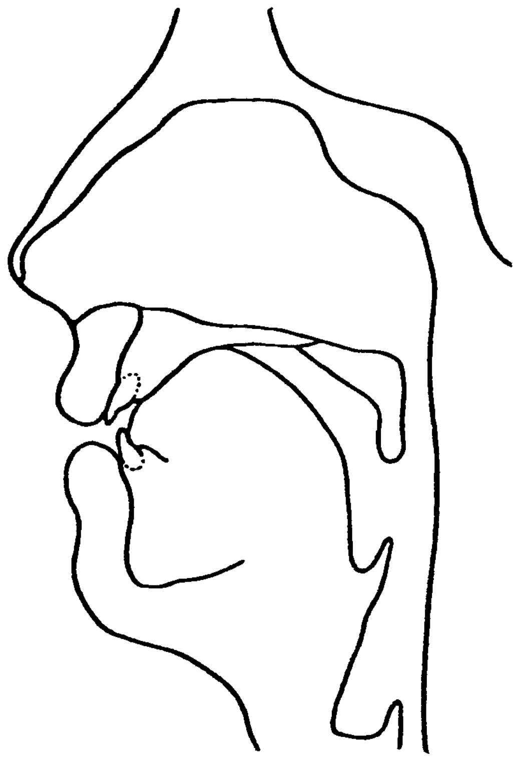


Figure 6. Palatal nasal stop in Czech, after Hála (1962).

the occlusion for stops is about the length of a velar constriction, but quite front; the blade touches just behind the alveolar ridge. Thus the stop occlusion itself is coronal, and nonanterior. At the same time, there is extensive side to side and front to back lateral contact as for [j], and the entire front of the tongue is extremely close to the palate. Non-stops have more open constrictions covering about the same area.

One basic observation here is that palatals have a very large constriction area, probably the largest of any outside the pharynx. A second basic observation is that palatals are articulated much further forward in the mouth, and on the tongue, than has often been assumed. Although the palatal place of articulation is "next to" velars, these are very far apart in practice. Palatals are further forward along the palate even than velars fronted in a front vowel context. There is room along the roof of the mouth for three different places of articulation, with fronted velars in between palatals and velars. In Keating (1988a) I proposed that the SPE tongue body feature values assigned to palatals be used instead for fronted velars. In particular, the value [-back] would refer to an articulation on the hard, rather than the soft, palate; thus fronted velars would be [-back] while nonfronted velars would be [+back]. The representation of palatals will be discussed below.

3.7. Palatalized Coronals

Thorough coverage of all the secondary articulations that can affect coronals is beyond the scope of this paper. However, in the case of palatalization, the secondary articulation can effect a change in the primary place and/or manner of articulation, and thus needs to be considered here. As a technical phonetic term, palatalization refers to the superposition of a high front tongue body position on a separate primary articulation, e.g. a primary articulation with the tongue blade. However, Bhat (1978) emphasizes that "palatalization" is used as a cover term for any combination of three independent articulatory components: tongue fronting, tongue raising, and spirantization. He points out that the term palatalization, in its wider use, more often refers to restricted changes in certain primary places of articulation, as when velars palatalize to palatoalveolars. It less often refers to a general secondary articulation across all the primary places in a language, as in Russian, where labials, coronals, and velars may all come in surface contrasting pairs of palatalized vs. non-palatalized.

3.7.1. Anterior coronals. Russian has surface contrasts of plain vs. palatalized anterior coronals. Bhat shows that, across languages, anterior coronals are more likely to undergo tongue raising than either tongue fronting or spirantization. Tongue raising of coronals usually results in retracted and laminal articulations. The X-rays of Russian coronals in Oliverius (1974) show this effect quite clearly.

3.7.2. Nonanterior coronals. Secondary articulations involving the tongue are very rare with [-anterior] places of articulation. However, surface contrasts do occur; in Russian, between retroflex and palatalized retroflex fricatives, and in Polish, between palatalized retroflex, retroflex, and alveolopalatal fricatives. The Russian palatalized retroflex looks straightforwardly like a palatalized version of the plain (curled) retroflex. However, X-rays of Abkhaz reproduced by Ladefoged and Maddieson (1986, 77) show that in that language, the alveolopalatal looks like a palatalized version of the retroflex, which is of the flat-tongued, apical type. Ladefoged and Maddieson therefore analyze it as such. Furthermore, the palatalized retroflex of Polish shown in Wierzchowska (1965) looks very much like the alveolopalatal of Abkhaz, supporting Ladefoged and Maddieson's analysis. On this account, there is no separate *place* of articulation for alveolopalatals; they are collapsed with the retroflexes, and only the palatalization distinguishes the two. The change from apical retroflex to laminal palatoalveolar would be a natural concomitant of palatalization. The problem, however, is the fact that Polish also has an alveolopalatal, contrasting with its palatalized retroflex. Since the alveolopalatal then cannot be just a palatalized retroflex, how are these to be analyzed? Tokens vary, but overall the three Polish fricatives lie on a continuum of tongue body raising. The retroflexes have a flat tongue, the alveolopalatals have a very raised and fronted tongue, and the palatalized retroflexes fall in between. Since the Abkhaz alveolopalatal looks like the Polish palatalized retroflex, the Polish alveolopalatal represents a more extreme palatalization. These relations are summarized in the table below. It might be possible to vary the feature values used to represent the palatalization so as to distinguish these two Polish types; for example, whether both Back and High are used.

TONGUE PROFILE BEHIND CONSTRICTION

LANGUAGE	flat	raised	palatalized
Polish	ṣ̌	ṣ̌ʲ	ɕ
Abkhaz	ṣ̌	ɕ	

It is not clear that palatoalveolars are ever palatalized. Reported cases, as in Slavic, instead seem to involve retroflexes.

4. Features Proposed for Coronals

Coronal segments have as their active articulator the tongue blade, and therefore can be specified with a positive value for the Coronal feature[4]. To distinguish the various coronal places of articulation, further features are needed. Though the issue will not be addressed here, these features must also be capable of expressing the natural class relations among coronals. We will consider first some of the standard SPE features (for further related discussion see Keating (1988b)), then some others.

4.1. Anterior

The feature Anterior describes the place of articulation, not the active articulator. It divides coronals into more-front and more-back categories, determined by their place of articulation along e.g. the roof of the mouth. The operational definition provided by Chomsky and Halle (1968) is that alveolars are [+anterior] while palatoalveolars are [-anterior]. The phonetic basis of this division has scarcely been discussed in the literature and has not received a precise articulatory description. It is often described in terms of the alveolar ridge: [-anterior] segments are formed behind the alveolar ridge, or more exactly, behind the corner of the alveolar ridge. Alveolars are said to be articulated in front of this point, and palatoalveolars behind it (e.g. Ladefoged (1989, 48)). For speakers with prominent alveolar ridges, this would be a clear articulatory distinction, and thus a clear boundary between the values of Anterior.

However, examination of X-ray data shows that this characterization is incorrect. Both values of Anterior can be found in front of the corner. Alveolars are articulated on the frontmost part of the ridge. Palatoalveolars are generally articulated at about the corner, either centered there, or extending into the part of the ridge in front of the corner. (See Ladefoged and Maddieson (1986, 65-67).) English readers can feel this for themselves by saying "chop" -- the stop component of the

---

4 In Chomsky and Halle (1968), this meant a [+coronal] feature value; in more recent feature hierarchies, it means the presence of a Coronal articulator node.

affricate is not made behind the corner, but instead at a point just behind where the /t/ in "top" is made. Both palatoalveolars and retroflexes can be made at or just in front of the corner of the alveolar ridge, so that they are only minimally different in place from alveolars. Thus palatoalveolars look like alveolars, but with the whole tongue moved back and up just a little; retroflexes can also look like alveolars, but with the blade curled back just a little more. (Sublaminal retroflexes can also be made well behind the corner, of course.) Thus the true dividing point between the values of the feature Anterior appears to be the midpoint of the part of the alveolar ridge between the upper teeth and the corner. This point is summarized in Figure 1.

Considered only in terms of millimeters difference between constriction locations, the difference between [+anterior] and [-anterior] is incredibly subtle. However, the corner of the alveolar ridge provides a more definitive landmark to which the tongue may orient itself for the [-anterior] articulations.

In light of this discussion and Figure 5, it seems phonetically incorrect to view the alveolopalatals of Polish as [+anterior], as Halle does (Chomsky and Halle (1968), Halle (1988)). These segments are at best variable in anteriority. For the most part they are articulated at the corner of the alveolar ridge, which counts as [-anterior]. Phonologically, they pattern as palatalized variants of dentals, and though the dentals are of course [+anterior], a change in anteriority under palatalization is in accord with the cross-language observations of Bhat (1978).

4.2. Distributed

The feature Distributed uses a description of length of the consonantal constriction (i.e. a manner property) to represent differences in place of articulation. Chomsky and Halle (1968) proposed this feature as subsuming the traditional apical/laminal distinction, with apical articulations having shorter ([-distributed]) constrictions and laminal articulations having longer ([+distributed]) ones. (Sometimes the name Apical/Laminal is used instead of Distributed, e.g. Clements (1989).) Chomsky and Halle make clear that there is no intended a priori correspondance between the values of

Distributed and place of articulation; both dentals and alveolars may be distributed or nondistributed, and it is left to the low-level phonetic rules in a language to specify the exact place of articulation of any coronal segment. It is even possible that in some particular case a laminal articulation might be shorter than an apical one (for example, if a speaker with a very sharp corner of the alveolar ridge made a laminal palatoalveolar, but a sublaminal retroflex). In this case, the usual correspondance between Distributed and apical/laminal would be reversed.

In general, dentals and alveolars do differ in other ways besides their place, and apicality is one of the differences observed. With stops, as Ladefoged (1989) discusses, dentals are more likely to be laminal, and alveolars, apical[5], and thus Distributed can usually be used to distinguish these places. Ladefoged and Maddieson (1986) report only one case of anterior coronals contrasting in place but not in apicality or any other feature, namely, apical fricatives in certain Amerindian languages. Dart (1988, forthcoming) studied the dental - alveolar stop contrast in Papago, where both places are said to be apical. However, all of the speakers who made any contrast used at least somewhat different articulators: either the only difference was in apicality, or apicality varied along with place (the dentals were tip down laminals, or tip up apicolaminals). Furthermore, the "alveolar" stops were usually actually post-alveolar, so that only the dentals are in fact [+anterior]. That is, the Papago case turns out to support Chomsky and Halle's claim that place alone never distinguishes anterior coronals.

The same result holds of another case presented by Ladefoged and Maddieson. They note that the two apical laterals of Albanian differ not only in place but also in tongue body backness. However, it appears from their figure that they would also differ in their values for Anterior, as happened in Papago.

In general, the use of the tip vs. the blade is often not consistent

---

[5] Surprisingly, Chomsky and Halle (1968, 314) describe dentals as most usually [-distributed], but I believe this is not common usage.

enough to be relied on as the basis of phonological distinctions. It is important to note that while this is a highly salient aspect of coronal articulations, it is largely a matter of speaker choice, not definition of sound types. Dentals, alveolars, and palatoalveolars can be made either apically or laminally, and retroflexes can be made either apically or sublaminally. Dart (1988; forthcoming) shows that French and English, languages without a contrast in apicality, permit great speaker variability in dental or alveolar stops and fricatives. Neither the place nor the apicality is invariant across speakers within a language. However, as Catford (1977) points out, an apical vs. a laminal articulation will have acoustic effects within the "same" place of articulation category. In particular, the size of any sublingual cavity will vary with the position of the tongue blade, and this in turn will affect the resonance frequencies of obstruent noise.

In its original form, where Distributed describes constriction length quite generally, it is equally well used for other constriction types. Chomsky and Halle employ it to distinguish alveolopalatals from other places of articulation in Polish. Alveolopalatals have the tongue front raised up behind the blade, and so may have rather longer constrictions than otherwise similar laminal coronals. Since Chomsky and Halle considered alveolopalatals to be [+anterior], they used Distributed to distinguish them from the dentals[6]. With alveolopalatals as [-anterior], Distributed would instead distinguish them from the Polish retroflexes.

It is useful to ask how much the coronal articulations actually differ in constriction length, i.e., whether the phonetic definition of Distributed in its SPE usage is supported empirically. Chomsky and Halle, after all, rely on very little data in this regard. I therefore measured the length of contacts or constrictions from tracings of a wide set of coronals; to allow comparison across speakers, these were compared with velars where possible. Alveolo-palatals and especially palatals usually have quite long

---

6 Thus Distributed can be seen to be a relative property: when dentals form a phonological contrast with alveolars, they might be [+distributed], but when they contrast with alveolopalatals, they are [-distributed].

constrictions, longer than those of velars. Retroflex stops have constrictions about as long as those of velars -- longer than expected, given their usual classification as [-distributed]. Beyond this, no clear differences emerge. Though laminal constrictions are longer than the shortest apical constrictions, apicals can also be long. Palatoalveolars can sometimes have among the shortest constrictions, in speakers with sharply defined alveolar ridges. Thus there appears to be little available physiological support for this phonetic definition of Distributed. This finding supports limiting the feature Distributed (by this or some other name) to the apical/laminal distinction.

4.3. Sublingual cavity

Stevens and colleagues (Perkell et al. (1979)) have called attention to the importance of the presence of a cavity under the tongue blade during the articulation of palatoalveolars, because of its lowering effect on acoustic resonances. The same is true for retroflexes. At first glance, then, the sublingual cavity would seem to be a correlate of [-anterior] segments. However, some [-anterior] coronals lack it. In particular, the absence of a sublingual cavity is a consistent and key characteristic of palatals. Also, Ladefoged and Maddieson (1986) discuss a rare sibilant fricative in Abkhaz, described by Catford, which is palatoalveolar but has the tip down and no sublingual cavity ("hissing-hushing"). The [-anterior] coronals can be arranged in order of increasing size of sublingual cavity, from the palatal and hissing-hushing, to the alveolopalatal, the palatoalveolar, the apical retroflex, to the sublaminal retroflex.

Halle (1988) proposes a new tongue feature, Lower Incisors Contact, to encode this property, with this contact implying no sublingual cavity. Halle thus distinguishes alveolopalatal from dental/alveolar (all as [+anteriors]) and palatal from palatoalveolar (all as [-anterior]). We have already noted that phonetically alveolopalatals are in fact usually [-anterior]. Thus Lower Incisors Contact plays no contrastive role among the true [+anterior] places (dental and alveolar). Furthermore, most X-rays of Polish alveolopalatals show at least a small sublingual cavity, implying no contact between tongue and teeth (see Figure 5). In most alveolopalatals, the tongue tip points at, but does not touch, the lower teeth. If anything, then, the presence of Lower Incisors Contact distinguishes most alveolopalatals from palatals, taking both as [-anterior] and [+distributed].

Halle's name for this sublingual cavity feature, Lower Incisors Contact, suggests a correspondance with another traditional phonetic descriptive dimension for coronals we have referred to, namely, "tip up" vs. "tip down". When there is lower incisors contact, clearly the tip is down. (However, for interdentals, the tip rests on the lower teeth, blocking off any sublingual cavity, yet might be considered "up".) In contrast, to guarantee a cavity large enough to affect the acoustic output substantially, the tip is best raised above the lower teeth; this is what is observed for most palatoalveolars and retroflexes. In these two cases, then, Lower Incisors Contact would correlate well with tip position. The only question would be whether there are cases where the tip is down but does not make lower incisors contact. We have suggested that this is the case with some palatoalveolars: they are reported as tip down but nonetheless have a sublingual cavity. Therefore the feature Lower Incisor Contact, or sublingual cavity, is not exactly equivalent to tip up or down, unless by tip "up" we mean any position above the base of the lower teeth.

Ladefoged and Maddiesion (1986) instead equate tip position with apical/laminal: tip up is apical, while tip down is laminal. However, a similar discrepancy is met here. Tip position does correlate with apicality if the tip is down: then the articulation must be laminal. But the reverse is not necessarily true. The tip may be raised only to the level of the upper teeth, and so be "up", while the constriction is formed laminally on the palate. Palatoalveolars are an example of this. (The flat retroflexes described as laminal by Ladefoged and Maddieson (1986) would also be examples, but it was suggested above that these are in fact apical.) Therefore apical or laminal is not exactly equivalent to tip up or down, unless by tip "down" we mean any position below the base of the upper teeth. In sum, then, none of these descriptive parameters - sublingual cavity, apicality, tip position - quite co-vary.

4.4. Tongue shape features

Ladefoged and Maddieson (1986) offer additional descriptive parameters for coronals, which provide phonetic detail which is redundant rather than contrastive in nature. One of these is constriction width (from side to side); a narrow constriction, as found for [s] sounds, involves grooving the tongue blade. Another parameter is pitting of the tongue behind the grooved constriction, again found for [s] sounds. That is, as Ladefoged and Maddieson point out, the grooving and pitting of the tongue in the formation of [s] sounds are important components of their articulation; feature descriptions in terms of Anterior and Distributed (or Laminal) alone do not give a complete phonetic description. The redundant detail is necessary to say exactly how the [s] sound is to be made. I would suggest that these parameters might be related to the feature Strident (or Sibilant): particular blade and body configurations, appropriate to the given place of articulation, are needed to produce the right kind of airstream jet for stridency. Thus, instead of themselves being features, they are phonetic parameters which are marshalled to help effect (or enhance) a phonological feature, such as [strident].

Ladefoged and Maddieson also use a new feature, flat vs. domed tongue shape, to distinguish retroflex from palatoalveolar fricatives. Both are "postalveolar" in place, and both are laminal on Ladefoged and Maddieson's account. The problem with using this phonetic parameter as a phonological feature is that the retroflex stops which correspond to the fricatives are domed, not flat, and thus the stops are grouped with the wrong set of fricatives. Instead, as I suggested above, the retroflex fricatives should be considered apical, like the corresponding stops.

Tongue shape features can also enter into the description of palatals. In Keating (1988a) I proposed that palatals are complex segments involving both coronal and tongue body articulations, with values for the tongue body features equivalent to palatalization. This complex representation makes the structure of palatals parallel to that of labial-velars, which also combine two major articulations. It also represents the direct articulatory relation between palatals and front vowels. However, another option in the representation of palatals is to treat them as simple coronals, and introduce at least one additional feature to distinguish them from the other [-anterior] coronals. This in effect is what Halle (1988) does with his new feature Lower Incisors Contact. Actually, both options could be exercised for more complete descriptive coverage. We have already seen that alveolopalatals as well as palatals might be viewed as palatalized, or complex, segments with a high front vowel component. As discussed above, phonetically speaking both are [-anterior] and [+distributed], so some further feature would be needed to distinguish them. Halle's cavity feature could be used in this way, with the palatals as [+lower incisors contact], and the alveolopalatals as [-lower incisors contact]. However, it must be noted that in the end, phonological evidence is needed to support the natural classes entailed by such proposals about features.

By using tongue body features, the proposal here is that palatals, and probably alveolopalatals, are treated as palatalized segments. We might ask, palatalized versions of what? We already discussed the palatalization relation between alveolopalatals and retroflexes. Palatals, by their feature values, would correspond to the Abkhaz hissing-hushing category. Both are [-anterior] and have the tongue tip behind the lower teeth with no sublingual cavity. The tongue lowering seen in the hissing-hushing fricative is replaced with extreme tongue fronting and raising in palatals.

5. Discussion

The main points of the paper can be summarized by showing how the features discussed above can characterize the coronal phonetic categories. Several ambiguities or inadequacies have been found.

Linguolabials and interdentals are both anterior. It was proposed that they differ in tip orientation, with linguolabials apical and tip up, and interdentals laminal and probably tip down.

Dentals and alveolars are also both anterior. When not in contrast with each other, they vary rather freely in apicality and tip position/sublingual cavity. When in contrast, they may be distinguished by apicality, stridency, or a secondary articulation.

Three types of retroflexes were discussed, all [-anterior] and all tip up with a sublingual cavity. The sublaminal retroflexes, attested most clearly for stops, would count as laminal (or [+distributed]) in most feature systems. The other two types of retroflexes are apical, occurring with either domed or flat tongue shapes. This distinction (which is never contrastive) poses a problem for current systems of phonetic description. A possible alternative description would use tongue body features such as Back.

Palatoalveolars are [-anterior], but vary in tip position and apicality. Most commonly, they are tip up but laminal. It seems likely that there are no apical palatoalveolars which are not at the same time also laminal (i.e. apicolaminal), and none which have the tip down to the point of lower incisors contact. The laminality distinguishes palatoalveolars from retroflexes.

A secondary articulation of palatalization was invoked to describe the palatals and alveolopalatals. These are both [-anterior] and laminal/distributed. They differ (though not always reliably) in tip position and presence of a sublingual cavity. The phonetic distinction here is problematic because of variability in the available data.

It can be seen that feature systems must be developed further to account for all of the possible coronal places of articulation. The problems of representation presented here, however, only serve to underline the great variety of coronals encountered in languages.

References

Bhat, D.N. Shankara (1978) "A general study of palatalization," in J.S. Greenberg, ed., Universals of Human Language, vol. 2: Phonology, pp. 47-92, Stanford University Press, Stanford, CA.

Catford, J. C. (1977) Fundamental Problems in Phonetics, Indiana U. Press., Bloomington.

Catford, J. C. (1988) A Practical Introduction to Phonetics, Clarendon Press, Oxford.

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

Clements, G. N. (1989) "A Unified Set of Features for Consonants and Vowels," Cornell University manuscript.

Daniloff, R. G. (1973) "Normal Articulation Processes," in F. D. Minifie, T. J. Hixon, and F. Williams, eds., Normal Aspects of Speech, Hearing, and Language, Prentice-Hall, Englewood Cliffs, NJ.

Dart, S. (1988) "Acoustic Correlates of Apical and Laminal Articulations," J. Acoustical Society of America 84, Suppl. 1, S112.

Dart, S. (forthcoming) Articulatory and Acoustic Properties of Apical and Laminal Articulations, UCLA dissertation.

Dem'janenko, M.J. (1966) Vstupnyi Fonetyko-Hrafichnyi Kurs Francuz'koi Movi Laboratorii Eksperimental'noi Fonetiki.

Hála, B. (1962) Uvedení do fonetiky češtiny: Na obecně fonetickém základě, Nakladatelství Československé Akademie Věd, Prague.

Halle, M. (1988) "Features," to appear in W. Bright, ed., Oxford International Encyclopedia of Linguistics, Oxford U. Press, Oxford.

Halle, M. and K. N. Stevens (1979) "Some Reflections on the Theoretical Bases of Phonetics", in B. Lindblom and S. Ohman, eds., Frontiers of Speech Communication Research, Academic Press, London.

Keating, P. (1988a) "Palatals as Complex Segments: X-ray Evidence", UCLA Working Papers in Phonetics 69, 77-91.

Keating, P. (1988b) A Survey of Phonological Features, distributed by Indiana University Linguistics Club, Bloomington.

Ladefoged, P. (1982) A Course in Phonetics, 2nd ed., Harcourt, Brace, Jovanovich, New York.

Ladefoged, P. (1989) Representing Phonetic Structure, UCLA Working Papers in Phonetics 73.

Ladefoged, P. and M. Halle (1988) "Some Major Features of the International Phonetic Alphabet," Language 64, 577-83.

Ladefoged, P. and I. Maddieson (1986) Some of the Sounds of the World's Languages (preliminary version), UCLA Working Papers in Phonetics 64.

Ladefoged, P. and Z. Wu (1984) "Places of Articulation: An Investigation of Pekingese fricatives and Affricates," Journal of Phonetics 12, 267-278.

Maddieson, I. (1984) Patterns of Sounds, Cambridge U. Press, Cambridge.

Miletič, B. (1960) Izgovor Srpskochravataskich Glasova, Belgrade.

Oliverius, Z. F. (1974) Fonetika Russkogo Jazyka, Státní Pedagogické Nakladatelství, Prague.

Ohnesorg, K. and O. Švarný (1955), Études Experimentales des Articulations Chinoises, Rozpravy, Vol. 65, No. 5, Československé Akademie Věd, Prague.

Perkell, J., S. Boyce, and K. N. Stevens (1979) "Articulatory and Acoustic Correlates of the [s-š] Distinction," in J.J. Wolf and D.K. Klatt, eds., Speech Communication Papers presented at the 97th Meeting of the Acoustical Society of America, Acoustical Society of America, New York.

Simon, P. (1967) Les Consonnes Françaises, Bibliothèque françaises et romane, Serie A, No. 14, Klincksieck, Paris

Wierzchowska, B. (1965) Wymowa polska, Panstwowe Zaklady Wydawnictw Szkolnych, Warsaw.

Wierzchowska, B. (1967) Opis Fonetyczny Języka Polskiego. Panstwowe Wydawnictwo Naukowe, Warsaw.

Wierzchowska, B. (1980) Fonetyka i Fonologia Języka Polskiego, Ossolineum, Warsaw.

Some reflections on the IPA

Peter Ladefoged

This paper is a commentary on both the newly revised International Phonetic Alphabet, and on the 1989 Kiel Convention of the International Phonetic Association at which it was produced. The new IPA chart is examined, and the pragmatic but conservative attitudes of the Association are described. It is shown that the IPA has a phonological basis, and that an IPA transcription has two parts: a text containing IPA symbols, and a set of conventions (rules) for interpreting the symbols. The paper concludes with a personal view of the problems of whether there is a finite set of speech sounds, and whether a sound in one language can be equated with one in another, suggesting that for the phonetician there is no universal truth independent of the observer.

The International Phonetic Alphabet (IPA) was revised at a convention in Kiel, Germany, in August 1989. This paper is a personal view of some aspects of the new IPA, and the convention that led up to it. It is in no way an official account of either. The convention was clearly An Event, with capital letters, which everyone who was there will recall with a sense that history was being made in the field. There had been a great deal of preparation, involving the soliciting of opinion by mail on a wide range of topics, and considerable research on the present status of the IPA. At the convention itself there were five major groups which met almost continuously for the first two days, discussing and arguing about every conceivable issue concerned with phonetic transcription. Three of the five main groups considered consonants, vowels and suprasegmentals; the other two considered computational aspects of the IPA, and the needs of speech pathologists and others for extensions of the IPA. In addition there were groups that met for shorter periods to discuss the principles on which the IPA should be based, the form of presentation of the IPA symbols on a chart, and past successes and failures of the IPA. All the groups reported back to plenary sessions that were held at intervals; and on the last day the convention met in continuous plenary sessions to consider and vote on the final working reports. A complete account of the form of the convention and of its decisions has been published in the *Journal of the International Phonetic Association*, 19.2, (1989), and the decisions taken have been officially approved by the Association's Council.

Perhaps the first thing to note about the revised version of the IPA is that it is very much like it predecessors, both in the particular symbols that were approved, and in the Association's beliefs concerning these symbols. The symbols themselves are easy to summarize. They are as shown in the chart in Figure 1. There are a number of new symbols, notably [ʙ] for a voiced bilabial trill, [ʟ] for a voiced velar lateral, [ʝ] for a voiced palatal fricative, and the diacritics for creaky voiced, linguolabial apical and laminal, and advanced and retracted tongue root. All of these sounds have been documented in the phonetic literature only comparatively recently (although they have been noted in the descriptions of individual languages for far longer). There are also some changes in particular symbols, notably the dropping of approval for [ɩ] and [ɷ] in favor of the more traditional [ɪ] and [ʊ], and the revising of the mid-high back unrounded vowel, changing [ɤ] to [ɤ], a symbol with curved upper branches making it look more like a ram's horn, and thus more different from the voiced velar fricative [ɣ].

The prevailing mood was to avoid making changes in specific symbols, unless a very strong case could be made. In this spirit, after considerable debate, the symbols for clicks were changed. The convention voted to approve [ǀ, !, ǂ,ǁ] on the grounds that these symbols, together with [ʘ] which was already an IPA symbol, were what were actually used by nearly all the scholars working on the Nguni and Khoisan languages. But this spirit of bowing to prevailing use did not extend to accepting the widespread American use of [š, ž, č, ǰ]. The convention was not sufficiently impressed by arguments (which to me, seemed logical) to the effect that these sounds formed a natural class, and thus it would be appropriate to recognize this by maintaining a common aspect to their symbolism.

Clearly the most controversial innovations were the symbols for the voiceless implosives. Within the phonetic literature there have been very few accounts of these sounds. Supple and Douglass (1949) note their occurrence in Tojolabal, and Pinkerton (1986) provides instrumental data in various other Mayan languages. None of the Mayan languages contrasts voiceless implosives with voiced implosives. Contrasts of this kind have been documented only in Owerri Igbo (Ladefoged et al 1976). At the convention, Constance Kutch-Lojenga noted that voiceless implosives were used in a number of other languages spoken in different parts of Africa; she gave as one instance Lendu, a Nilo-Saharan language spoken in Zaire. There is, however, no instrumental evidence supporting this assertion; and to my ears the recordings of Lendu so-called voiceless implosives sounded like laryngealized stops, perhaps accompanied by a glottal closure, but without involving a descending larynx producing suction. Many members of the convention felt that voiceless implosives were sufficiently rare not to need a whole new row of symbols. There were proposals (which I favored) for symbolizing these sounds by a diacritic, such as the

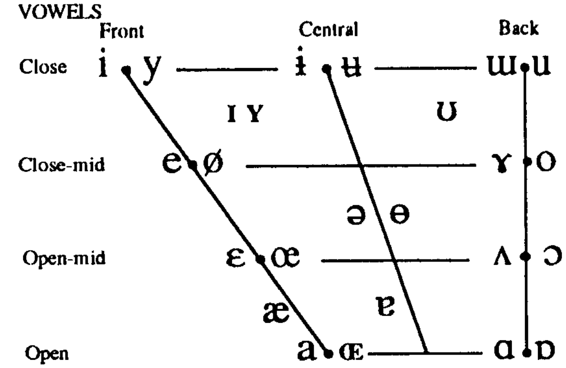
THE INTERNATIONAL PHONETIC ALPHABET (revised to 1989)

CONSONANTS

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			
Ejective stop	p'			t'		ʈ'	c'	k'	q'		
Implosive	ƥ ɓ			ƭ ɗ			ƈ ʄ	ƙ ɠ	ʠ ʛ		

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative
w Voiced labial-velar approximant
ɥ Voiced labial-palatal approximant
ʜ Voiceless epiglottal fricative
ʢ Voiced epiglottal plosive
ʡ Voiced epiglottal fricative
ɧ Simultaneous ʃ and x
ɜ Additional mid central vowel

ʘ Bilabial click
ǀ Dental click
ǃ (Post)alveolar click
ǂ Palatoalveolar click
ǁ Alveolar lateral click
ɺ Alveolar lateral flap
ɕ ʑ Alveolo-palatal fricatives

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

DIACRITICS

̥ Voiceless n̥ d̥	̹ More rounded ɔ̹	ʷ Labialized tʷ dʷ	̃ Nasalized ẽ
̬ Voiced s̬ t̬	̜ Less rounded ɔ̜	ʲ Palatalized tʲ dʲ	ⁿ Nasal release dⁿ
ʰ Aspirated tʰ dʰ	̟ Advanced u̟	ˠ Velarized tˠ dˠ	ˡ Lateral release dˡ
̤ Breathy voiced b̤ a̤	̠ Retracted i̠	ˤ Pharyngealized tˤ dˤ	̚ No audible release d̚
̰ Creaky voiced b̰ a̰	̈ Centralized ë	̴ Velarized or pharyngealized ɫ	
̼ Linguolabial t̼ d̼	̽ Mid centralized e̽	̝ Raised e̝ ɹ̝ (ɹ̝ = voiced alveolar fricative)	
̪ Dental t̪ d̪	̘ Advanced Tongue root e̘	̞ Lowered e̞ β̞ (β̞ = voiced bilabial approximant)	
̺ Apical t̺ d̺	̙ Retracted Tongue Root e̙	̩ Syllabic ɹ̩	̯ Non-syllabic e̯
̻ Laminal t̻ d̻	˞ Rhoticity ə˞		

SUPRASEGMENTALS

ˈ Primary stress ˌfoʊnəˈtɪʃən
ˌ Secondary stress
ː Long eː
ˑ Half-long eˑ
̆ Extra-short ĕ
. Syllable break ɹi.ækt
| Minor (foot) group
‖ Major (intonation) group
‿ Linking (absence of a break)
↗ Global rise
↘ Global fall

LEVEL TONES

̋ or ˥ Extra-high
́ ˦ High
̄ ˧ Mid
̀ ˨ Low
̏ ˩ Extra-low
↓ Downstep
↑ Upstep

CONTOUR TONES

̌ or ˩˥ rise
̂ ˥˩ fall
᷄ ˧˥ high rise
᷅ ˩˧ low rise
᷈ ˧˦˧ rise fall
etc.

Figure 1.

voiceless diacritic under the more usual voiced implosive symbol, e.g. [ɓ̥] for a voiceless bilabial implosive. But the symbols that now appear on the chart were voted in.

The Association's beliefs concerning the general nature of phonetic representation are more difficult to discern, because they reflect the views of a group of individuals who are, on the whole, more pragmatic than dogmatic. They certainly do not all hold to the same view of the relation between phonetics and phonology. Nevertheless they were all able to work together in a remarkable way. The convention was dominated by a delightful willingness to compromise that rested on a premise that virtually everybody shared, that it was more important to agree on and promulgate a standard set of symbols than it was to haggle over matters of phonetic or phonological theory.

There are explicit statements reflecting this pragmatic view in the revised set of Principles that the Association has adopted. They begin by noting that the IPA "is designed primarily to meet practical linguistic needs." The Association also decided that in exemplifying the use of the IPA: "Several styles of transcription should be illustrated, and the point should be made that these are all valid IPA transcriptions." In the new set of illustrations of the use of the IPA that the Association is preparing, the American English section notes that "Four different forms of transcription of the vowels are given in the list of Key words." Perhaps now that the Association has been explicit in its eclectic approach, outsiders to the Association will no longer speak of *the* IPA transcription of a given phenomenon, as if there were only one approved style.

The most interesting ambiguities in the Association's theoretical concepts concern what the symbols are symbolizing. The newly adopted *Principles* that form the basis of the IPA begin by saying that:

> "The IPA is intended to be a set of symbols for representing all the possible sounds of the world's languages."

This seems to regard a sound (or perhaps, in other terms, a phone) as the basic unit. But the next few sentences present a different point of view:

> "The representation of these sounds uses a set of phonetic categories which describe how each sound is made. These categories define classes of sounds that operate in phonological rules and historical sound changes. The symbols of the IPA are shorthand ways of indicating certain intersections of these categories. Thus [p] is a shorthand way of designating the intersection of the categories voiceless, bilabial, and plosive;

[m] is the intersection of the categories voiced, bilabial, and nasal; and so on."

These statements indicate that the symbols are *not* symbols for phones; they are simply shorthand for what a phonologist would regard as a bundle of features. The basic units are the phonetic categories which the symbols represent. The Principles go on to say:

"The sounds that are represented by the symbols are primarily those that serve to distinguish one word from another in a language."

This is clearly a phonological view, one that is further supported by the statement on the following page concerning the establishment of the set of symbols:

"When two sounds occurring in a given language are employed for distinguishing one word from another, they should whenever possible be represented by two distinct symbols without diacritics."

Both these sentences suggest that the Association's alphabet is intended to have a phonological basis, rather than what what might be called a purely phonetic approach. This is indeed the Association's long held view. In its first *Exposé des Principes,* published in 1900, it noted that its alphabet included "les *sons distinctifs* de toutes les langues étudiées jusqu'ici". (My emphasis.) Similarly, the 1912 English version, in a section headed "principles of transcription for languages hitherto not transcribed," notes, long before the phoneme became a popular notion: "It is necessary to ascertain what are the *distinctive* sounds in the language, i.e. those which if confused might conceivably alter the meanings of words." (Emphasis in the original.) The corresponding section in the 1922 *L'Écriture phonétique internationale* uses the then new term 'phoneme' saying: "Pour chaque langue, on représente les *phonèmes* ou sons distinctifs, et ceux-là seuls." (Emphasis in the original.)

The Association is deliberately not explicit about what is meant by a phonological contrast. Perhaps this is because of the British influence within it, which makes it behave somewhat like the Church of England – a body whose doctrine is so diffuse that one can hold almost any kind of religious belief and still claim to be a member of it. In this spirit, the Association in its Kiel Convention followed its long tradition, and worked towards establishing an alphabet that would be maximally useful to all, whatever their theoretical persuasion. Inevitably they did not entirely succeed. There was, for example, a request from Cathe Browman and Louis Goldstein (who were unable to be present at the meeting) that the Association should provide a standardized means of symbolizing articulatory gestures. They drew attention to a pre-publication version of their paper (Browman & Goldstein 1990), in which they noted:

"We have been trying different approaches to the question of what gestural symbols should be; our present best estimate is that gestures should be

> treated like archiphonemes. Thus our current proposal is to use the capitalized form of the IPA symbol for oral gestures, capitalized and diacritized {H} for glottal gestures, and {±N} for [velic close] and [velic open] gestures respectively. In order to distinguish gestural symbols from other symbols, we enclose them in curly brackets: { }. This approach should permit gestural descriptions to draw upon the full symbol resources of the IPA, rather than attempting to develop additional symbols."

This proposal, which seemed to me to be very sensible, fell on deaf ears, largely because most of those present had not yet been introduced to the notions of articulatory phonology, and did not understand the need.

A somewhat related proposal (of my own) was also rejected. I had wanted to include on the consonant chart some of the terms traditionally used for describing features, and perhaps also to include some of the notions of a feature hierarchy. With that in mind I suggested that there should be cover terms above the places of articulation, Labial above bilabial and labiodental, Coronal above dental, alveolar, post-alveolar, and retroflex, Dorsal above palatal, velar and uvular, and Radical above pharyngeal and (potentially) epiglottal. If this had gone through, I would have tried for an even more hierarchical organization. But it was not to be. Inherent conservatism triumphed (perhaps quite rightly) over something that is not yet fully established and agreed; I had, for example, considered Dorsal to include palatal on the grounds that what most IPA phoneticians call "palatal" is an articulation defined in the 1949 *Principles of the IPA* as having a tongue position similar to that in the cardinal vowel [i] in which it is the part of the tongue below the hard palate, and not the blade of the tongue that is the active articulator. But many phonologists consider palatal consonants to involve the blade of the tongue, and therefore to be Coronal (Keating 1988).

I have tried to argue so far that the Association views its set of symbols as having a phonological basis. But others might disagree. Thus Gösta Bruce, who was the coordinator of the suprasegmentals section, said in his preliminary report summarizing the views of those interested in this section: "Assuming with J.C. Catford (Ann Arbor) that the primary purpose of the IPA is to provide symbols and diacritics for the notation of primarily phonetic - and *not* [his emphasis] phonological - entities, the kind of suprasegmental notion we are aiming at will be diacritic symbols for suprasegmental categories added to a segmental, phonetic transcription." Nevertheless, Bruce then goes on to note that: "The proposed phonetic notation must, however, be related to the needs of phonology, so that phonological surface contrasts in the languages of the world can be symbolized." Furthermore he joins others in pointing out that "it is usually not possible or meaningful to create a transcription which is completely independent of some linguistic interpretation. This means that a suprasegmental transcription somehow has to be model-based and based on at least some knowledge of the language to be transcribed." (The above quotations are all from documents informally circulated at the Convention.)

Members of the Association are thus somewhat schizoid in their view of phonetic transcription. There is still a lingering spirit of the founding fathers such as Jespersen, and, only a little later, Jones, neither of whom recognized Trubetzkoy's distinction between phonetics, the continuous substance of the sounds that signify meanings, and phonology, the discrete, formal basis of the images that combine to form concepts. It was not that Jones and Jespersen did not acknowledge the distinction between Saussure's "langue" and "parole," which formed the basis for Trubetzkoy's view. It was rather that they preferred to look at the world in a different way. One of Jones's major works is called *An Outline of English Phonetics,* although it is concerned with what we would now call phonemics. The book describes both the set of items that can contrast (the phonemes) and the actual sounds (the allophones) that occur in particular contexts. Jones viewed phonetics as *containing* phonemics, making the phoneme a unit within the more encompassing notion of a phonetic description of a language.

Probably very few of the phoneticians at Kiel would still use the term "phonetics" in exactly the same way as Jones. But most of them would hold to something like his view of phonetic transcription. The contemporary view of possible types of transcription is still basically that of Abercrombie (1964). Much of what follows in this section is simply a paraphrase of this classic paper. Abercrombie's terms are summarized in Figure 2.

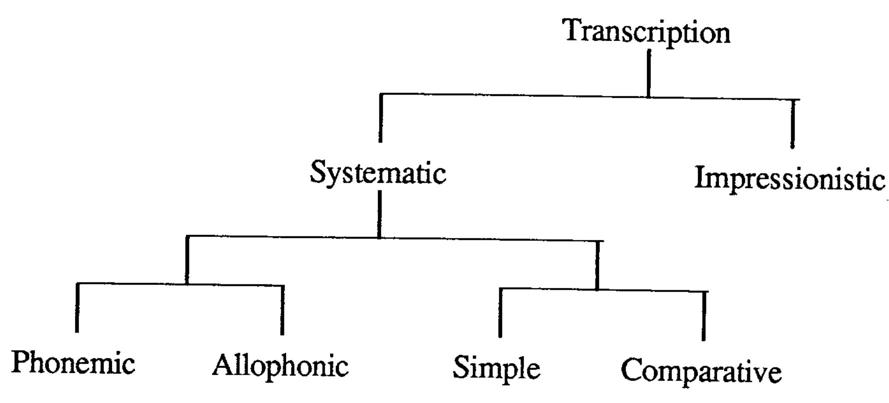


Figure 2. A schematic representation of the terms used by Abercrombie (1964) for different types of phonetic transcription.

The first distinction is between a transcription which in some way reflects the systematic, linguistic, facts of the utterance being described, as opposed to one in which the symbols are used just to provide an impressionistic record of the sounds as heard — the kind of record that might be made by a linguist totally without any preconceptions hearing the first few words in a language that had never been transcribed before. In theory, an impressionistic transcription is one in which the symbols represent intersections of general phonetic categories.

Few of us could ever make a totally impressionistic transcription. Generally, within a few minutes of starting to transcribe a language we have not worked on before, we begin to use symbols that rely on our linguistic hunches and preconceptions. We very soon stop noting small differences between repetitions of the same utterance, particularly if they are of the kind of which the speaker seems to be unaware. Virtually the only occasion when a completely impressionistic transcription is necessary is in the investigation of an infant's pre-linguistic babbling.

Within the class of systematic (phonetic) transcriptions there are two independent divisions. First a transcription may be phonemic or allophonic. There is little point in discussing here the meaning of the classic term phoneme, as current phonological theories are largely in a different domain. We may take it that a phonemic transcription is one in which, as Abercrombie (1964:19) says, "the smallest possible number of different letters [symbols] ... distinguish unambiguously all words of different sound in the language." An allophonic transcription is one that uses a larger number of distinct symbols, so that it can differentiate among systematic, allophonic, differences in the sounds of an utterance.

The IPA tradition for the last 60 or 70 years has been that phonemic and allophonic transcriptions are related to each other by a set of conventions (we would nowadays say "rules") that, by their application, convert the one form of transcription into the other. A phonemic transcription plus its conventions conveys exactly the same information as an allophonic transcription. The difference between these two types of transcription is simply whether the detailed phonetic information is made explicit within the transcription itself, or within the set of rules that accompanies the transcription.

The other kind of distinction among systematic phonetic transcriptions is that between a simple and a comparative use of particular symbols. The simplest IPA symbols are those that are ordinary letters of the roman alphabet. More exotic letters are often used to convey greater phonetic detail. As Abercrombie (1964:20) puts it: "romanic shapes have, by tradition, acquired a more *general* phonetic value than the exotic shapes which are alternatives to them." Again the IPA tradition allows phoneticians to make a simple transcription (which may, or may not, be phonemic), using the most general symbols possible in the circumstances, and giving an account of the values of the symbols in the set of conventions (rules) accompanying the text; alternatively, phoneticians may use more specific symbols and have less information in the accompanying set of conventions. The use of more specific symbols is often a convenient way of expressing part of what phonologists regard as phonetic interpretation conventions (Keating 1985). Abercrombie (1964) uses the term comparative for this type of transcription on the grounds that the use of more specific symbols implicitly reflects a comparison between the general phonetic value of the simple symbols, and the more precise value of the exotic symbols.

We should also note that if we follow the views outlined above concerning the nature of transcription (as many members of the IPA would do), then the term "systematic phonetic transcription" is a cover term for several types of transcription, including a simple phonemic transcription, and, by extension, current phonological representations. Once again it seems worthwhile emphasizing to phonologists that many phoneticians think of phonetics as, in some extended Jonesian sense, including phonology. Alternatively, this is simply a restatement of the view that phonology and phonetics are inextricably intertwined. It is true that phonology must rest on phonetic observations. But it is equally true that most phonetic observations are made in terms of a phonological framework. The only pure phonetic description is the instrumental data derivable from a high quality recording. As soon as the data is segmented or described in any way, then phonological considerations are bound to be present.

Next I will turn to another point that the IPA leaves unresolved. This is the question of whether there is a finite set of speech sounds, a matter that has been cogently discussed in a recent paper by Lindblom (1990). Most members of the Association would probably answer that there is such a set, and that it can be determined by observing the sounds to be found in the world's languages. They would thus reject Lindblom's notion that the study of what is a possible speech sound can only be pursued anthropophonically – starting from the study of what the human vocal organs are capable of producing (and the auditory system is capable of identifying). Of course there are prominent members of the Association, including J.C. Catford who gave us the term anthropophonics (Catford 1977), who clearly believe in a general phonetics divorced from the study of the sounds of languages. I do not know whether they believe that there is a finite set of sounds that could be used in languages, in the sense that they could be part of a linguistic system. But the majority of phoneticians probably would agree that a very substantial proportion of the possible sounds of the world's languages have now been recorded, and that, although there are doubtless a number of sounds as yet unobserved, this number is becoming smaller every year. As we working linguistic phoneticians know, it is becoming harder and harder to mine the phonetic dross and come up with something new. There seems to be no danger that "the accumulation of more data will eventually make the IPA and various D[istinctive] F[eature] frameworks unmanageable and ultimately bring about their collapse." (Lindblom, 1990).

It is nevertheless true that the description of a very large proportion of the sounds of languages (or, a description of the features that characterize these sounds) is no substitute for an explanation of why these sounds (or these features) should be used and others not. Lindblom is correct in emphasizing that an explanation of these facts, and of nearly all the observations of phonology, must come from outside linguistics. An explanation of something is an account of that event in terms of general principles that are not themselves dependent on the event. This is, of course, what large numbers of phoneticians have been trying to do for years, to explain phonology in terms of the general physiological principles of articulatory phonetics or the properties of the auditory system that permit us to identify speech sounds. But we are at the moment a long way from being able to show whether the set of possible speech sounds is finite or not, and whether it has a particular form.

Very much related to the problem of deciding whether there is a finite set of speech sounds is the problem of deciding whether a sound in one language is the same as a sound in another. Most phoneticians would agree that it is possible to make such decisions, but the basis for them is by no means obvious. In answering the question whether a sound in one language is the same as that in another, one cannot test whether the use of the one rather than the other would change the meaning of a word. The two sounds in question do not exist in the same language, so any subtle differences between them cannot be heard in the same linguistic context. The best that one can do is to ask whether, if the one sound were to be used in the other language, it would cause the user to be considered as a non-native, or deviant speaker. But this is not the same as asking whether the two sounds are different in the sense that, if they did occur in the same language could they be used to distinguish words. Take the case of dental versus interdental [θ], for example, which we know (Ladefoged 1979) are used consistently differently by British (RP) and American (Californian) speakers of English; 90% of RP speakers use a dental [θ] without tongue protrusion in initial position, for example in 'think' and 'thin', whereas 90% of Californians have an interdental [θ], with the tongue clearly protruding between the teeth in these words. Nevertheless the use of an interdental [θ] by a British RP speaker is not regarded as marked in any way. It would certainly not cause the speaker to be considered to be a foreigner. So these two variants can presumably be considered to be the same sound.

The difference between bilabial [ɸ] and labiodental [f] is non-contrastive in English, and in that respect has the same status as the difference between dental and interdental [θ] in English. Most speakers of English use labiodental [f], but use of bilabial [ɸ] is not regarded as marked for speakers of any kind of English. I habitually use [ɸ] and [β] in virtually all contexts in which others use [f] and [v]; but nobody has ever commented on it, even in phonetic circles discussing accents. But we have long known that bilabial and labiodental fricatives are phonologically distinct in many languages (e.g. Ewe; Westermann 1930), so we regard them as distinct speech sounds. Given these facts, why should dental and interdental [θ] not be regarded as distinct speech sounds, just because no known languages uses them contrastively?

Cases of differences that are often noticed as socially marked but are probably never distinct speech sounds within a language also occur. For example the Canadian vs. western American pronunciation of the vowel in ‘out’, or various shades of London vs RP pronunciations of the vowel in ‘say’, are clearly marked for those who know the speech of these groups. But these small variations in diphthongal quality are not known to be phonologically contrastive in any language. It may be that in order to be contrastive, a difference has to be above some auditory threshold. But for most speech sounds we do not know what this threshold is, nor why bilabial [ɸ] vs labiodental [f] is above it, noe whether dental vs interdental [θ] is not. There is no principled way in which we can determine whether two sounds in different languages are sufficiently similar to be considered to be the same sound.

We have already noted that the general atmosphere at the Kiel convention was one of pragmatism. We were there to make practical changes and not to argue theoretical points. Certainly issues of the philosophy of science were very far from our minds. But behind even the most mundane description there are philosophical assumptions about the nature of knowledge. Facts do not exist in isolation but only as part of an interpretive whole, as has been cogently argued recently by Rosaldo (1989). His view is that we can understand what we are observing only by recognizing that we are part of what we are observing. Nowhere is this truer than in linguistics. For us there is no absolute scientific reality. Once a language has been learned one is living in a room with a limited view. Even the greatest polyglot, who can think clearly in a dozen languages, has only a limited set of windows through which the world can be observed. There is no way in which one can answer questions such as which speech sounds are most alike, or what articulations are most difficult to make without being severely affected by one’s linguistic biases. A Navaho can make an ejective more easily than a dental fricative. A !Xóõ Bushman can distinguish over 80 different clicks, but would have problems with [v] and [w].

It may be theoretically possible to set up a procedure for measuring the degree of articulatory effort involved in producing a sound. One might, for example, claim that sounds that require a greater deviation from the neutral position required a greater degree of articulatory effort, as Lindblom (1990) has advocated. But procedures for making measurements of this kind are virtually impossible to put into practice. Even if we knew whether the neutral position of the speech organs was itself a language dependent notion, there is virtually no way in which we can assess whether a movement of the lips required more or less effort than a movement of the tongue; or even whether raising one part of the tongue is more difficult than lowering another. Theoretically we could assess the calories involved in producing each action, but this is beyond the scope of present day physiological techniques. Phonetic principles based on pseudo procedures of this sort are uninteresting, as they are not scientifically testable.

It is not even technically possible to devise a measure of auditory distinctiveness among speech sounds without becoming entangled in the problem of observer bias; which speech sounds are most distinct depends on the observer's linguistic background. Even skilled phoneticians will fail to recognize auditory distinctions to which they are completely unaccustomed. The nearest approach to an unbiased observer is a new born infant. Observations of the order in which sounds are learned by native speakers of different languages are of some relevance in investigating questions of articulatory ease and auditory distinctiveness. The fact that Bushman children learn to articulate click sounds later than many other consonants (Traill, p.c.) presumably shows that these sounds are in some sense harder to say. But this source of evidence has its own problems. The semantic weight and frequency of occurrence of the words containing these sounds will affect the rate of learning; and it is always difficult to tell whether a sound is learned later because it is more difficult to produce or more difficult to hear. In addition, even infants are not living in an unbiased phonetic environment. We cannot devise a universal hierarchy of articulatory difficulty or auditory similarity from scattered observations of very beginning language acquisition in discrete languages, none of which uses more than a small proportion of the total available sounds.

For the phonetician there is no universal truth independent of the observer. What we choose to represent in our phonetic transcriptions is a product of our biases, just as our whole view of language and society depends on our observational stance. Elsewhere (Ladefoged 1982) I have argued that language is like morality in that it is a property of an institution, not of an individual mind. I also suggested that, just as a moral code may be seen as the product of conflicting forces such as the pressure for individual liberty versus the prosperity of the society as a whole, so speech communication depends on the balance between the need for auditory distinctiveness and the desirability of articulatory economy. But when looking at morality cross-culturally we find that there is no universally applicable notion of individual liberty, nor of what is needed to make a society prosperous. And, just as there is no absolute morality valid for all cultures, so there is no linguistically useful notion of auditory distinctivesness or articulatory economy in absolute terms. There is no phonetic absolute valid for all languages. There are a myriad different events that might have been speech sounds needing phonetic representation.

This leaves us with three possibilities. We could either follow the earlier linguists such as Hockett (1955) and Joos (1950), who advocated ad hoc descriptions for each language. Hockett thought it "impossible to supply any general classificatory frame of reference from which terms can be drawn in a completely consistent way for the discussion of every individual language." The second possibility is that we could adopt the anthropophonic approach advocated by Lindblom (1990). Taken to its extreme, which Lindblom does not do, but which he logically should do, this would lead to considering as distinct speech sounds every possible articulation, including presumably every conceivable vowel at every conceivable pitch and at every conceivable loudness and with every conceivable voice quality. This does not seem an appropriate set of data to study if we are concerned with linguistic phonetics. Nor would the situation be sufficiently simplified if we restricted ourselves to studying the dimensions in what Lindblom (1990) terms a 'universal phonetic space.' There would still be a plethora of dimensions, including all the diverse activities of the different parts of the glottis, the epiglottis and root of the tongue, the gnashing of the teeth, the vibrations of the cheeks, and the twisting of the lips.

Many phoneticians (myself included) would agree with Lindblom's view that it is best to consider contrasting sounds as occupying ranges within each of the phonetic dimensions (parameters, features, scales, or however else the variables are named). Keating (1984) has suggested that certain ranges form major phonetic categories within a given dimension; and Stevens (1989 and references therein) has discussed why certain ranges are more favored by languages than others. All these views are in some form part of the stock in trade of most phoneticians. But none of them is likely to be able to answer the question: what is a possible speech sound?

The third possibility, listing the speech sounds that have been observed in all the world's languages, is all that is left to us It is a useful approach for those of us concerned with studying linguistic phonetics. If we take it that we now know virtually all the possible sounds, we can then start looking for reasons why this set of sounds is all that can be found in our circumstances. We must assume that there is some balance between the various functions such as articulatory ease, auditory distinctiveness and other factors that lead to languages being what they are. We can then look at the data and try to derive what these functions must be. It is the biased observed data that will lead us to the locally appropriate principles that determine the phonetic structure of languages. There is no way in which we can start from a set of principles that will lead us to delimiting speech sounds. It is not possible to begin by considering the nature of functions such as articulatory ease and auditory phonetic similarity in a language independent way. Of course there are a few valuable general constraints of articulatory and auditory physiology that will grossly limit the class of possible speech sounds; but these will not help us unravel the delicate locally woven fabric of individual languages, so that we can understand why things are the way that they are. All we can do is observe as many languages as possible, describe the data in terms of categories that seem appropriate for the situation, derive the principles that produce these data, and wonder at what else might have been.

Where, then, does this leave the IPA? Is it bound to be an unscientific enterprise? If one believes that there is such a thing as absolute scientific truth, the answer is yes. But this is not an appropriate view of science. Science, like beauty, is in the eye of the beholder; and the comment made by Keats (1820) on the possession of beauty is equally true of scientific knowledge:

"When old age shall this generation waste,
Thou shalt remain, in midst of other woe
Than ours, a friend to man, to whom thou say'st
'Beauty is truth, truth beauty,' – that is all
Ye know on earth, and all ye need to know."

Acknowledgements

I would like to acknowledge useful comments on this paper by Caroline Henton, Patricia Keating, Thegn Ladefoged, Mona Lindau and Ian Maddieson, even though they may not share my views on scientific knowledge, truth and beauty.

References

Abercrombie, David (1964) *English phonetic texts*. London: Faber and Faber.

Browman, Cathe & Goldstein, Louis (in press) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: Between the grammar and the* *physics of speech* (John Kingston and Mary Beckman, editors), Cambridge: Cambridge University Press.

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

Hockett, C. (1955) *A manual of Phonology.* Bloomington: Indiana University Publications in Anthropology and Linguistics.

International Phonetic Association (1990) Report on the 1989 Kiel Convention, *Journal of the International Phonetic Association*,**19**, 67-80.

Joos, M. (1950) Description of language design. *Journal of the Acoustical Society* **22**, 701.

Keating, Patricia. (1984) Phonetic and phonological representation of stop consonant voicing. *Language* **60**, 286-319.

Keats, John. (1820) Ode on a Grecian urn. *Annals of the fine arts.*

Keating, Patricia. (1985) Universal phonetics and the organization of grammars. In *Phonetic linguistics* (Victoria A. Fromkin, editor) pp. 115 - 132. Orlando: Academic press.

Keating, Patricia. (1988) *A survey of phonological features.* Bloomington: Indiana University Linguistics Club.

Ladefoged, Peter (1979) Review of "Fundamental problems in phonetics" by J. C. Catford. *Language,* **55,** 904-907.

Ladefoged, Peter (1982) 'Out of chaos comes order': Physical, biological, and structural patterns in phonetics. In *Proceedings of the tenth international congress of phonetic sciences* (M. P. R. Van den Broecke & A. Cohen, editors) pp. 83-96. Dordrecht-Cinnaminson: Foris Publications.

Ladefoged, Peter, Williamson, K. ,Elugbe, B., and Uwalaka, A. (1976) The stops of Owerri Igbo, *Studies in African Linguistics,* Supp. 6, 147-163.

Lindblom, Björn (1990). On the notion of "Possible speech sound", *Journal of Phonetics.* **18.**

Pinkerton, Sandra (1986) Quichean (Mayan) glottalized and nonglottalized stops: A phonetic study with implications for phonological universals. In *Experimental phonology* (John J. Ohala and Jeri J. Jaeger, editors) pp. 125-139. Orlando: Academic Press.

Rosaldo, Renalto (1989) *Culture and truth: The remaking of social analysis*. Beacon Press.

Stevens, K. N. (1989) On the quantal nature of speech. *Journal of Phonetics,* **17**, 3-46.

Supple, Julia, and Douglass, Celia (1949) Tojolabal (Mayan): phonemes and verb morphology. *International Journal of American Linguistics*. **15**. 168-174.

Traill, Anthony, personal communication.

Westermann, Diedrich H. (1930) *A study of the Ewe Language,* [trans. by A.L. Bickford Smith] Oxford: Oxford University Press.

Jaw Position in English and Swedish VCVs*

Patricia A. Keating
Björn Lindblom, University of Texas, Austin
James Lubker, University of Vermont
Jody Kreiman, VA Medical Center, West Los Angeles

0. Introduction

Lindblom (1983) proposed that consonant segments differ in their propensity to coarticulate jaw heights with neighboring segments: some consonants adopt the jaw height of their vowel nucleus, while other consonants do not. The latter consonants, he proposed, are those that tend to occupy syllabic positions that are further from vowels, with other consonants intervening between the incompatible consonants and vowels. Lindblom presented experimental data on the height of the jaw for different consonants before /a/. These data were taken to indicate the consonants' degree of coarticulation with the vowel. Keating (1983) took issue with this interpretation. She noted that different observed heights directly reflect degree of coarticulation only if all consonants have some identical inherent height very different from vowels' heights. She also pointed out that the position for the vowel was not reported, and could possibly be varying even more than the consonant positions. Keating cited data of Condax (1980) on Fijian, in which vowels appear to vary more than consonants, with overall variability of a segment seemingly correlated with its overall openness. She also presented new data on English consonant clusters suggesting that vowels and some consonants adopt jaw heights to accommodate other consonants such as /s/. Keating (1988) introduced the term *anchor* to refer to a segment which determines aspects of articulation over a segment span, to which neighboring segments accommodate themselves. Anchor segments by definition are less variable than non-anchors.

Lindblom viewed vowels as anchoring jaw height in a syllable, with consonants accommodating the vowels, while Keating viewed some consonants as anchoring jaw height, with vowels and other consonants accommodating. The issue then can be framed as one of relative variability: do consonants vary contextually more than vowels, or the reverse? The present experiment addresses this question. Data on jaw height were collected for combinations of consonants and vowels, and segments compared across contexts. The experiment also addresses another possibility: Lindblom and Keating studied different languages, which perhaps behave differently in the relevant respects. Therefore Swedish and English are both examined here.

1. Method

1.1. Speakers

The speakers were all visiting or working at Stockholm University. They included two of the authors and some others familiar with the purposes of the experiment. The Swedes were three men and two women, while the Americans were two men and three women. All speakers served as subjects only for their native language, but all of the Swedes knew English as well, and four of the Americans knew Swedish.

1.2. Speech materials

Most of the items used in the experiment were VCVs, uttered in isolation. The first vowel received weak stress and the second strong stress; the resulting pattern was like the usual pronunciation of English "Aha!". Broadly speaking, the vowels were /i/, /e/ (English) or /ɛ/ (Swedish), and /ɑ/, and the consonants were /f,b,t,d,s,n,l,r,k,h/ in both languages. However, speakers in each language used their own appropriate qualities of these phonemes for each context, resulting in phonetic differences across the languages, the phonetic contexts, and the individual speakers[1]. The most obvious differences in the consonants involved /r/ and /l/ across the languages. The American /r/ is an approximant, while the Swedish /r/ is most often an apical trill. The American /l/ is most often strongly velarized, while the Swedish /l/ is not. The vowels also differed somewhat across the two languages[2]. With front mid vowels it is difficult to match the phonetic qualities across the two languages without phonotactic violations. The Swedish high vowel used was [ɩ] in the unstressed initial syllable but [i:] in the final stressed syllable. The English mid vowel phoneme used was /e/, typically diphthongized to [eɩ] in both syllables. The Swedish mid vowel phoneme used was /ɛ:/ in the final stressed syllable, which occurs only as short [ɛ] (occasionally [e]-like) in the unstressed initial syllable. Throughout this report the mid front vowel of both languages will be represented simply as /e/. Since Swedish long vowels occur only in stressed syllables, there is noticeable variation in length, as well as quality, between the two vowel allophones in the Swedish VCVs. Particularly noticeable to Americans is the difference between initial [a] and final [ɔ:], [ɒ:], or [ɑ:]. Throughout this report the low vowel of both languages will be represented simply as /a/. The American vowels did not differ as much auditorily in initial compared to final position; the initial vowels, with weaker stress, were not audibly reduced or centralized.

A subset of these VCVs, the ones with vowel /a/, was also produced by a few speakers in a loud-voice condition, intended to induce more extreme jaw excursions and thus magnify any contextual variation for /a/.

Also included was a set of six VCCV utterances in which the two consonants formed a cluster: /isti, isli, ibli, asta, asla, abla/. These items allow the jaw positions for sequences of consonants to be compared with their component singleton consonants, both in terms of extreme positions and in terms of the time courses of segment interactions.

Finally, each subject (except for one American) also produced the three vowels in isolation, in long and short versions. For Swedish, it was necessary to get both phonemic lengths in isolation to compare with the two positional allophones in each VCV. For English, parallel durational differences were elicited, with shorter vowels checked by glottal stops, and longer vowels with a more typical citation pronunciation.

Each test condition except the isolated vowels included 6 repetitions of each item in a random order which was the same for all speakers. For the isolated vowels, the Swedes produced 4 repetitions and the Americans produced only 3.

1.3. Equipment

Data were collected using the Movetrack magnetometer system designed and built at the Phonetics Laboratory of Stockholm University (Branderud 1985). This system sets up a magnetic field around the subject's head and then tracks the position of a small magnetic coil within that field. The subject wears a helmet that holds in position two large parallel coils, one in front of the head and one behind, that set up the field. In this experiment the small coil was attached with soft wax at the midline on the gum ridge at the lower teeth. For most subjects, simultaneous audio and movement signals were recorded on FM tape. For all subjects these signals were also printed by an oscillomink.

1.4. Procedure

Each subject served during one experimental session. At the beginning of a session, the procedure was described, the speech materials gone over, and a reference position for the jaw found and practiced. This was described to subjects as "starting to swallow, letting your teeth touch, and holding the teeth in that position without completing the swallow", and all subjects seemed able to use this kind of description to find a repeatable reference position. Next, still outside the experimental room, the magnetic coil was attached in the mouth and the helmet put on. The subject then entered the room with the experimenters, sat in the chair and was hooked up to the apparatus. After the connections etc. had been tested, the experimental recording began. One experimenter remained in the room with the subject (either Lindblom or Keating, depending on the native language) and first triggered an automatic calibration device, whose output was recorded. Then the experimenter asked the subject to assume the clench reference position. Next, the experimenter read the speech items one at a time, and the subject repeated each one. The original motivation for this modeling technique was that the helmet worn by the subjects made reading from a list awkward; however, it soon became apparent that modeling offered the additional advantages of standardizing the prosody used by the speakers, and eliminating confusions about the phonetic transcription of the nonsense items. These advantages seem to outweigh the potential danger of artificial homogeneity in the data. The two models themselves also served as subjects, reading the by-then familiar test materials. Usually after every 10 items, the experimenter asked the subject to repeat the clench position.

After the 180 VCV items, the order of conditions was clusters (6 items, 6 repetitions each), isolated vowels (3 or 4 repetitions each of long and short vowels), and for 4 subjects, the loud condition. At the end of the session, the calibration was repeated.

For 7 of the 10 speakers, the jaw movement signal was printed out in two forms. One was taken directly from the output of the tape recorder, and the other was filtered. These are virtually identical except for a slight difference in gain. For two subjects only unfiltered data taken at the input to the tape recorder was printed., while for the remaining subject only the filtered data was printed. For the 7 speakers, a rectified and integrated audio signal was also printed, but for the other three, only the direct audio signal was printed. However, the printed audio signal was not used in the data analysis reported here.

1.5. Analysis

Figure 1 reproduces a sample token of a VCV item taken from the paper oscillomink print-outs made during each experimental session. From such tokens, the most extreme jaw position for each V and C was measured by the first author. Where two jaw movement signals were available (as in Figure 1), the filtered (lower) one was used. Measurements were made to the nearest half millimeter of paper, which after calibration usually corresponded to about a quarter millimeter of spatial position. For each of the two vowels in each token, the most open position was measured. For the consonant in each token, the most closed position was measured. That is, the measurement for vowels is a maximum opening, while the measurement for consonants is a minimum opening, each often corresponding to a turning point in the jaw movement trace. No special effort was made to locate the measurement point for each segment within the acoustic signal corresponding to that segment, and thus there may be errors of mistaking vowel high points for consonant positions.

The actual measurements made from the printout were distances from the jaw tracing to a reference line printed by the oscillomink (see Figure 1). The clench positions were also measured with respect to this line, in between every block of repetitions. Because of variations in the reference line and/or the clench position, this measure varied by up to 3 - 4 mm (more typically 1 mm) across a subject session. Therefore the average value of the clenches before and after each block was subtracted from the jaw measurements in that block, giving jaw positions with reference to this averaged clench. Finally, all measurements for a given subject were multiplied by the calibration factor for that subject, to convert millimeters paper to millimeters vertical jaw position. Thus all the jaw positions reported here are distances from a clench position (i.e. opening, not height), such that larger numbers represent lower jaw positions. We note that some other studies have *not* measured jaw position relative to an arbitrary reference position, as we have, but instead have measured the displacement from each segment to the next, so that each segment in effect serves as the reference for the next. Such a method makes it very difficult to say how adjacent segments influence each other's positions.

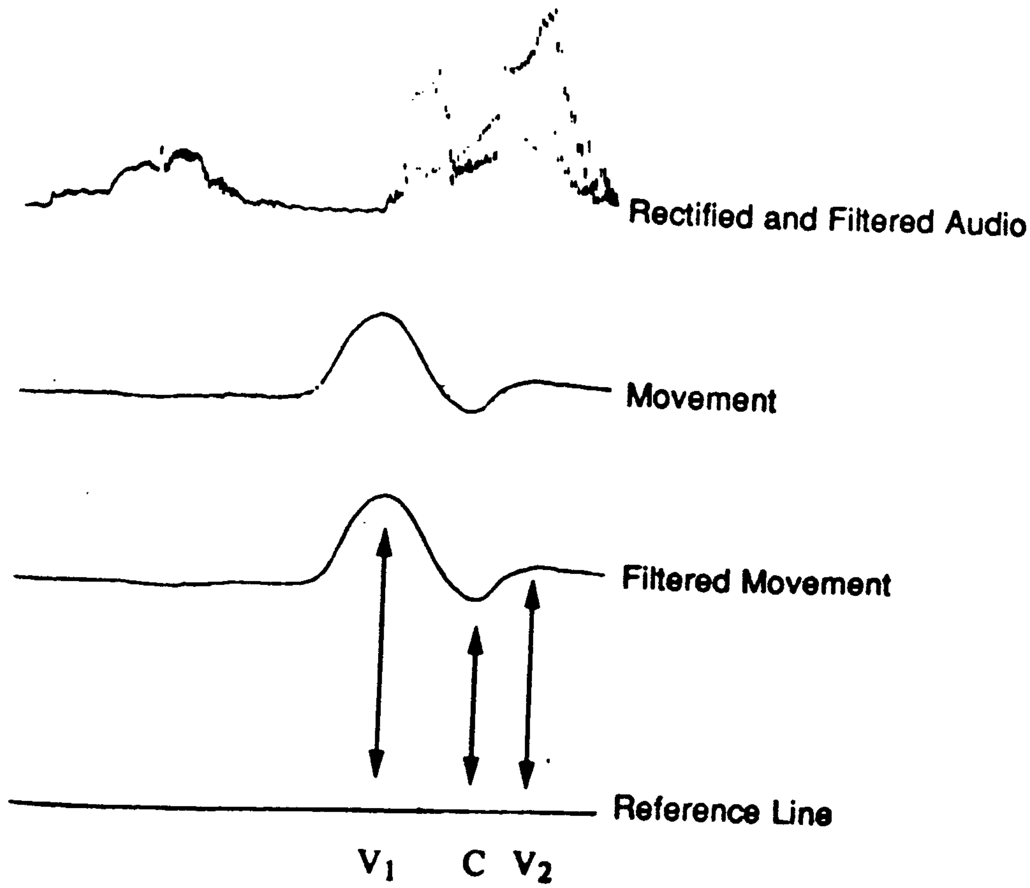


Figure 1. Example of printed raw data and measurements for a VCV token. See text for further description.

Measurements were entered into computer files so that statistical tests could be performed. Normalization of the data is described below. Analysis of Variance was used to determine which of the experimental conditions (language, segments, contexts) significantly affected measured jaw positions. Other statistical measures are described below.

2. Results

2.1. Isolated Vowels

Table 1 gives the mean maximum opening for each vowel for each speaker, and averaged for each language. E1 through E4 are four English speakers, E is the mean for English, S1 through S5 are the five Swedish speakers, and S is the mean for Swedish. Long and short vowels are distinguished. Most striking are the differences across the individual speakers, which are surely in part due to anatomical differences, including sex-related differences. For English, the shorter vowels tend to be less open than the long vowels, and the height difference between /e/ and /a/ is obscured when they are short. In Swedish, the long vowels tend to be more peripheral than the short ones. These data can also be compared with all the results below for vowels in context. The isolated vowels are generally slightly more open than normal-volume vowels in context, and somewhat less open than the loud vowels in context.

	i:/i	e:/e	a:/a
E1	5.6	9.0	10.5
E2	3.9/3.8	6.7/7.9	9.3/6.2
E3	4.7/4.8	13.8/10.0	15.6/9.4
E4	9.7	20.8	31.3
E	6.0/4.3	12.6/9.0	16.7/7.8
S1	7.1/6.7	9.6/10.1	9.2/10.4
S2	7.6/8.3	14.2/13.6	22.1/21.1
S3	7.6/8.6	12.3/9.5	13.2/13.4
S4	4.2/4.3	5.8/5.2	4.4/5.2
S5	5.2/5.6	8.7/7.3	10.8/14.3
S	6.3/6.7	10.1/9.1	11.9/12.9

Table 1. Mean jaw openings in mm (raw measures) for isolated vowels.

2.2. VCV Utterances

2.2.1. Normalization procedure

A repeated-measures ANOVA on the raw measurements (in mm) revealed a significant effect of subject on jaw opening (F (4,40) = 647.99, $p < 0.001$); additionally, all interaction terms involving subjects were significant. We therefore transformed the data by dividing each measurement for each subject by that subject's maximum observed jaw opening; i.e., all measurements were converted to a percentage of the maximum observed opening for a given subject. All measurements reported below are therefore percentages, not mm: for example, a measurement of .21 refers to a jaw opening that is 21% of the maximum opening for that subject. This normalization was intended to correct for confounding of subject size differences with the other factors in the experiment. Specifically, 1) the two language groups had different numbers of males, so group effects could be confounded with size differences among subjects; 2) since all measurements were made relative to a clench position, any size differences among subjects would not evenly affect all segments; i.e., higher segments (nearer the clench position) would vary less across subjects, regardless of size, while lower segments would vary more.

It should be noted that earlier work such as Lindblom (1983) and Keating (1983) did not use any normalizing procedure. Single-subject experiments, and analyses of position rather than of variability, do not require such normalization.

	Swedish			English		
ibi	.361	.314	.390	.389	.287	.395
ebe	.527	.408	.668	.676	.351	.630
aba	.708	.365	.617	.748	.328	.655
idi	.354	.263	.340	.334	.224	.320
ede	.486	.300	.575	.604	.246	.523
ada	.611	.243	.499	.632	.200	.534
iti	.353	.273	.360	.335	.187	.314
ete	.445	.273	.610	.569	.188	.507
ata	.592	.240	.495	.615	.161	.523
iki	.382	.375	.461	.351	.318	.421
eke	.519	.449	.671	.612	.392	.621
aka	.694	.475	.603	.671	.398	.611
ifi	.411	.325	.395	.404	.281	.378
efe	.545	.350	.643	.668	.282	.571
afa	.646	.268	.584	.720	.219	.553
isi	.316	.215	.360	.335	.164	.325
ese	.443	.220	.600	.571	.154	.508
asa	.557	.194	.488	.596	.160	.541
ihi	.371	.377	.378	.380	.377	.421
ehe	.595	.618	.663	.652	.538	.648
aha	.743	.636	.623	.665	.639	.664
ini	.394	.333	.445	.441	.307	.401
ene	.471	.351	.596	.608	.342	.519
ana	.673	.316	.556	.676	.318	.540
iri	.360	.314	.366	.384	.233	.375
ere	.512	.367	.652	.633	.272	.593
ara	.659	.305	.571	.660	.234	.630
ili	.411	.429	.451	.401	.298	.367
ele	.548	.519	.719	.654	.309	.558
ala	.684	.497	.668	.708	.342	.601

Table 2. Mean jaw openings (normalized to proportion of maximum opening) for VCVs

2.2.2. Analyses of Variance

(Normalized) jaw positions for the test syllables, averaged across subjects within a group, are shown in Table 2. Two analyses of variance were undertaken. The first included only (normalized) measurements on vowel segments, with repeated measures on subject, vowel identity (i, e, a), vowel position/stress (initial weak stress, final main stress), and consonant context (10 consonants); language (Swedish/English) was treated as a grouping factor. The second ANOVA included only measurements made during consonants, with repeated measures on subject, vowel context, and consonant identity, and a language grouping factor. (The position/stress factor did not apply to consonant measurements.)

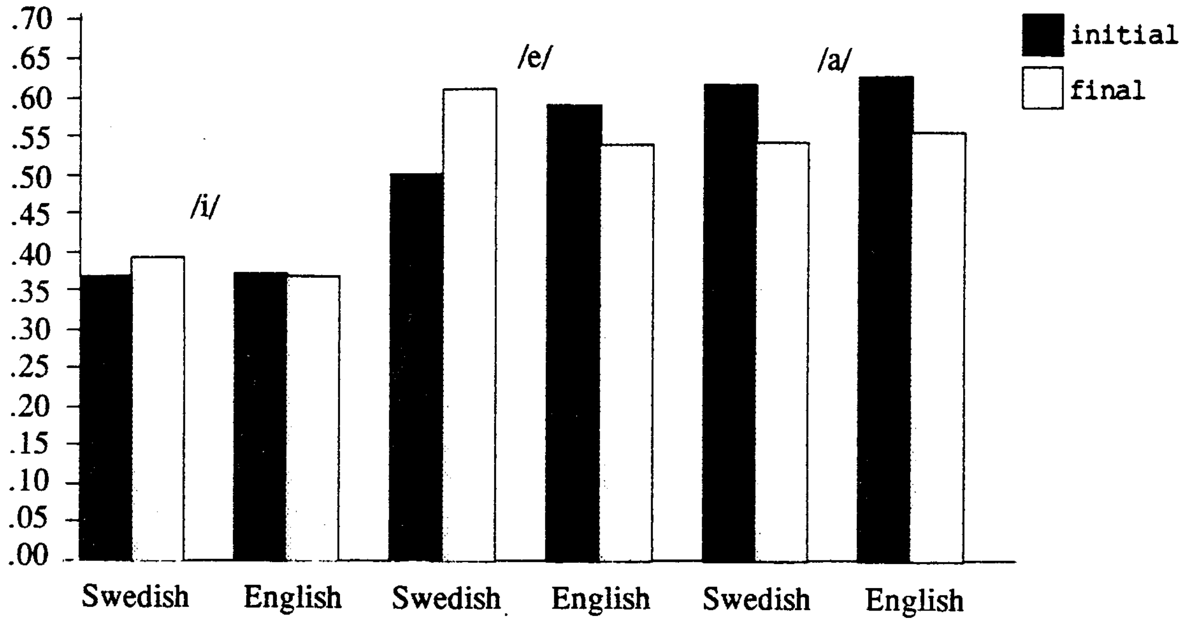


Figure 2. Mean jaw positions (normalized to proportion of maximum opening) for the vowels of English and of Swedish.

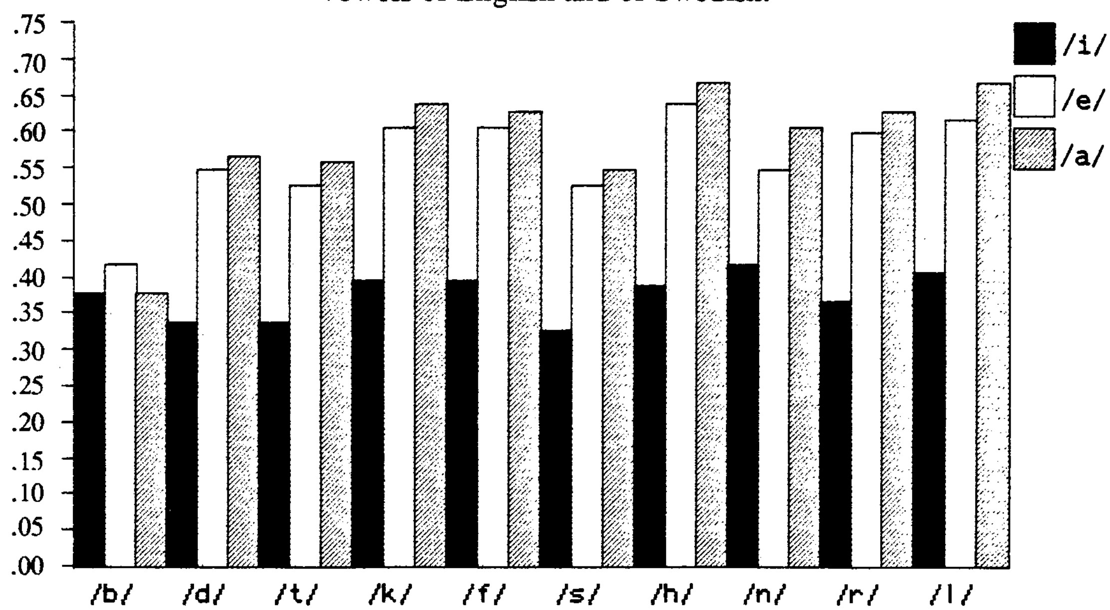


Figure 3. Mean jaw positions (normalized to proportion of maximum opening) for vowels according to consonant context; data from the two languages combined.

2.2.2.1. Vowel height measurements

Figure 2 displays jaw positions for the vowels of English and Swedish. Analysis of variance results are given in Table 3. On the average, English and Swedish speakers did not differ significantly in jaw position during vowels, although a significant position/stress x language interaction was observed: overall, Swedish vowels are more open finally, and English vowels are more open initially. When individual vowels are examined, we find the difference is largely due to differences in position during /e/ (Figure 2). The differences for /e/ may be linked to that vowel's allophones in Swedish: the large difference between initial (higher) [e] and final (lower) [ɛ:] is clearly audible. The two languages differ from each other very little in jaw position during /i/ and /a/. English and Swedish differ in how much variation in jaw position during vowels is contributed by each position: in English, initial vowels vary significantly more than final vowels (F (899, 899) = 1.428, $p < 0.01$), but no difference in variability by position was observed for Swedish (F (899, 899) = 1.059, n.s.).

Source	F	df	p
Language	2.94	1,10	n.s.
Subject	57.57	4,40	< 0.01
Subj x Lang	98.19	4,40	"
Consonant Context	76.77	9,90	"
Cons x Lang	4.76	9,90	"
Cons x Subj	7.53	36,360	"
Vowel Identity	85.79	2,20	"
Vow x Lang	3.42	2,20	"
Vow x Subj	75.54	8,80	"
Vow x Cons	3.39	18,180	"
Position/Stress	18.17	1,10	"
Pos x Lang	137.23	1,10	"
Pos x Subj	19.34	4,40	"
Pos x Cons	15.02	9,90	"
Pos x Vow	225.46	2,20	"

Table 3. Analysis of Variance results: Main effects and first order interactions for measurements made on vowel positions

All higher order interactions were also significant at the 0.01 level, except Cons x Vow X Lang (n.s.) and Cons x Vow x Pos x Lang (F (18,180) = 1.92, $p < 0.02$).

Consonant context has a highly significant effect on jaw position during vowels (Table 3), though the precise nature of the effect depends on the particular vowel/consonant combination. Figure 3 shows the interactions; data from Swedish and English are combined since the consonant x vowel x language interaction did not achieve statistical significance at even the 0.05 level. This figure shows that the jaw is highest during /i/ vowels for all consonant contexts except /b/; position tends to be quite similar for /e/ and /a/ across consonant contexts. The average jaw position during vowels is clearly affected by consonant context, however: the jaw is highest overall in VCV's containing bilabial and (to a lesser extent) alveolar obstruents, and lower for other segments.

This figure also shows how similar the three vowels are one to another in each consonant context. The similarity of the three vowels' positions can be quantified as a range, a simple and intuitive measure of variation. The range is the difference between highest and lowest values in a sample, in this case the mean positions of the highest and lowest vowels. Thus in the /b/ and /n/ contexts, the three vowels differ least in jaw positions, with their means spanning a range of 0.04 (with /b/) and 0.19 (with /n/); in the /r/, /l/, /h/ contexts the three vowels differ more in position,

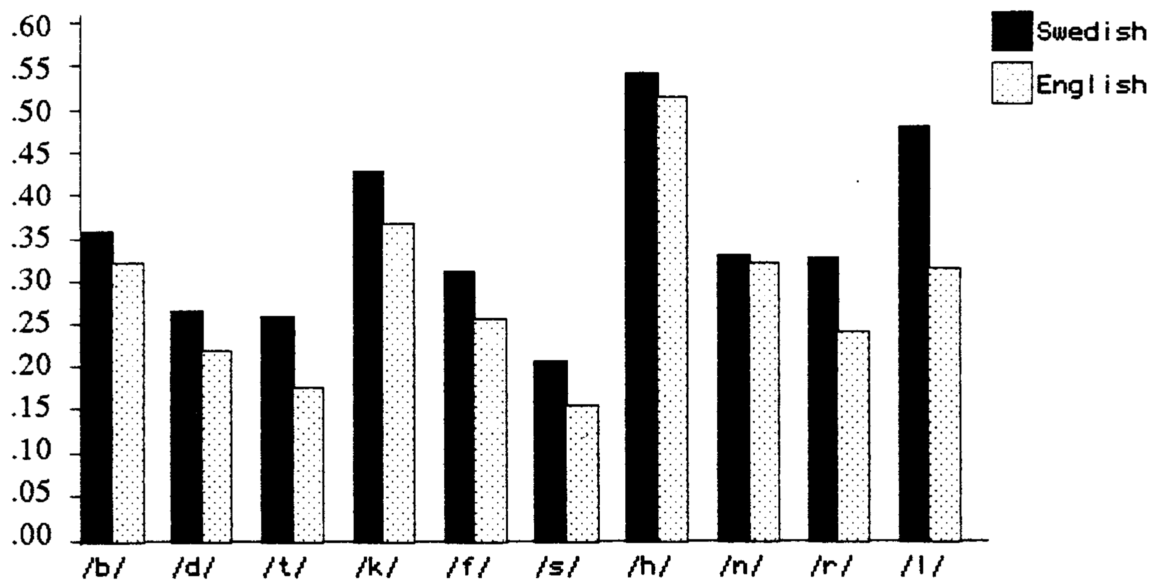


Figure 4. Mean jaw positions (normalized to proportion of maximum opening) for the consonants of English and of Swedish.

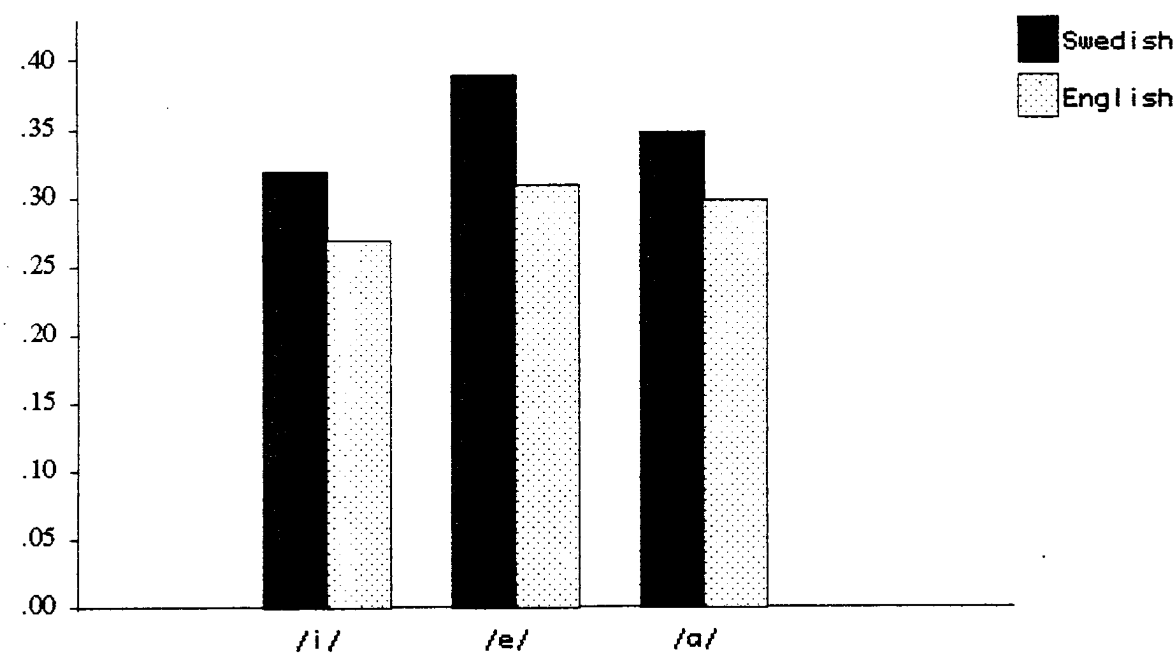


Figure 5. Mean jaw positions (normalized to proportion of maximum opening) for consonants according to vowel context in English and in Swedish.

with mean positions differing by 0.26, 0.26, and 0.29, respectively. The figure also allows us to compare the range of values covered by each vowel individually across the consonant contexts. The range of means for /i/ is 0.09 (the difference between the /s/ context, where /i/ is highest, and /n/ context, where /i/ is lowest), for /e/ is 0.22, and for /a/ is 0.29. Clearly in these data /i/ varies less as a function of consonant context than do the other two vowels.

2.2.2.2. Consonant height measures

Table 4 gives results of the ANOVA for jaw position during consonants. Overall, Swedish consonants are more open by about 6% (of the maximum observed position) than English consonants. Figure 4 shows the average jaw position for each consonant in each language; the pattern is similar for all cases except /n/ (where the two languages do not differ) and /l/ (where they differ quite substantially). The language factor is still significant when /l/ is removed from the analysis (F (1,8) = 107.85, $p < 0.001$) indicating that the observed differences are not due to this segment alone. Mean jaw position is given for each consonant segment in each language in Table 5; the data are rank-ordered within language from highest to lowest. In both languages the alveolar obstruents, /r/, and /f/ are in the higher half, while /b/, /n/, /l/, /k/, and /h/ are in the lower half (although details of the rankings differ somewhat). These rankings are highly correlated across the two languages, with a Spearman correlation coefficient of 0.912.

Source	F	df	p
Language	163.72	1,9	< 0.01
Subject	63.21	4,40	"
Subj x Lang	178.94	4,40	"
Consonant Identity	616.11	9,90	"
Cons x Lang	27.08	9,90	"
Subj x Cons	26.24	36,360	"
Vowel Context	77.23	2,20	"
Vow x Lang	5.42	2,20	ns/p < 0.02
Subj x Vow	28.56	8,80	p < 0.01
Cons x Vow	47.26	18,180	"

Table 4. Analysis of variance results: Main effects and first order interactions for measure-ments on consonant positions. (Higher order interactions were all significant at the 0.01 level except Consonant x Vowel x Language, p < 0.03.)

Swedish		English	
s	0.210	s	0.159
t	0.262	t	0.179
d	0.269	d	0.223
f	0.314	r	0.246
r	0.329	f	0.261
n	0.333	l	0.317
b	0.362	n,b	0.322
k	0.433		
l	0.482	k	0.369
h	0.544	h	0.518

Table 5. Mean (normalized) jaw position, rank ordered highest position to lowest

The significant consonant x vowel interaction (Table 4) indicates that consonant height depends in part on vowel context. The significant vowel context x language interaction is due to the fact that vowel context affects consonant position more in Swedish than in English (Figure 5). Correspondingly, jaw position during consonants varies more for Swedish than for English (F (899, 899) = 1.26, $p < 0.01$).

Figure 6 shows the interaction of consonant identity and vowel context for the two languages combined (since the consonant x vowel x language interaction failed to reach significance). Contrary to intuition, consonants are not consistently highest in /i/ contexts and lowest in /a/ contexts. For example, /d/, /t/, /f/, and /s/ are highest in the /a/ context, while /k/, /h/, and maybe /l/ are lowest in the /a/ context, on the average.

The data from Figure 6 are replotted in Figure 7 to compare the variation of individual consonants across vowel contexts. The exact effect of vocalic context on consonant position depends on the particular combination of segments: some consonants vary more than others; some are lower in /e/ contexts than in /a/ contexts, and vice versa. Overall, however, consonants differ one from another least in /i/ contexts and most in /a/ contexts: positions vary by an average of 19% in /i/ contexts, by 39% in /e/ contexts, and by 46% in /a/ contexts. The consonants that are consistently highest in this plot (compared to other consonants) are highest in the /a/ context, and vary relatively little across vowel contexts. The consonants whose positions are very low in the /a/ context vary much more across vowel contexts, and are fairly high in /i/ contexts.

2.2.3. Consonant versus vowel variability

Having presented the data on jaw positions for consonants and vowels, we turn now to consideration of the relative variability in position of consonants vs. vowels: Is one class of segments significantly more variable than the other?

2.2.3.1. Variances

The traditional measure of the variability of a data set--the standard deviation -- is given by the following formula:

$$s = \sqrt{\frac{\Sigma\ (x - \bar{x})^2}{n - 1}}$$

where x is a data value, $\bar{x}$ is the mean, and n is the number of data values in the data set. It thus measures the average departure of data values from the mean. Since the mean is an integral part of this measure of variability, the measure varies with the mean: given two data sets with different means but equal dispersion about the mean, the set with the higher mean will have a higher standard deviation. In the present case, vowels have lower mean jaw positions than consonants; correspondingly, simple comparisons of the variability of jaw position for consonants and vowels show vowels as more variable overall. Individual vowels in both languages are more variable than individual consonants, with most consonants varying over an average range of 56% across contexts, and vowels over an average range of 90%. The greater variability of vowels than consonants is seen even when vowels are limited to one or the other of their two position/stress conditions. A statistical test of relative variability can be made with a ratio of variances, submitted to an F-test. In both languages consonants as a group vary less than either initial or final vowels as groups (Swedish: initial vowels, F (899, 899) = 1.214, $p < 0.01$); final vowels, F (899, 899) = 1.285, $p < 0.01$. English: initial vowels, F (899, 899) = 1.727, $p < 0.01$; final vowels, F (899, 899) = 1.21, $p < 0.01$). This result is particularly striking in view of the fact that there are more consonants than vowels in the data: we might have expected more segments in a group to result in more variation within that group.

However, we might equally well expect more segments in a group to cause more contextual variation in the other group. Each vowel occurred in many more consonant contexts than each

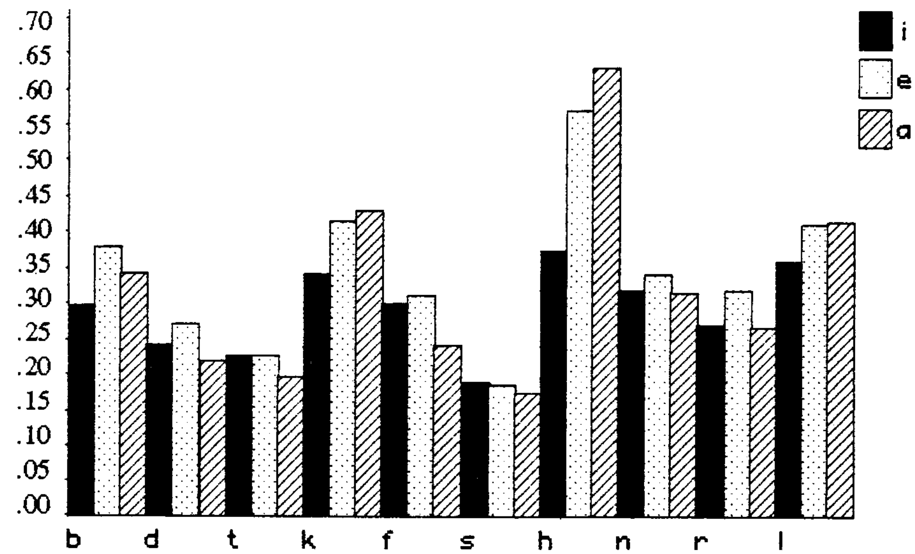


Figure 6. Mean jaw postions (normalized to proportion of maximum opening) for consonants according to vowel context; data from the two languages combined.

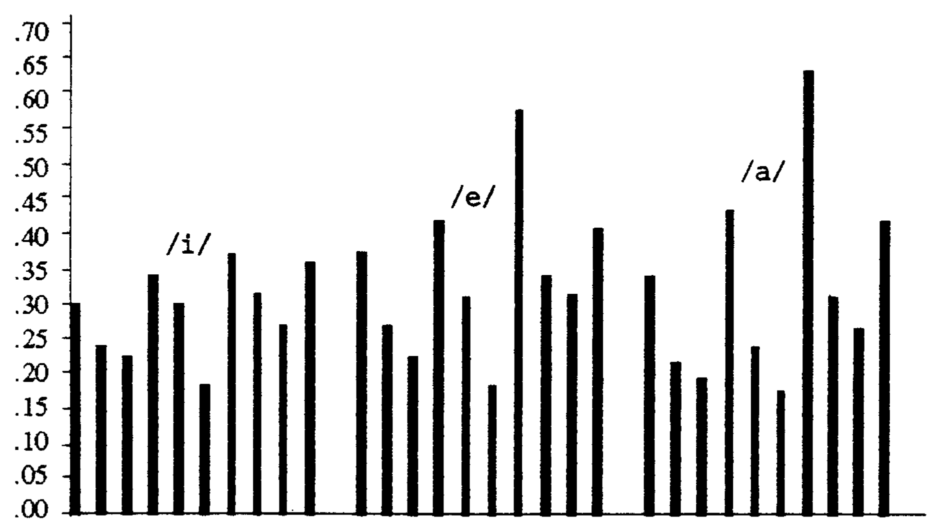


Figure 7. Same data as in Figure 6, replotted to compare variation of individual consonants across vowel contexts.

consonant occurred in vowel contexts; it might be proposed that the greater number of contexts accounts for the greater vowel variability. However, if we select sets of three consonants, and compute vowel and consonant standard deviations for just these six segments taken together, then the vowels are still usually more variable than the consonants. For example, if the three consonants are /s, t, d/, then the standard deviation for the vowels is .113, while the standard deviation for the consonants is .082. If the three consonants are /k, n, b/, then the vowel and consonant standard deviations are .134 and .109, respectively. In these cases, the vowels vary more, but the consonant sets have either high or low overall jaw positions. When a mixed group of consonants, /l, s, k/, is selected, the consonants vary more: vowel and consonant standard deviations are .132 and .159, respectively. In sum, then, comparison of consonant and vowel variability by means of standard deviations shows that jaw positions are usually more variable for vowels than for consonants.

2.2.3.2. Coefficients of variability

However, this difference in variability between consonants and vowels may well be an artifact of the differences in mean positions between consonants and vowels: by the argument given above, the segment class with the more open mean position would have a higher standard deviation (and thus variance) simply as an artifact of the way standard deviations are calculated. We want to ascertain that there is some difference in variability beyond the fact (which we already knew) that vowels are lower than consonants overall. We therefore calculated the coefficient of variability for each segment and used these values to compare the variability of consonants and vowels. The coefficient of variability is the standard deviation s divided by the mean $\bar{x}$. It in effect corrects the sample variability around the mean for the value of the mean, though it is still proportional to the variance. In our situation, where jaw height, but not jaw openness, is physically limited, use of this coefficient also partly corrects for "ceiling" effects.

Figure 8, which plots coefficients of variability against (normalized) openings, shows that for both consonants and vowels, segments with higher average positions vary more across contexts than do segments with lower average positions. Thus high vowel /i/ is more affected by consonant context than are the lower vowels, and the high alveolar obstruents in general vary more than other consonants. Also, low consonants and high vowels, which are most similar in height, are also most similar in variability. As suggested by the figure, mean height is negatively correlated with overall variability as measured by the coefficient of variability $s/\bar{x}$ (Spearman's rho = -0.62, $p < 0.01$).

Thus use of coefficients of variability reverses the overall pattern found by examining standard deviations: once overall means are corrected for, higher segments are seen to be generally more variable. It is illustrative to reconsider the comparison of vowels with subsets of consonants given above in terms of standard deviations. Examination of the coefficients of variability for different subsets of three consonants shows that the consonants are always more variable than the vowels, though some subsets are more variable than others. If, for example, the three consonants are /k,s,l/, then the coefficient of variability for the vowels is 0.254, while the consonant coefficient of variability is 0.484: the consonants are twice as variable as the vowels. If the consonants are /b,k,n/, then the vowel coefficient of variability is 0.249, while the consonant coefficient of variability is 0.306: the consonants are only somewhat more variable than the vowels. And if the consonants are /d,t,s/, then the vowel coefficient of variability is 0.238, while the consonant coefficient of variability is 0.380: an intermediate level of consonant variability.

2.2.4. Loud Condition

For this condition, data were gathered from one Swede and three Americans, who produced six repetitions each for the 10 consonants in the /a_a/ context (Swedish /a_a:/). Results are shown in Figure 9. An ANOVA on the measurements made during vowels showed significant main effects of subject (F (3,15) = 232.25, $p < 0.01$), consonant context (F (9.45) = 289.43, $p < 0.01$),

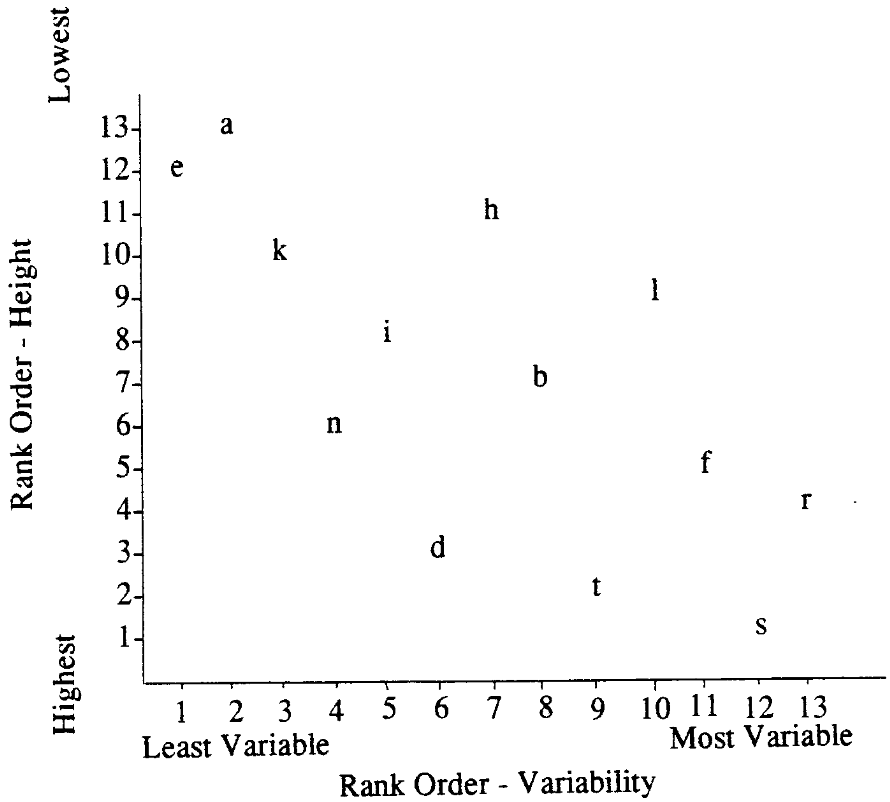


Figure 8. Coefficients of variability against (normalized) openings for consonants and vowels; data from the two languages combined.

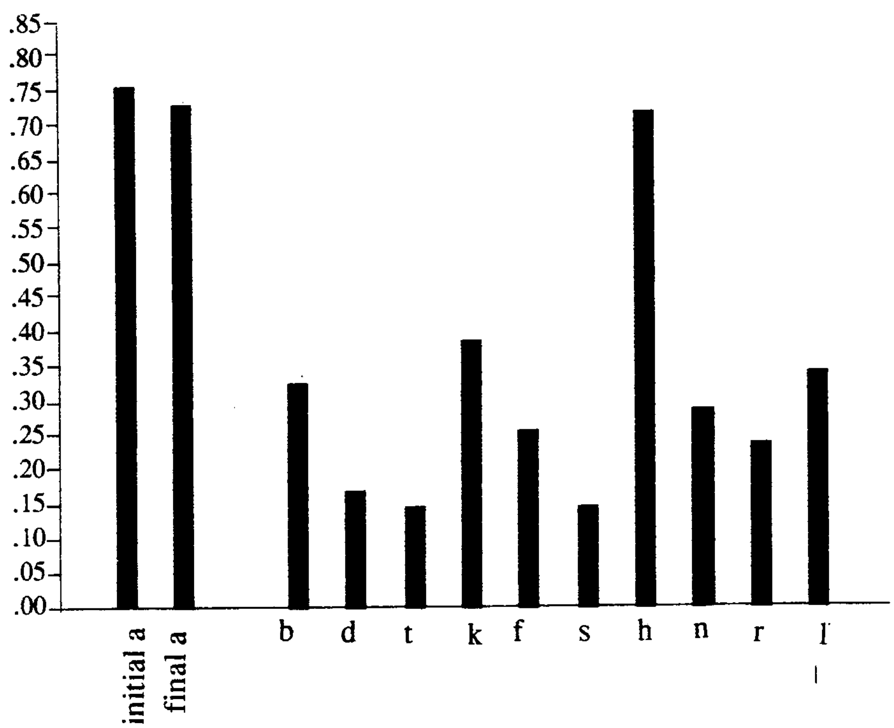


Figure 9. Mean jaw positions (normalized to proportion of maximum opening) for consonants in "loud" condition.

and position/stress (F (1.5) = 220.10, p <0.01) on jaw position during loud /a/; all interaction terms were also significant (Table 6). These findings parallel those for non-loud vowels (Table 3), except that the present ANOVA had no vowel identity term (since only /a/ was used) and no language term (since only one Swede was recorded, all speakers were combined in the present analysis). Figure 9 further suggests that, while loud segments are more open overall than non-loud segments, no other differences in patterns of jaw position occur (see also Schulman 1989).

Source	df	F	p
Subject	3,15	232.25	< 0.001
Consonant context	9,45	289.43	"
Subj x cons cont	27,135	250.27	"
Position/stress	1,5	220.10	"
Subj x pos	3,15	387.95	"
Cons x pos	9,45	328.02	"
Subj x cons x pos	27,135	344.69	"

Table 6. Analysis of Variance results: loud VCV items

2.3. VCCV Utterances

Subjects also produced six VCCV utterances in which /t/ and /l/ enter into consonant clusters with /s/ and /b/: isti, isli, ibli, asta, asla, abla. The single highest jaw position associated with the consonant cluster was measured for these items as for the VCV items, and similarly calibrated and normalized. Figure 10 shows these data, compared to results for the same, singleton, consonants. It is clear that the highest jaw position for the cluster generally represents a compromise between the positions of either of its two members alone. It appears that the highest jaw position for the cluster is more similar to the non-cluster position of the consonant with greater degree of constriction, than to the consonant with lesser degree of constriction (/b/ and /s/ over /l/, and /t/ over /s/).

2.4. Summary of results

This study compared jaw positions in two languages, English and Swedish. Several language differences were found. For vowels, the two languages were quite similar; however, the allophonic variation in Swedish /e:/ contributed to a language difference, and English showed a difference in variability of initial vs. final position that Swedish did not show. The two languages do not differ in how consonant context affects vowel position, and the rank orderings of consonants by height are significantly correlated in the two languages. However, for consonants, Swedish jaw positions are overall more open (even after normalization), and the rank orderings of consonants by height do differ slightly in the two languages. Vowel context affects consonants more overall in Swedish, and consonant positions are more variable in Swedish, but no other differences in how vowel context affects consonants were found. Overall, then, these are relatively minor differences--the two languages are generally quite similar re jaw position.

In most consonant contexts, as in isolation, /i/ is the highest vowel, and the positions for /e/ and /a/ are rather similar to one another. Overall, positions for vowels are higher when the intervocalic consonant is a bilabial or alveolar obstruent, and the three vowels have their most similar positions in high-consonant contexts.

Consonants are not consistently highest between /i/s and lowest between /a/s. Some consonants -- those that are overall highest -- are actually highest between /a/s, and some are lowest between /e/s. Generally, however, consonants are most similar in position between /i/s, and least similar between /a/s.

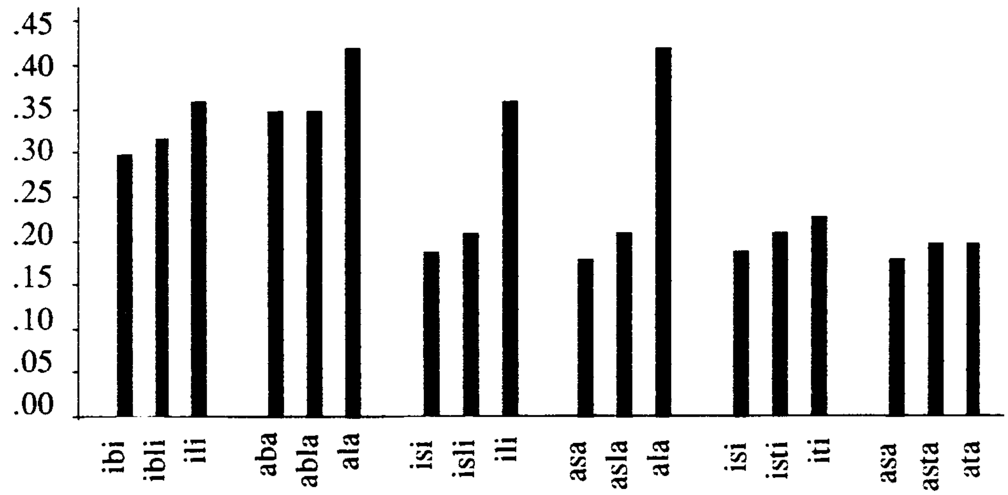


Figure 10. Mean highest jaw positions (normalized to proportion of maximum opening) of singleton consonants and consonants in clusters.

This study was also concerned to compare the behavior of vowels and consonants. When variability is measured by the coefficient of variability, which corrects for the value of the mean for each dataset, then higher segments are seen to be more variable than lower segments, both within and across the classes of consonants and vowels.

Other findings were that while loud segments are overall more open, they pattern otherwise like normal-volume segments, and that the maximum jaw opening for consonant clusters is a compromise between the positions for the consonants as seen in VCV tokens.

3. Discussion

3.1. Comparison with the literature on jaw positions

Other researchers have reported data on the relative jaw positions for consonants and/or vowels. Perkell (1969) examined several consonants before /ɛ/ for one speaker of English, and found alveolars to have higher maximum displacements than /p/ or /k/. Similarly, Kim (1972) found more displacement from /a/ to /t/ than from /a/ to labial stops in /aCa/ tokens (that is, presumably /t/ was higher than labials, though positions per se are not reported). The consistently high jaw position of /s/ was pointed out also by Amerman et al. (1970). Perkell also examined several vowels after /t/, and found /u/ the highest (due to lip rounding), followed by other high vowels, then /ɛ/, /a/, and /æ/ lowest of all. These results are all consistent with the present data, to the extent that the same segments are examined. On the other hand, contrary to our own earlier findings (Lindblom 1983, Keating 1983), with five speakers of each language and three vowel contexts, we now find /r/ in both languages to have relatively higher positions than reported earlier. The most general finding common to all these studies is that alveolars, especially alveolar obstruents, tend to have high jaw positions. This is not surprising, given that the tongue must be positioned so as to form a constriction directing airflow at the teeth. In our data, /f/ also has a fairly high jaw position, presumably due to the requirement of touching the lower lip to the upper teeth.

Data have also been previously reported on the effects of context on jaw position. Some earlier studies bear on the question of whether some segments are more variable than others. Inspection of Perkell's figures indicates that /ɛ/ varied greatly across the different preceding-consonant environments, while /t/ maintained a fairly consistent position across following-vowel contexts. Kiritani et al (1983) compared positions for /p,t,k,s/ in different vowel contexts, and found that only /p/ and /k/ varied with context. Imagawa, Kiritani, Masaki, & Shirai (1985) found that vowels have higher positions in CVCs than they do in isolation, and that high vowels /i/ and /u/ vary much less across contexts than do non-high vowels /e/ and /a/. These results are similar to our own when ranges and standard deviations are considered. However, it must be borne in mind that such results confound overall positions with variability, and that analysis with the coefficient of variability can reverse such findings.

Other studies on the effects of phonetic context bear on the question of which segments do or do not induce variability in other segments. Abbs, Netsell, & Hixon (1972) found that in CVC, /æ/ was lower when surrounded by /p/ or /k/ than by /t/. Sussman, MacNeilage, & Hanson (1973) combined three vowels and three bilabial stops in VCV utterances, and measured changes in jaw position from segment to segment (i.e. not absolute position, as we did). Movement between consonant and vowel was greatest with /m/ and least with /p/. Imagawa et al. (1985) found that vowels were higher adjacent to /s/ or /t/ than to /p/ or /k/. They also found that vowels /a/ and /e/ were more variable in frames which also contained /a/ vowels than in frames which also contained /i/ vowels. This finding was due to a result also found in our study -- some consonants, such as /s/ and /t/, are actually higher in /a/ contexts than in /i/ contexts. In Imagawa et al.'s test items, the higher /s/ or /t/ carried over into higher /a/ or /e/, resulting in greater overall variability of /a/ and /e/ in /a/ contexts. The fact that high-position consonants are higher in low-position contexts than in high-position contexts seems paradoxical, but Imagawa et al. relate it to the high velocity used to travel from a low to a high position. In effect, the high velocity caused the jaw to overshoot the usual high position for /t/ or /s/. The same may be true of our data, though we did not measure velocity to test this possibility.

3.2. Lindblom and Keating

The disagreement between Lindblom (1983) and Keating (1983) centers around the issue of whether consonants or vowels are more variable. If consonants are more variable, it makes sense to view consonants as accommodating to contextual vowels, whereas if vowels are more variable, then the accommodation is in the reverse direction. Previous discussion of the available data failed to take into account the overall differences in mean jaw position between consonants and vowels, and thus confounded overall position with variability. Under this confounded analysis, vowels are more variable than consonants in both English and Swedish. However, when mean position is corrected for, the clear pattern emerges that higher segments are more variable than lower segments, and thus consonants are more variable than vowels. This result supports Lindblom's proposal that consonants accommodate their jaw positions to those of neighboring vowels.

Footnotes

* This report presents an analysis of data collected by the first three authors in May 1984 at Stockholm University. We gratefully acknowledge the assistance of Peter Branderud, the designer of the Movetrack system, who helped us gather the present data. This report will also appear in PERILUS, Stockholm University's working paper in phonetics.

1. One American did, on request, substitute an [ɑ] quality for his more natural final [ɔ].
2. In Lindblom's notation, the Swedish vowels were [I] and [i:], [ɛ] and [ɛ:], [a] and [ɑ:].

References

Abbs, J. H., Netsell, R., and Hixon, T. J. (1972). "Variations in mandibular displacement, velocity and acceleration as a function of phonetic context." *Journal of the Acoustical Society of America*, **51**, 89.

Amerman, J. D.,Daniloff, R. and Moll, K. L., (1970). "Lip and jaw coarticulation for the phoneme /æ/." *Journal of Speech and Hearing Research*, **13**, 147-161.

Branderud, P. (1985). "Movetrack--A movement tracking system." PERILUS IV, Stockholm University, pp. 20-29.

Condax, I. D. (1980). "Correlation of mandible position and vowel durations." Journal of the Acoustical Society of America, **67**, Supplement 1, S93 (Abstract).

Imagawa, H., Kiritani, S., Masaki, S., & Shirai, K. (1985). "Contextual variation in the jaw position for the vowels in /CVC/ utterances." *Ann. Bull. Rilp*, **19**, 7-19.

Keating, P. A., (1983). "Comments on the jaw and syllable structure." *Journal of Phonetics*, **11**, 401-406.

------------------, (1988). "The phonology-phonetics interface." In F. J. Newmeyer, ed, *Linguistics: The Cambridge Survey. Vol. 1. Linguistic Theory: foundations*. Cambridge: Cambridge University Press.

Kim, J. (1972). "Studies of segmental aarticulatory dynamics. Doctoral Dissertation, University of Texas.

Kiritani, S., Tanka, S. and T., Hashimoto, K., Masaki, S., & Shirai, K. (1983). "Contextual variation of the jaw movement for the intervocalic consonant in VCV utterances." *Ann. Bull. Rilp*, **17**, 45-53.

Lindblom, B. (1983). *Economy of Speech Gestures. The Production of Speech.* (P.F. MacNeilage, ed..) pp. 217-246. New York: Springer-Verlag.

Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study.* Cambridge, Mass.: M.I.T. Press.

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America*, **85**, 295-312.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). "Labial and mandibular dynamics during the production of bilabial consonants: preliminary observations." *Journal of Speech and Hearing Research*, **16**, 397-420.

Some of the acoustic characteristics of Xhosa clicks

Bonny E. Sands

Paper presented at the 118th Meeting of the Acoustiocal Society of America, St. Louis, Mo.

Xhosa is a Bantu language spoken in South Africa. The consonant inventory of this language includes fifteen phonemic clicks. Clicks are typologically unusual sounds only found in languages of Southern Africa and parts of East Africa. Clicks are also unusual in that they are made with two closures, one of them being a velar closure. This is represented in the click transcription with a consonant symbol preceding the click. Clicks are made with a velaric ingressive airstream. In Xhosa, each click is made with a primary articulation, which may be dental, alveolo-palatal, or alveolar lateral. Each of these three click types can occur with one of five different accompaniments. Clicks can be voiceless, aspirated, nasalized, breathy voiced, and nasalized breathy voiced. The set of clicks and the symbols used for their transcription are laid out in Table 1.

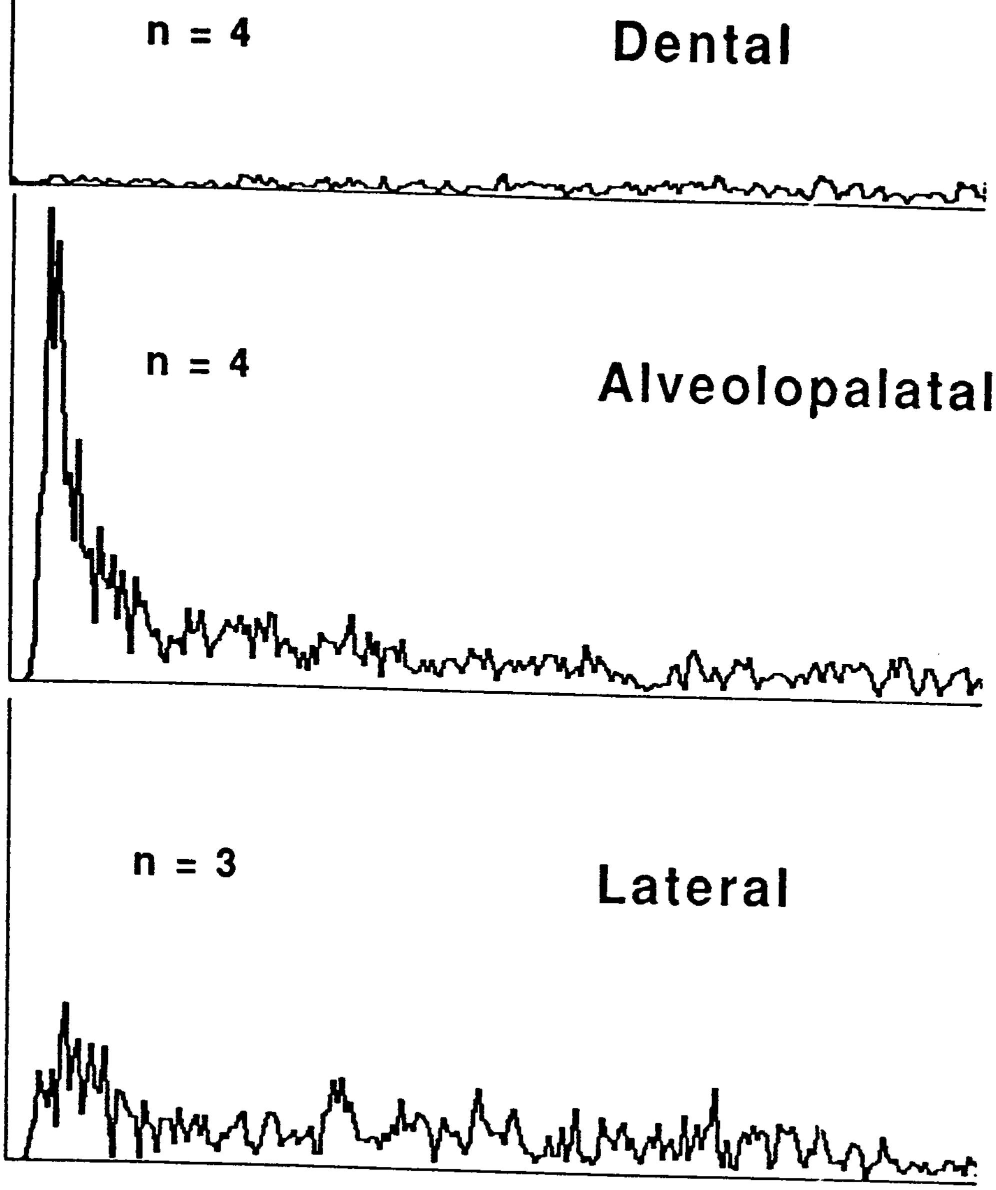
Table 1. Symbols for Xhosa clicks

	Voiceless	Aspirated	Breathy Voiced	Nasalized	Nasalised Breathy V.
Dental	k|	k|h	g|ɦ	ŋ |	ŋ |ɦ
Alveolo-palatal	k!	k!h	g!ɦ	ŋ !	ŋ !ɦ
Lateral	k||	k||h	g||ɦ	ŋ ||	ŋ |ɦ

Little is known about the acoustics of clicks, the clicks of Xhosa being no exception. This study hopes to provide a starting point in answering some of the questions, including:

•Are some clicks affricates?

Figure 1
Rectified integrated waveforms of the dental, alveolopalatal and lateral clicks taken from 3 or 4 speakers. The waveform is from 100 msec from release of the click.



•Are the spectra of click bursts similar to those of pulmonic sounds made at corresponding places of articulation, such as dental, alveolo-palatal, and alveloar lateral. In other words, are the cues for place of articulation similar in this respect for both velaric and pulmonic consonants?

•Do clicks show coarticulation effects of neighboring vowels, as is seen with non-click consonants.

The answers to these questions will help determine the natural class groupings of the clicks.

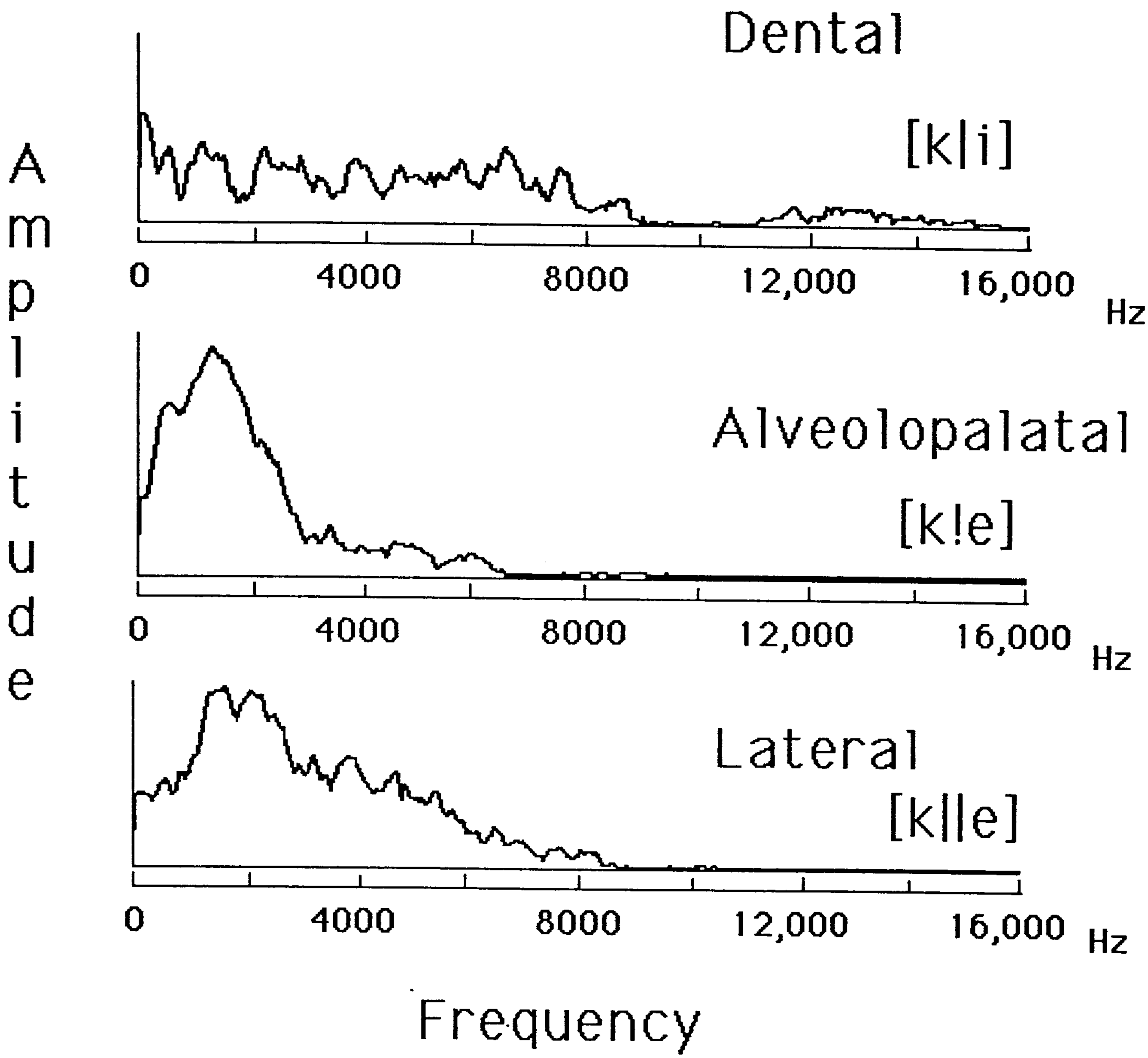
The data analyzed in this study were taken from a recording of four speakers saying words containing each of the 15 phonemic clicks before each of the five vowels of Xhosa, /i,e,a,o,u/. The DSP Sona-Graph was used to produce wideband spectrograms, waveforms and spectra of these utterances. Waveforms and spectra were averaged using a Macintosh computer.

Rectified integrated waveforms showing the mean of the 4 speakers over a span of 100 milliseconds starting just before the release of the plain voiceless clicks are shown in Figure 1. Since some of the VOT values were less than 80 milliseconds, a portion of the vowel is included in the averages in this display. Hence only the left hand portion of this display should be considered truly indicative of click type. The three types of clicks can be distinguished by characteristic patterns in these waveforms.

The dental clicks are made with an affricated release. They are made with the tip of the tongue placed against the lower front teeth with the blade of the tongue being sucked away from the upper front teeth.The dental burst is not characterized by a sharp onset, but the release has an extended period of friction following the burst. The amplitude of the burst and friction is typically low relative to the amplitude of the following vowel.

The alveolo-palatal clicks are produced with the tip of the tongue at a post-alvelolar or retroflex place of articulation and have a release with a very sharp onset, high amplitude and little affrication. The alveolo-palatal clicks bursts are high in amplitude relative to the following vowel. These clicks are characterised by

Figure 2
Spectra of the voiceless clicks before the vowels /i,e,a/. The spectra represent the means of 4 speakers each producing one token of each vowel.



a sharp initial spike which typically lasts fewer than 5 msec. After this spike the amplitude decreases quickly.

The lateral type of click is produced with the tip of the tongue on the alveolar ridge and the sides of the tongue against the side teeth. The release is affricated, being made by lowering one or both sides of the tongue. The click noise is similar in duration to the dental click, but is louder. The lateral click consists of an initial section characterized by strong amplitude, followed by a period of lessened amplitude. The lateral click is characterized by a comparatively sharp onset, but there is no initial spike of energy as in the alveolo-palatal click, and amplitude remains at roughly the same level for the first part. It seems that the lateral and the dental clicks are made with a long constriction release, which is released more gradually, causing the release to be affricated.

In addition to being distinguished in the temporal domain, clicks can also be distinguished in the spectral domain. Spectra of a 25 msec window centered at the release of the stop closure were analyzed. Averages of such power spectra for four speakers for clicks before the vowels /i, e, and a/ are shown in Figure 2. The click types vary in both range of frequency and in amplitude.

The power spectra of the dental click bursts are characterized by energy in a wide range of frequency which can be considered to consist of two main frequency bands. The first band is in the range of roughly 0 to 9000 Hz and is characterized by a relatively steady amplitude level, lower than that of either the lateral or the alveolo-palatal clicks. The second band ranges from roughly 10,000 Hz to above 16,000 Hz. The amplitudes of noise in this band are roughly one-third the amplitude of noise in the first band. The dental click can therefore be described as having a diffuse spectrum.

The power spectra of the alveolo-palatal click bursts are very compact, they are characterized primarily by one main band of energy. This main band ranges from roughly 0 to 3000 Hz and consists of a high amplitude peak of energy centered around 1000 Hz. Energy may continue up to 9000 Hz or beyond, but it is typically of very low amplitude relative to the first band.

The power spectra of the lateral click bursts are compact. Although less compact than the alveolo-palatal click bursts, they are not as diffuse in frequency range as are the dental click spectra. The lateral click has significant energy in a wide frequency band which ranges from about 0 to 6000 Hz with a strong prominence from about 1500 to 3000 Hz.

One can investigate the question of coarticulation by looking for differences in the power spectra before different vowels. Power spectra of the voiceless dental clicks from one speaker are shown in Figure 3. There are no apparent differences for the power spectra of the clicks before the vowels /i, e, a/. In particular, no effect of the high front vowel /i/ is seen. This is the vowel which commonly causes extensive coarticulation effects with other consonants. Preceding the rounded vowels, the dental clicks show a concentration of energy in the lower spectral region resulting from attenuation of amplitudes in the higher frequency range, due to the anticipation of lip rounding. The effect of rounded vowels on clicks is the only coarticulation effect seen.

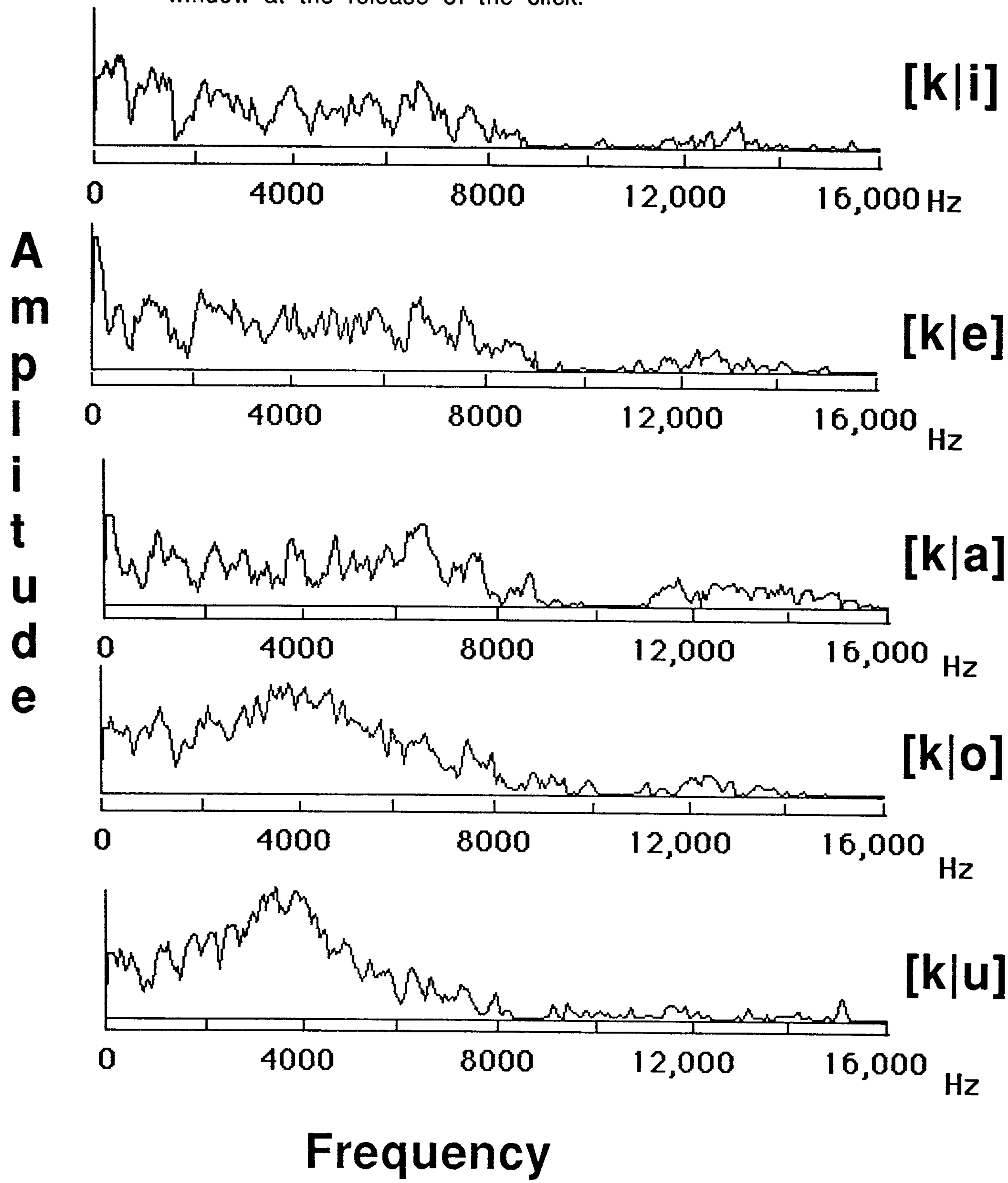
In summary, the click types, or places of articulation can be distinguished by information contained in the spectra of their bursts and in their temporal patterns. The dental clicks have spectra which show a diffuse frequency bandwidth and waveforms which show affrication. The alveolopalatal clicks have very compact spectra and a burst which has a sharp onset and high amplitude noise. The lateral clicks have compact spectral shapes. They have a highly affricated release.

To answer the questions posed earlier in this paper, it has been clearly shown that some of the clicks are affricates. The question of exactly how clicks correspond to other consonants has been left to further research, although it would seem that the click burst spectra are not similar to pulmonic stops made at these primary places of articulation, namely dental, alveolo-palatal, and lateral. Finally, it has been shown that coarticulatory relations between clicks and vowels seem less extensive than for between other consonants and their following vowels.

Acknowledgements

This work was supported by NSF Grant: BNS-8418580 (Peter Ladefoged, Principal Investigator). The project would not have been possible without the assistance of Dr. Rosalie Finlayson and Professor J. A. Louw of the Department of

Figure 3
Spectra of one speaker for the voiceless dental clicks before each of the five vowels. Spectra are taken from a 25 msec window at the release of the click.



African Languages, University of South Africa. The recording was supervised and made available to UCLA by Dr. Finlayson. Professor Louw suggested the word list and provided helpful comments on Xhosa phonology.

Bibliography

Beach, D.M. (1938). *The Phonetics of the Hottentot Language.* Cambridge: W. Heffer & Sons Ltd.

Kagaya, Ryohei. (1978). "Sound spectrographic analysis of Naron clicks: A preliminary report". *Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, Annual Bulletin* 12. 113-125.

Kohler, Oswin, Peter Ladefoged, Jan Snyman, Anthony Traill & Rainer Vossen. (1988). "The symbols for clicks". *Journal of the International Phonetic Association* 18: 140-142.

Ladefoged, Peter and Anthony Traill. (1980). "Phonological features and phonetic details of Khoisan languages". *Bushman and Hottentot Linguistic Studies 1979*, ed. by Jan W. Snyman. Pretoria: University of South Africa.

______ and Anthony Traill. (1984). "Linguistic phonetic description of clicks". *Language* 60.1: 1-20.

Louw, J.A. (1977). "Clicks as loans in Xhosa". *Bushman and Hottentot Linguistic Studies 1975 (papers of seminar held on 25 October 1975).* Department of Bantu Languages. ed. by J.W. Snyman. Pretoria: University of South Africa.

______with A.S. Davey, and P.C. Taljaard (compilers). (1976. rev. issue 1980). *Xhosa I. Guide 1.* Pretoria: University of South Africa.

Traill, Anthony. (1985). *Phonetic and Phonological Studies of !Xóõ Bushman.* (Quellen zur Khoisan-Forschung, 5.) Hamburg: Helmut Buske.

Wilkes, A. (compiler, revised ed. by Miss B.P. Putu, Mr. Z. Jama). (1987). *Course Xho100, Block A, Language Description: Phonetics.* Department of African Languages, Vista University.

Updating UPSID

Ian Maddieson and Kristin Precoda

Paper presented at the 118th Meeting of the Acoustical Society of America

St Louis, November 27-December 1 1989

The purpose of this paper is to introduce to an audience of potential users a "new and improved" database for studying the structure of phonological systems, and a special-purpose program for accessing and modifying it. The database is known by the acronym UPSID—standing for the UCLA Phonological Segment Inventory Database. An earlier version was described and analyzed in the book *Patterns of Sounds* (Maddieson, 1984).

The expanded and corrected second version of the database is now in the final stages of preparation and will be available early in 1990. This version improves the sample, both in terms of the selection of languages included and in terms of the accuracy of the data. There is increased coverage of previously undersampled language families and a few oversampling errors have been corrected. The expanded database will contain data on (approximately) 460 languages. The entire set of sources for the data has been re-examined and additional sources consulted, resulting in some errors in individual language inventories being corrected. A new custom-written software package for DOS systems provides an economical and flexible means of storing and examining this enhanced database and outputting subsets of the data for further analysis.

The database is stored in four modules. One module contains a list of the names of the languages in the database and their identification numbers. Identification numbers are four digits long. For each language, the first digit indicates the continental landmass on which the majority of its family is located, the second digit indicates the family affiliation and the final two digits identify the individual language. UPSID is designed to enable statistically valid generalizations to be made about the frequency of occurrence and co-occurrence of particular segment types in the world's living languages. Hence, the languages included are selected to provide a valid statistical sample. For various reasons the appropriate sample is a quota sample, based on genetic relationship. The procedure adopted is to target for inclusion one and only one language from each group of closely related languages. The total number of included languages amounts to perhaps about 6% of the world's languages. Over 90% of the quota design is satisfied in the new database.

The second database module contains character codes for each distinct segment type occurring in the database. These may be thought of as equivalent to phonetic symbols, but use only standard ASCII characters to maximize compatibility. Each code is paired with a standard phonetic description in words and a description in terms of a full set of feature values assigned to that segment. For the sake of compactness, the feature description is stored as a hexadecimal number. The features specify roughly the categories of standard international phonetic theory; they remain largely the same as those used in *Patterns of Sounds*. Note that this means that they are considerably more redundant than the feature sets used by phonologists. For example, rounding of vowels and labialization of consonants are indicated by presence of separate features "rounded" and "labialized".

The third module contains records which each consist of a language number and a character code, so each record represents a segment in a particular language. The list of segments given for each language reflects a level of analysis that is similar to the set of "lexical segments" posited in some recent phonological theories. The fourth module contains a list of the features available for describing segments in UPSID. We will see a partial listing of these in figure 5 below. This module pairs fully spelled out feature names with abbreviated names used in output files, and organises the features into groups concerned with place of articulation, phonetion type, manner of articulation, etc.

The program comes with two front ends. One variant - UPSID - allows a user to modify the data in any reasonable way, the other variant - PHONEME - prevents users from making changes. This version is more suitable for locations where many people have access to the system where it is installed. An overall view of the structure of the database package is given in figure 1.

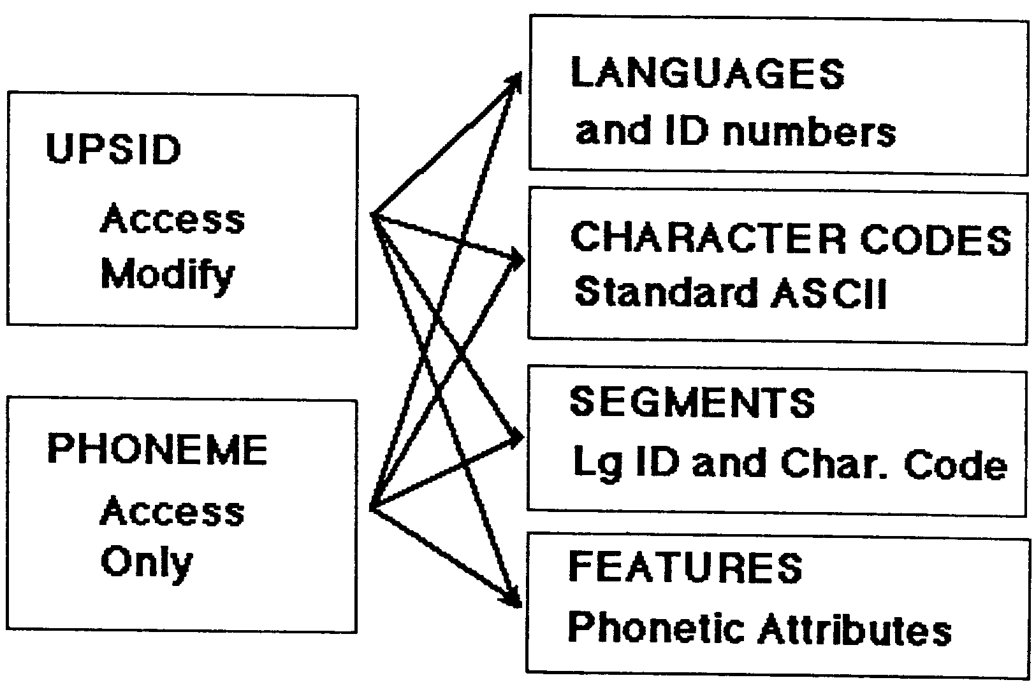


Figure 1. Outline of UPSID program package design.

The main program menu has sections headed Language, Character code, and Feature, and options to select or count subsets of the data, as shown below. In the UPSID version one can also add to, delete from or edit the data. For example, new languages, character codes or features can be added. We will not discuss the details of these functions in this presentation, but they are fully described in documentation available with the program.

Main Menu

Language - **display segments**
- **general information**

Character Code

Feature

Select Subset of Database

Miscellaneous counts

The program functions mainly by mating information from the various modules. One basic way of using the program is to display the segments of an individual language. We can enter either the language name or the number. For example, if we request Hawaiian, the display shown in figure 2 will appear. The program finds the language name from the languages file, looks up the number and locates the segments in the segments file with that language number. It then displays the character codes for the segments and the total of segments for the language.

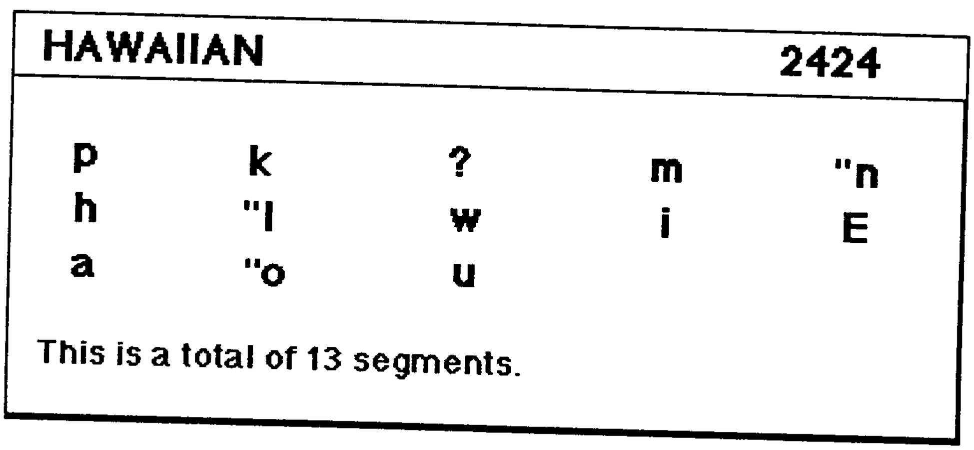


Figure 2. Language display for Hawaiian

Many of the common character codes are reminiscent of the standard IPA symbols. A nonstandard usage illustrated here is the use of quotes before some letters, which indicates, for coronal consonants, uncertainty as to whether these are dental or alveolar, and, for mid vowels, no indication that they are higher mid or lower mid. Several other nonstandard but consistent symbolizations are used in UPSID to create character codes for the approximately 800 distinct segments which need to be distinguished.

Each language also has a text file of general information associated with its name and ID number. This can also be viewed from the main menu. For example, if we request the information file for the language Iai we would obtain the display in figure 3. The general information files contain information on the genetic classification of the language and complete references for the sources used. They often contain comments on interpretation of the sources and other useful notes.

Language Name: Iai
Language Number: 2422
Alternate Names: Iaai
Classification: Austro-Tai, Austronesian, E. Malayo-Polynesian
Comments: Spoken in New Caledonia

Sources:

Ozanne-Rivierre, F. 1976. Le Iaai. SELAF, Paris

Tryon, D. T. 1968. Iai Grammar (Pacific Linguistics, Series B, 8). Australian National University, Canberra.

Haudricourt, A.G. 1971. New Caledonia and the Loyalty Islands. In Current Trends in Linguistics (T. Sebeok, ed.) 8: 359-396. Mouton, The Hague.

Figure 3. Language Information file for Iai.

Selecting the Character code section and entering a code, for example, **t.** (t followed by period), enables one to view its properties in the database, that is, its phonetic description and its assigned features. If you didn't know, you would learn this was the code for a voiceless retroflex plosive, as shown in figure 4. The features are assigned from a list of 60. The full listing, with a checkmark by the features selected for the particular character code, may also be viewed. A partial listing is given in figure 5. Such a display would normally only be viewed when adding to or editing the Character code module.

Character code: t.   **Number of occurrences: 26**
Phonetic description: voiceless retroflex plosive
Summary of features currently set to 1:
plos ret vl

Figure 4. Information on a character code.

Type <spacebar> to toggle a feature value, <enter> to exit with changes
Character code t.

√ plosive	implosive
ejective stop	click
fricative	ejective fricative
affricate	ejective affricate
affricated click	unspecified r-sound
tap	flap
trill	approximant
nasal	simple vowel
diphthong	lateral
sibilant	bilabial
labiodental	linguolabial
dental	unspecified dental or alveolar
alveolar	palato-alveolar
√ retroflex	palatal
velar	uvular
pharyngeal	glottal
labialized	palatalized
velarized	pharyngealized
nasalized	nasal release
high	higher mid
mid	lower mid
low	front
central	back
nonperipheral	rounded
unrounded	lip-compressed
r-colored	backing
lowering	rounding
√ voiceless	voiced
aspirated	laryngealized
long	breathy
overshort	preaspirated

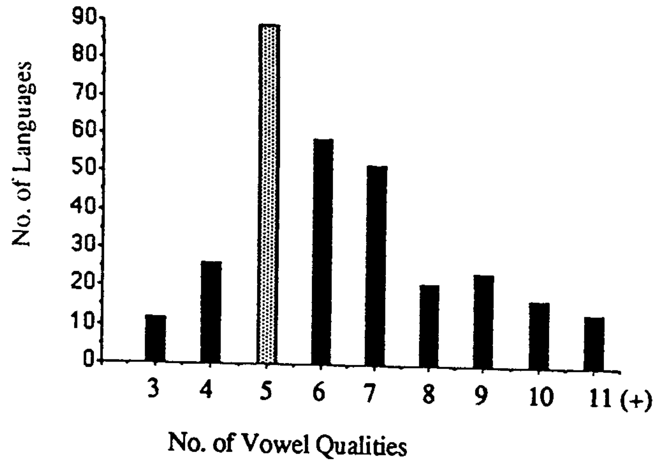
Figure 5. List of features, as displayed when editing a character code.

Finally, and probably most importantly, the program enables users to count or select and output to a file the particular subset of data that is crucial to the questions they want to address. Since the UPSID program itself performs only relatively simple selection and counting operations, many analyses will require use of other programs. The format of the output file is designed to make this easy. It is an ordinary ASCII file consisting of a list of abbreviated variable names followed by values for each individual segment output. Each segment is identified by language number, character code, and binary values for every feature. Such a file can easily be imported into a statistics package, such as SPSS or SAS. We will conclude this presentation with an example of the way the database can be used.

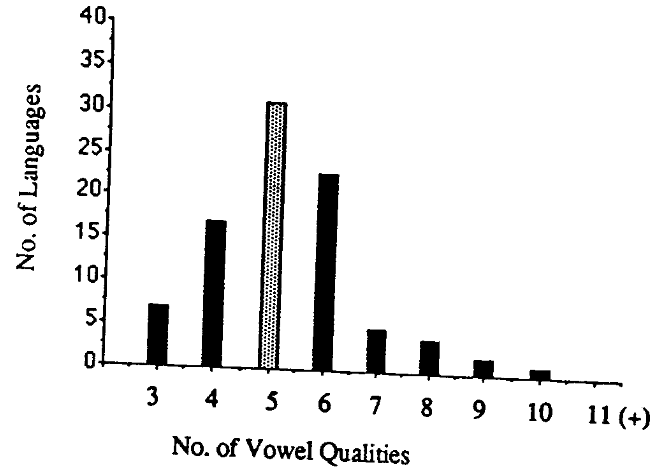
Suppose we wanted to investigate the hypothesis that there is a universal preference for vowel systems with just five qualitatively different vowels. We would select all vowels in the database - all segments with the value 1 for the feature "simple vowel" - and output these to a file. Then, reading the data into SAS, we would drop all features not related to vowel quality. Duplicates, such as vowels which are qualitatively the same but differ in nasality or length, must be eliminated from the file, so that each remaining entry in the file represents a distinct vowel quality in a particular language. Following this operation, we can now sum the number of vowel qualities for each language. Our provisional result, based on the partially completed database as it stood on November 23 1989, is shown in the form of a histogram in the upper left of figure 6. As may be seen, the modal system is indeed one with five vowel qualities.

However, before we conclude that this reflects a universal preference, we might heed the warnings of scholars such as Dryer (1987) and Perkins (1987), who have drawn attention to the widespread effects of areal convergence. Dryer suggests that a universal tendency can only be considered confirmed if is shown to occur independently in different regional and genetic language groupings. We sorted the languages into four large geographical groups, using the first digit of the language identification number. As noted earlier, this indicates the continental landmass on which the majority of languages in the family concerned is spoken — Eurasia, Africa, the Americas, or Australasia. We then reviewed the patterns of different vowel inventory sizes in each area.

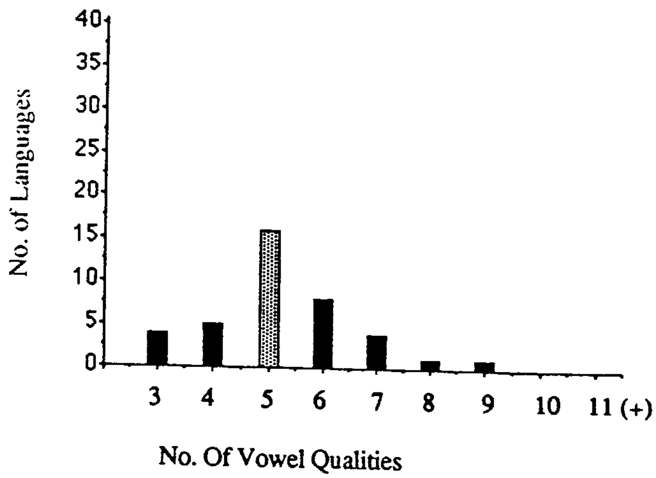
The pattern in indigenous languages of the Americas (upper right of figure 6) is largely similar to that seen in the overall sample. Five is the modal number here, shown by the stippled bar. The same is true of the languages of the Australian and Papuan families in our sample (middle left of figure 6). Languages of the Eurasian landmass show a somewhat different pattern (middle right of figure 6). Although five is also the modal number here, languages with larger numbers of vowel qualities are more common than they are in the preceding groupings. But African languages show a different modal preference (lower left of figure 6). Here seven, rather than five, is the



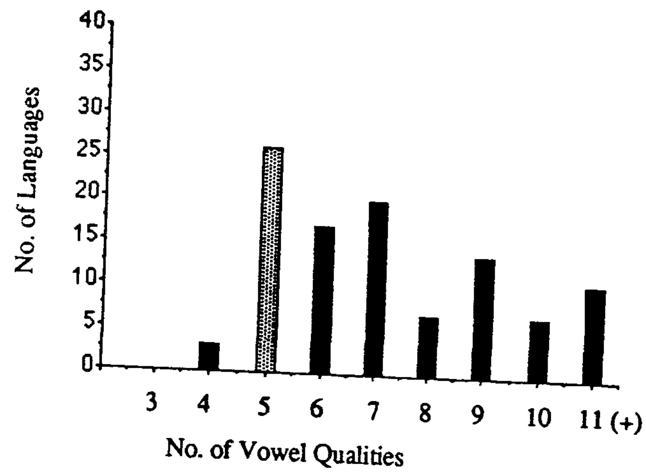
All Languages: Provisional Data.



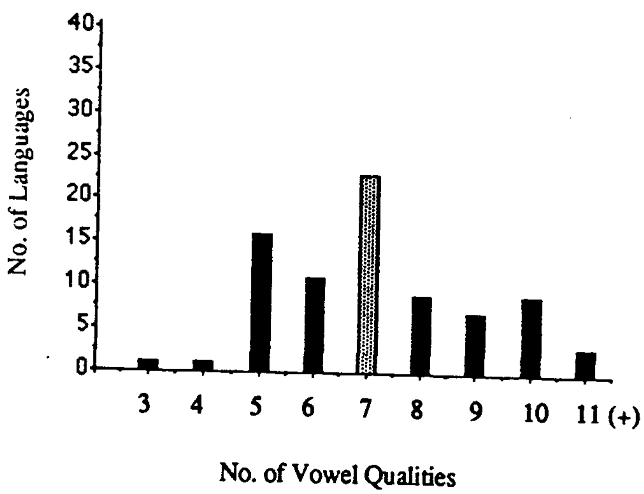
Americas (Amerind, Na-Dene, Eskimo-Aleut).



Australasia (Australian, Papuan).



Eurasia (Indo-European, Ural-Altaic, Sino-Tibetan, Austro-Tai, Austro-Asiatic, Dravidian, Caucasian, etc).



Africa (Niger-Kordofanian, Nilo-Saharan, Afro-Asiatic, Khoisan).

	Total	Eurasia	Africa	Americas
Eurasia	.87			
Africa	.71	.80		
Americas	.83	.49	.23	
Australasia	.90	.62	.39	.95

Figure 6. Sample analysis: size of vowel quality inventories.

most common number of vowel qualities. These groupings can be compared by examining correlations between the patterns of vowel inventory size represented in these histograms. If the patterns were comparable in the different groups, we would expect to see high correlations between the groups and high part-whole correlations between each group and the total language sample. The actual correlations are given in the lower right of figure 6. Africa has the lowest part-whole correlation, in the first column of the table. It has very low correlations with the American and Australian regional language groupings. We thus see that the question we asked has a less than straightforward answer.

So, to summarize, we have presented a description of an enhanced segment inventory database and outlined some characteristics of the program built around it. We have shown how data from the database can be readily accessed and manipulated to address questions of general interest for those investigating the structure of sound systems. Those interested in obtaining a copy of the program are invited to write to us at UCLA.

References

Dryer, Matthew S. 1987. Large linguistic areas and language sampling. Paper presented at the 1987 Linguistic Society of America Annual Conference, San Francisco.

Perkins, Revere. 1987. Statistical techniques for determining language sample size. Paper presented at the 1987 Linguistic Society of America Annual Conference, San Francisco.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge University Press, Cambridge.

The UCLA Language Database

Peter Ladefoged, John-Dongwook Choi and Stephan Schuetze-Coburn

[Modified version of a poster session paper presented at the 18th meeting of the Acoustical Society of America, November 1989]

The UCLA Language Database is a set of Macintosh Hypercard stacks, intended for both research and teaching purposes. Our aim is to illustrate, in a convenient way, a large percentage of all the sounds of the world's languages. There are, by our estimate, about 8,000 languages in the world. We are (obviously) not trying to illustrate every *language*, we are simply trying to illustrate as many of the *sounds* as possible.

We estimate that there are about 800 possible speech sounds. (This number is very dependent on what one considers to be a single phonetic unit, as opposed to a cluster.) The present collection contains data from 44 different languages, as shown in Table 1, and already illustrates several hundred different sounds. We have concentrated on trying to provide examples of the less well known sounds.

Table 1. Languages in the UCLA Hypercard database.

Agul	Akan	Avar	Bruu	Bura
Chinantec	Danish	Dedua	Eggon	Ewe
Gujarati	Igbo	K'ekchi	Kele	Korean
Liangshan Yi	Malayalam	Mandarin	Marathi	Melpa
Mid-Waghi	Mpi	Nama	Navaho	Nepali
Newari	Ngwo	Nunggubuyu	Polish	Quechua
Sindhi	Swedish	Temne	Titan	Tsakhur
Tsonga	Ubykh	Venda	V'enen Taut	Wangurri
Yanyuwa	Yeletnye	Yoruba	!Xóõ	

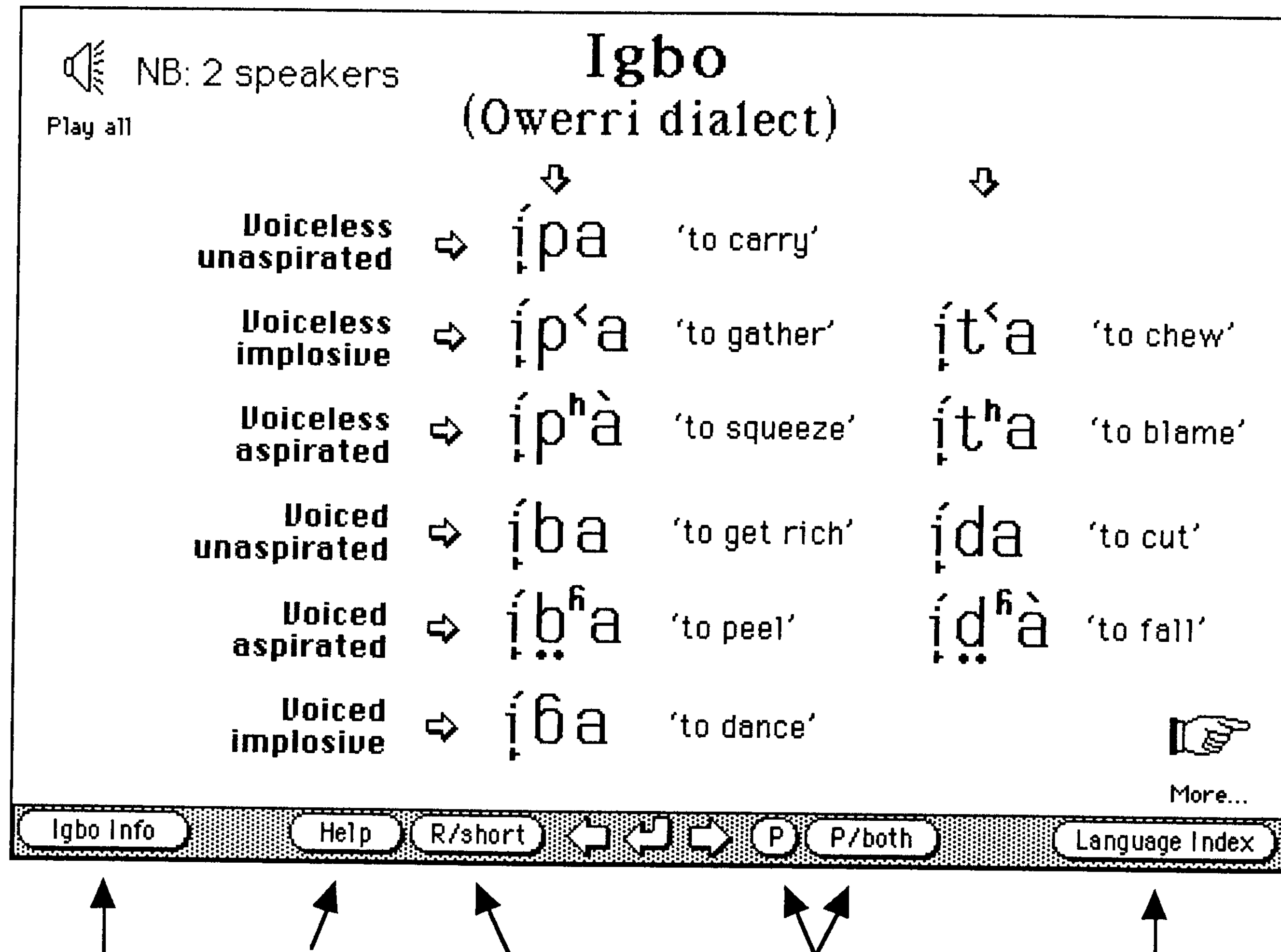
In each language there is a set of contrasting words in phonetic transcription, and an English gloss, as exemplified in Figure 1. Clicking with the mouse on any word plays a recording of that word as said by a native speaker or speakers. Clicking on a row or column of words plays all the words in that row or column. There are indexes allowing sounds to be located by type or by language. For each language there is also basic information on the language family, the number of speakers, and where the language is spoken. For some entries there are also reproductions of instrumental data. As an aid to students studying these sounds, at the bottom of each card there are

Example Card

Clicking on any word will reproduce a digitized recording of 2 native speakers saying that word.

Clicking on any arrow will play all the words in the row or column.

Clicking on a button in the button bar carries out the specified action.



Move to the card about this language.

Go to the Help card.

Record your pronunciation of these sounds.

Play your recording of these sounds (alone or with those of the native speakers).

Go to the language index.

Figure 1

buttons for use in conjunction with MacRecorder™ that enable students to record and play back their pronunciations of any word (or series of words), and to compare their pronunciation(s) with that of the native speaker.

The recordings are of varying quality, some having been made in a recording studio, others having been made in the field, in far from perfect conditions. They have all been digitized at 22 kHz, using 8 bit samples. Care was taken to ensure maximum use of the full 48 dB range (which was often better than the signal/noise ratio on the original recordings). There is very little difference evident in spectrograms made from the original recordings and those made from the database.

The number of stacks is growing all the time; at the moment the set takes up about 20 meg of disk space. The stacks can be used on a MacPlus or on any Macintosh that has Hypercard. While it is possible to use the stacks with a machine that has only floppy disk drives, a hard disk is virtually essential. The quality of the sound is very significantly improved if an external loudspeaker and amplifier is attached to the Macintosh. All the stacks are in the public domain, but users are asked to acknowledge the source of the data in any publication making use of the recordings.

The current version is available now; however, a substantially revised version will become available in March. Copies may be obtained for $50 from:

UCLA Phonetics Lab
Linguistics Department
UCLA
Los Angeles, CA 90024-1543

Please make checks payable to the Regents of the University of California.

Interarticulatory relationships in vowel production

Mona Lindau and Peter Ladefoged

We have been studying articulatory movements in English vowels using the University of Wisconsin x-ray microbeam system. This system allows us to observe particular points on the surface of the vocal tract. The points that we are concerned with in this paper are illustrated in figure 1 below, which shows the placements of six pellets: four on the tongue, one on the lower teeth as an indicator of the jaw position, and one on the lower lip. We studied the movements in the height dimension of the jaw, the lower lip, and the tongue. We are particularly interested in the degree to which speakers remain constant in repetitions of the same phrases, the variation between speakers in achieving similar vocal tract shapes, and such questions as whether the articulators move faster when the situation requires them to move further, or more intriguingly, whether they move further when the situation requires them to move faster.

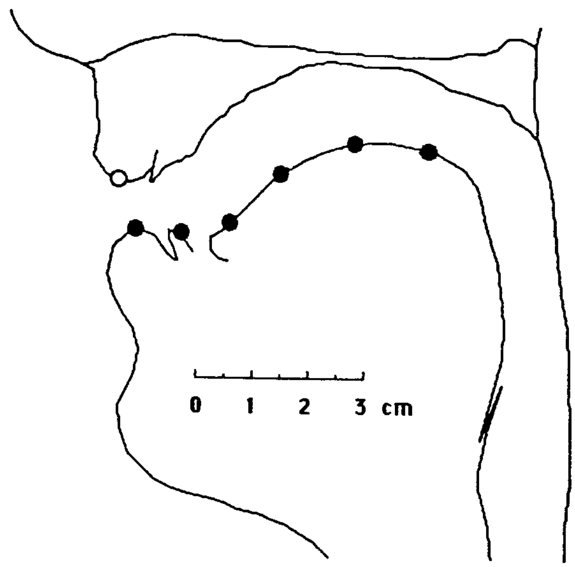


Figure 1: The placements of pellets for the x-ray microbeam recordings of the tongue and the jaw and the lower lip in this study.

From a larger data set we selected data from 4 speakers, 3 males and 1 female, who were all college students and native speakers of Mid-Western American English. These speakers produced 10 different vowels in monosyllabic words in 3 consonantal environments, with the vowels in each set being surrounded by consonants of the same place and manner, usually in the target words themselves, but sometimes in the sentence frame. The utterances are listed in Table 1 below. We tried to keep to real words, but the list does contain some nonsense words as well. Each word was spoken within a frame sentence and each such sentence was repeated 3 times, so there were 96 tokens for each subject in each run. Each of the four subjects repeated the whole procedure twice, making a total of 32 x 3 x 4 x 2 = 768 utterances.

Table 1. The set of phrases recorded.

Say	dee	to me	Say	bee	between	Say	see	serenely
	did			bib			sis	
	day			bay			say	
	dead			beb			cess	
	dad			bab			sass	
	Dodd			bob			soss	
	daw			baw			saw	
	doe			bow			sew	
	dood			---			soos	
	do			boo			sue	
	dud			bub			suss	

In this report we will consider only the position of the jaw, the lower lip, and the four points on the tongue measured in the middle of the vowel. The mid vowel position was selected as the point of maximum excursion of one of the tongue points, usually the pellet behind that on the tongue tip.

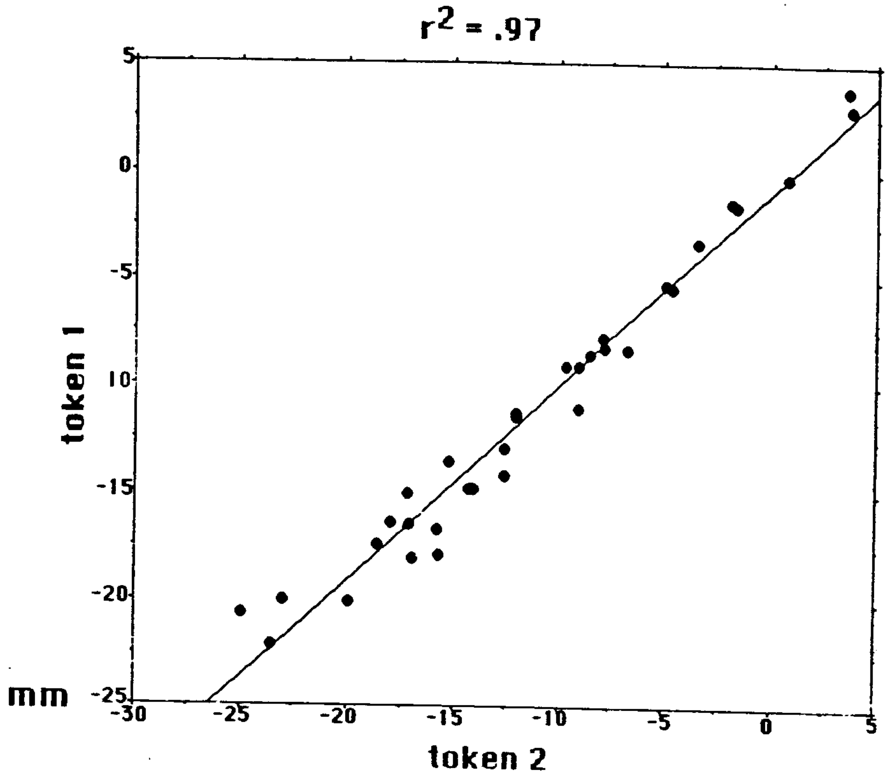


Figure 2. Comparison of the height values of the jaw, the lower lip, and one tongue-point in two tokens, one in the first run and one in the second for speaker 2. The measurements were taken from the mid position of the vowels in the s-s context.

The speakers had very little variation between repetitions of the same word. Regression analyses show a high correlation between tokens, even when they were recorded in different runs several minutes apart. Measurements of the height values of the jaw, the lower lip, and one point on the tongue in the mid position of the vowels were compared. Figure 2 shows a sample regression of the height values of the jaw, the lower lip and one point on the tongue for all the vowels in the s-s-context, comparing two tokens of the same word, one recorded early in the session, and one recorded in the latter part of the session. In this context, the r^2 values are between .95 - .99 for all the speakers. The token-to-token correlation within speakers was equally high for the b-b and the d-d contexts. These speakers did not exhibit any significant variation in repetitions of the same phrase. This consistency within speakers makes the large differences that we will report between speakers more valid.

In earlier studies (Ladefoged, DeClerk, Lindau, and Papcun 1972) we showed that within a language, i.e. English, the jaw and the tongue interact in different ways for different speakers in order to produce the vocal tract shapes required for a given vowel. Some speakers tended to behave in accordance with a jaw-based model of the tongue, where the tongue height was controlled by the height of the jaw. Other speakers moved the tongue and the jaw independently of each other. These studies were, however, limited to front vowels. Our x-ray microbeam data with all the vowels in three consonantal contexts confirm our earlier results. Table 2 below shows the results of multiple regression of the height of the jaw against the height of all four points of the tongue. The only speaker with high correlations between the height of the jaw and the height of the tongue for all three contexts is speaker 2 (S2). This speaker behaves in accordance with a largely jaw-based model of the tongue for vowels, attaining the required tongue height by adjusting the position of the jaw. The other speakers show lower correlations, i.e. more independence between the movements of the jaw and the tongue.

Table 2. r^2 values from multiple regression of jaw height vs the height of the tongue in 3 consonantal contexts (b-b, d-d, and s-s) for four speakers (S1, S2, S3, S4).

	b-b	d-d	s-s
S1	.50	.67	.82
S2	.96	.86	.93
S3	.72	.55	.80
S4	.69	.77	.77

In our earlier study the vowels were produced in an h-d context that minimized the influence of surrounding consonants. Our present results demonstrate that the interaction between the jaw and the tongue is to some extent context-dependent, as the goodness of fit within each speaker varies with the consonantal context. The movements of the jaw and the tongue are better correlated in the s-s context than in the other contexts for each of the speakers who move the tongue and jaw somewhat independently. This surprised us, as we thought that with the jaw tending to be in a high position for the initial and final /s/, it would have been necessary for the tongue to have made larger independent movements, resulting in a lower correlation between the jaw and the tongue movement.

Following these results, we then compared the vowel positions in the different consonantal contexts. We expected that the position in the middle of the vowel would be strongly affected by the active articulator of the consonant: if the lower lip, or the jaw, as the active articulator has to be in a high position for both consonants in the CVC words, then we expect the lower lip, or the jaw, to be in a relatively high position for the intermediate vowel than in a context of non-high consonants. Thus we expected the lower lip to be in a higher position in the middle of "bib" or "bab" than in the middle of "did" or "dad", respectively. Similarly, we expected the jaw to be in a higher position in "sis" or "sass" than in "did" or "dad", respectively.

We will begin by noting the coordination between the jaw and the tongue body in the s-s and d-d contexts, i.e. between words like "see, sis" and words like "dee, did". As is well known, in the production of [s], the mandible has to be in a high position so that the upper and lower teeth might be suitably juxtaposed for the high frequency friction (Keating, Lindblom, Lubker, and Kreiman 1990). Consequently, we expected that in the "s" series of words, the jaw would be in a higher position in the middle of the vowel than it would be in the middle of the vowel from the d-d context, due to either a carry over of the relatively high jaw position required for the initial consonant, or anticipation of the position required for the postvocalic consonant. Paired t-tests were used to compare the differences in the height of the jaw between the two consonantal environments.

Our expectation was born out for only two of the four speakers, namely Speaker 3 and Speaker 4, labeled S3 and S4, in figure 3. In this figure the jaw height for the d-set of words is plotted by the black triangles, and for the s-set by the open triangles. Each vowel is represented by three tokens. The gap in the chart of S3 occurs because this speaker did not have an /ɔ/ vowel. These two speakers show statistically significant differences in jaw height with the vowel in the s-s context being in the higher position. The differences are larger for Speaker 3 ($p < .0001$) than for Speaker 4 ($p < .001$). This is due to the fact that the jaw does not seem to move as far for Speaker 4, so the differences are restricted to the non-high vowels.

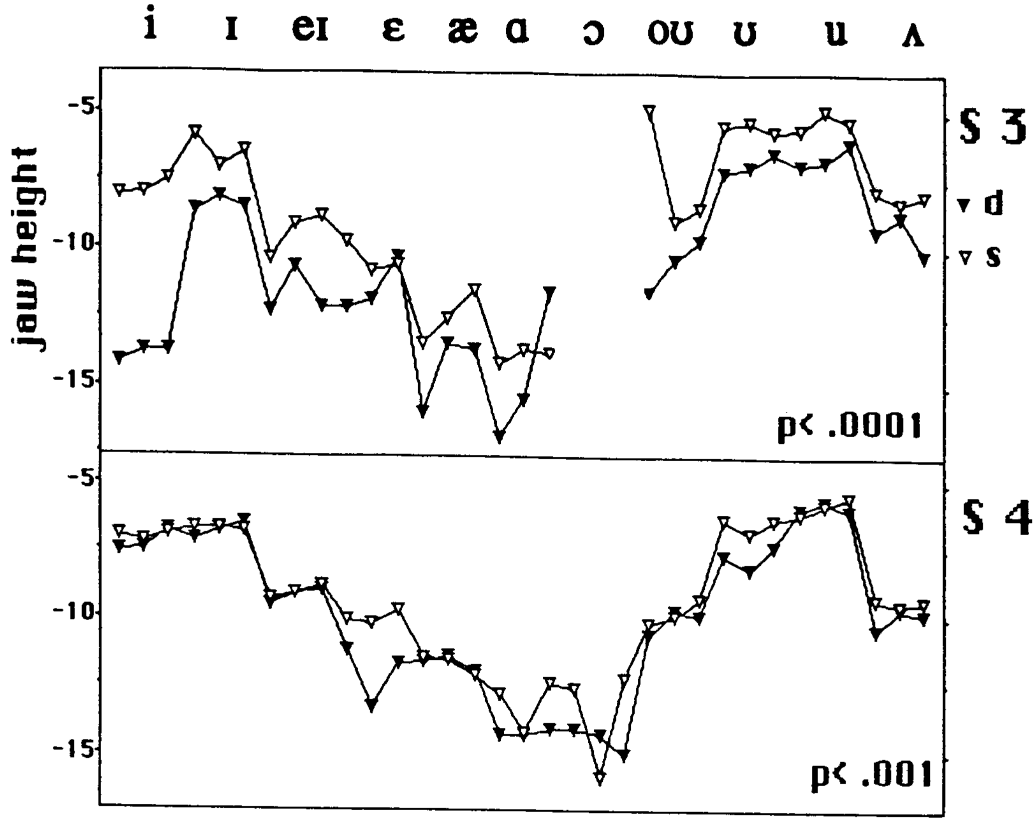


Figure 3. Jaw height for midvowel position of English vowels in a dVd-context (black triangles) and a sVs-context (open triangles) for speaker 3 (S3) and speaker 4 (S4). The p-values are from paired t-tests.

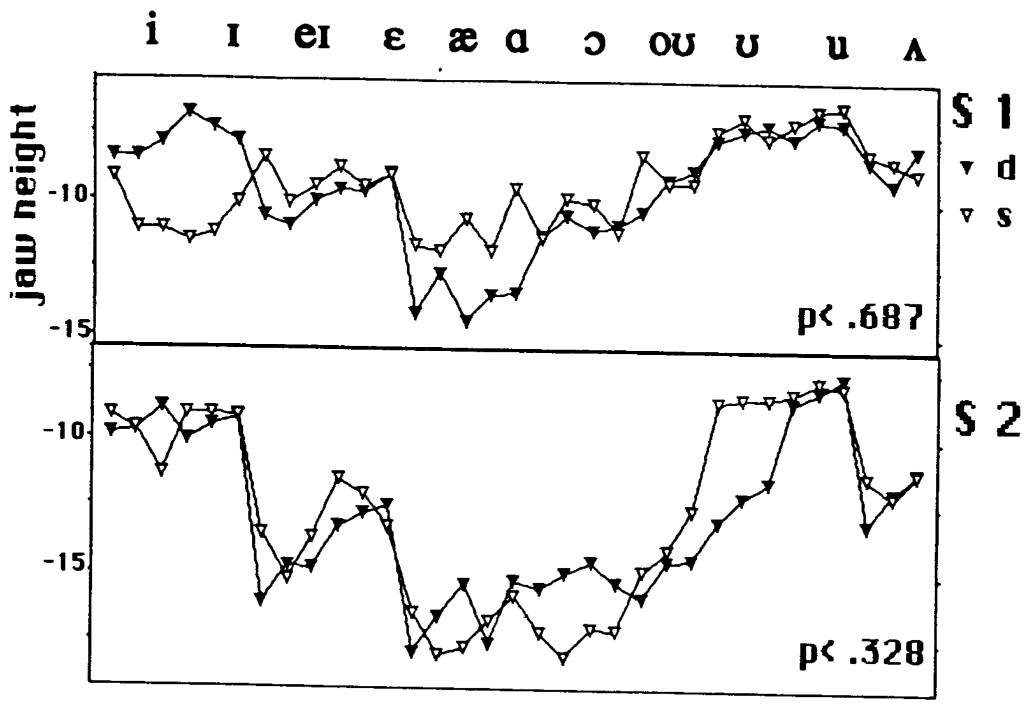


Figure 4. Jaw height for mid vowel position of English vowels in a dVd-context (black triangles) and a sVs-context (open triangles) for speaker 1 (S1) and speaker 2 (S2). The p-values are from paired t-tests.

Figure 4 shows the results for the remaining two speakers, labeled S1 and S2. These two speakers make no statistically significant difference in jaw position between the d-d context and the s-s context, and in fact, Speaker 1 has reversed the expected order for the high front vowels /i/ and /ɪ/, where the vowels the d-d context have a considerably higher jaw position than the vowels in the s-s context. However, for the open vowels which require a large mouth opening, this speaker does have a higher jaw position for vowels in the s-s context. Speaker 2 has a higher jaw position for some vowels, but the picture is confused for others.

The speaker differences between speakers in the /d-d/ and /s-s/ contexts may have to do with differences in laminal and apical articulation of the consonants. American English speakers differ widely in the way they articulate alveolar consonants. Some speakers produce alveolar consonants laminally, others apically (Dart, 1990). We assume that the tongue is higher for a laminal consonant than for an apical consonant, so it is possible that some speakers may use a laminal /s/ but an apical /d/. With a laminal /s/ and an apical /d/, which Dart has shown to be the most favored pattern among American speakers, the intervening vowels might reflect this aspect of consonant production, so that the vowels in the d-d context would be lower, as for speakers 3 and 4. The fact that the jaw is lower in the d-context than in the s-s context, may have to do with not only the high jaw position of the /s/, but may also be due to a faster action out of and into the plosive /d/.

Now consider the position of the lower lip in vowels in the /b-b/ context and the /d-d/ context. We expect the position of the lower lip to be higher in the /b-b/ context than in the /d-d/ context, because the lower lip is an active articulator and must be high for the labial consonant. Figure 5 is a plot of the results for speakers 1 and 3 (S1, S3), figure 6 is a similar plot for speakers 2 and 4 (S2, S4). The height of the lower lip in midvowel position is plotted with black circles for the /d-d/ context, and open circles for the /b-b/ context. The results show considerable variation, following our expectations for only one of the four speakers, namely speaker 3 (S 3). Figure 5 shows that the lower lip is significantly higher for the vowel in the b-context than the d-context for this speaker. The significant difference arises from the front and low vowels. The back and central vowels exhibit very similar lip heights in the two contexts. Out of the other three speakers, one (S1, figure 5, top) has no significant difference in lip position between the two consonantal contexts.

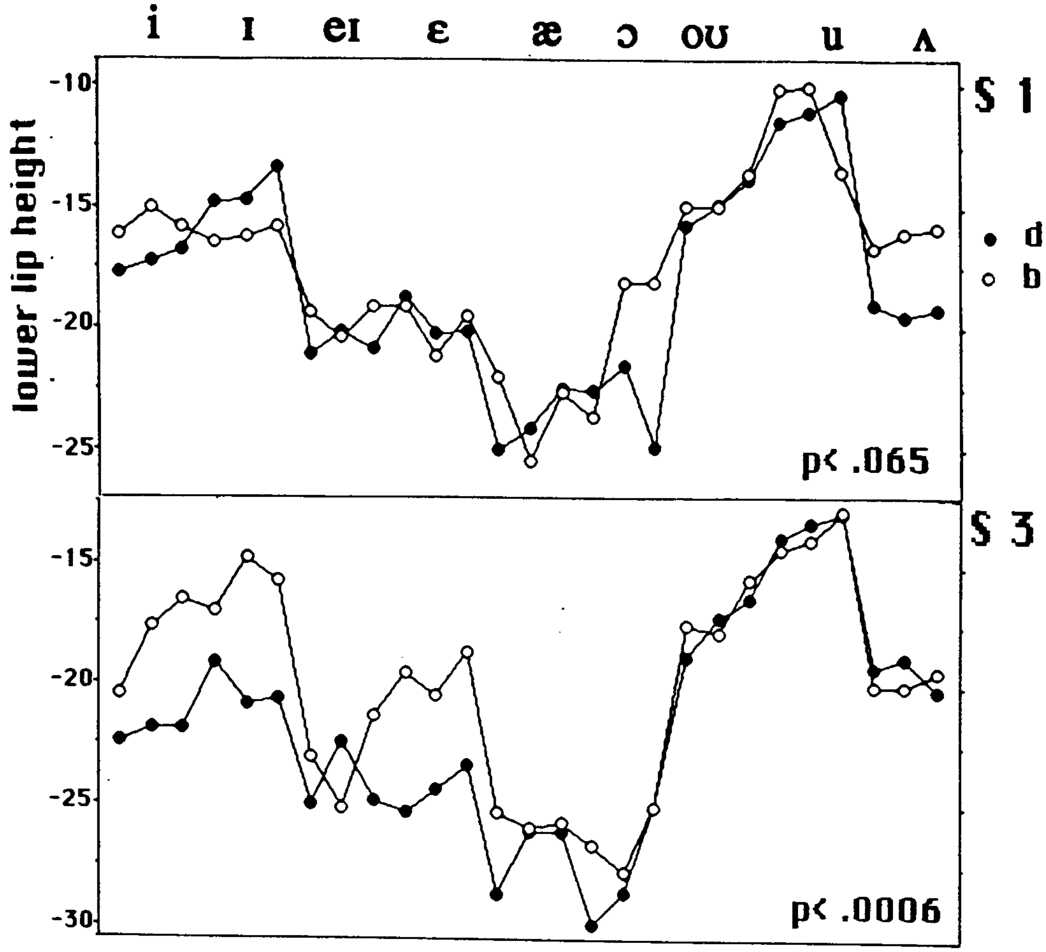


Figure 5. Lip height for midvowel position of English vowels in a dVd-context (black circles) and a bVb-context (open circles) for speaker 1 (S1) and speaker 3 (S3). The p-values are from paired t-tests.

Figure 6 shows the results for subjects 2 and 4. Here we find the opposite relationship: contrary to our expectations, the lower lip is significantly *lower* in the b-b context than in the d-d context. (This is in fact also the pattern of a 5th speaker with an incomplete data set). For these speakers, *all* the vowels, including the high vowels have a lower lip in the b-b context than in the d-d context.

We also looked at the height of the jaw in the b-b and d-d contexts. The results are similar to those for the lower lip. Speaker 3 has no significant difference in jaw height between the vowels in the b-b and the d-d contexts. The other three speakers showed a significantly *lower* jaw in the context of the bilabial consonants than in the context of alveolar consonants. These results are thus different from Keating et al. (1990), who found that the vowels in a b-context in both English and Swedish had a higher average jaw position than in all other consonantal contexts. In addition Keating et al. showed that the height of the jaw for the /b/ consonant was lower than for alveolar consonants. We are unable to explain these differences at this stage. The jaw positions of vowels in their study were analyzed from VCV utterances, not from CVC, but we doubt that can explain the difference.

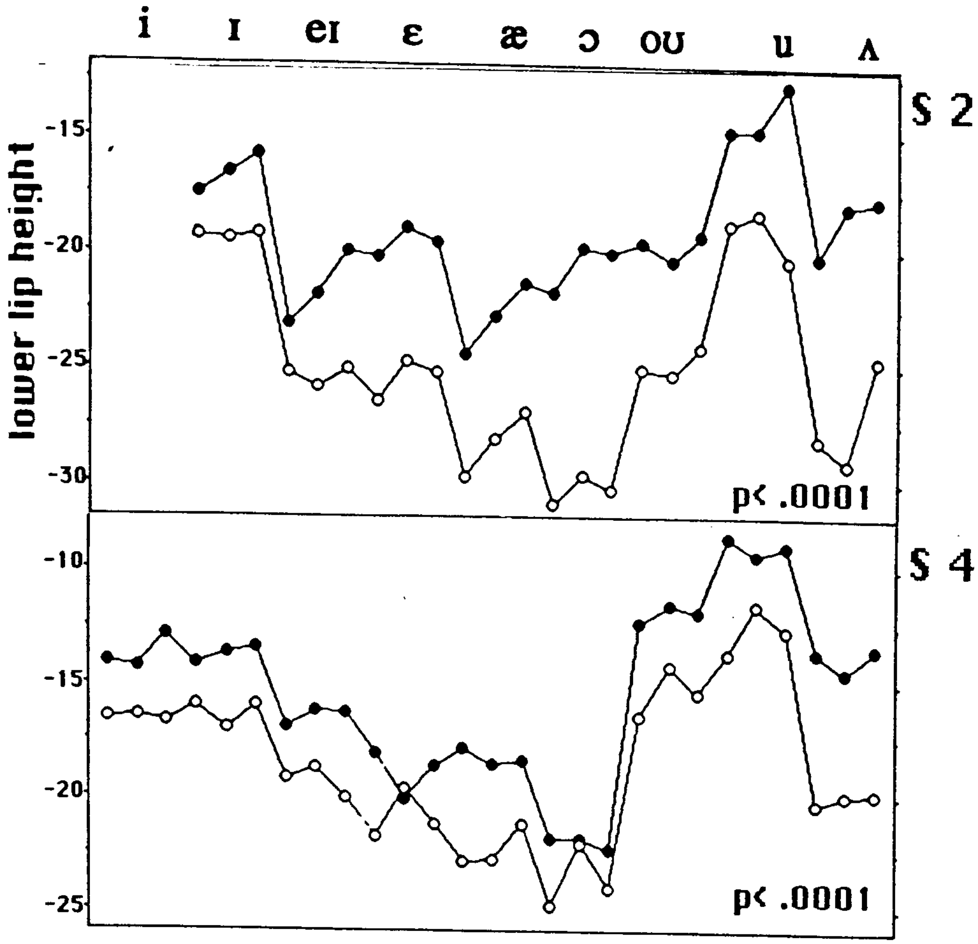


Figure 6. Lip height for midvowel position of English vowels in a d-d context (black circles) and a b-b context (open circles) for speaker 2 (S2) and speaker 4 (S4).

Why should the jaw position be lower in vowels in a b-b context than in a d-d context? Vowels tend to be shorter before labials than before dental/alveolars (Lehiste 1970), so if the lower lip articulator had less time to travel through the vowel, it should actually travel a shorter distance - unless the lower lip traveled at a *higher rate* through the vowel in the b-b context than in the d-d context. This appears to be the case. The movements in and out of plosives are relatively fast. In the b-b context, when the lower lip is also the active articulator, it travels faster during the vowel than in the d-d context. For most of our subjects, because it travels faster, it also travels a further distance. Many years ago Kent and Moll (1972) established the principle that, other things being equal, the further an articulator has to go, the faster it travels. We have found that on some occasions with stop consonants in which it is important for the lips to move fast, we have a reverse principle: the faster an articulator has to go, the further it moves. We seem to have a principle of "the faster, the further" operating here. When we look at the action of the lower lip in vowels in cases when the lower lip is also the active articulator of the surrounding plosives, this articulator apparently gets up speed during the vowel, and therefore moves further as well. Instead of the usual principle of phonetic undershoot when joining one gesture to another, a relatively high speed of an articulator results in an overshoot. Thus we can establish a new principle: other things being equal, if an articulator has to move faster, it also moves further.

References:

Dart, Sarah (1990). Articulatory and acoustic properties of apical and laminal consonants. UCLA Working Papers in Phonetics, forthcoming.

Keating, P., B. Lindblom, J. Lubker, and J. Kreiman (1990). Jaw position in English and Swedish VCVs. UCLA Working Papers in Phonetics 74:77-95.

Kent, R.D. and K.L. Moll (1972). Tongue body articulation during vowel and diphthong gestures. Folia Phoniatrica 24:278-300.

Ladefoged, P., J. DeClerk, M. Lindau, and G. Papcun (1972). An auditory-motor theory of speech production. UCLA Working Papers in Phonetics 22:48-76.

Lehiste, I. (1970). Suprasegmentals. Cambridge, Mass.: M.I.T.Press

Acknowledgments: We would like to acknowledge all the help we got from the X-ray Microbeam Facility at the University of Wisconsin, especially Bob Nadler. Financial support was provided by NIH.