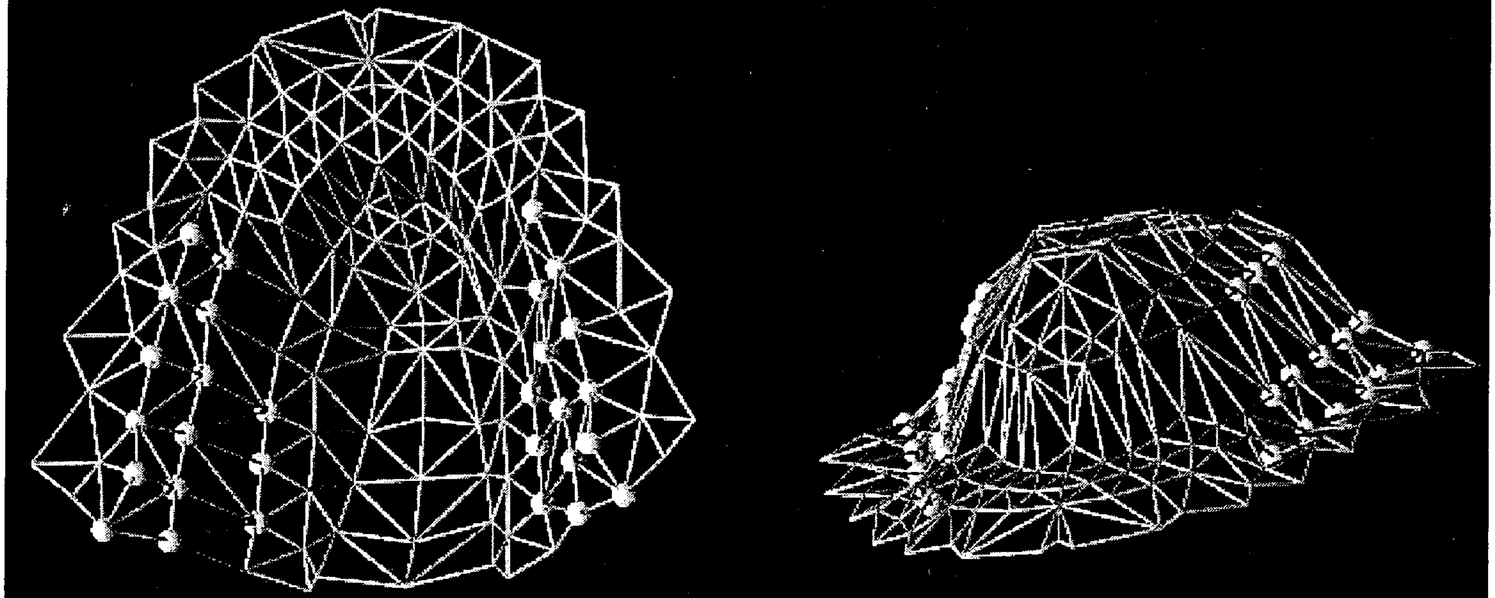
WPP, No. 85

1993-11-01

# UCLA Working Papers in Phonetics



## No. 85
## November 1993

# The UCLA Phonetics Laboratory Group

Zhoumagaly Abuov
Victoria Anderson
Barbara Blankenship
Dani Byrd
Court Crowther
Ken de Jong
Sandy Disner
Beatriz Dukes
Vicki Fromkin
Edward Flemming
Vicki Fromkin
Susan Meyers Fosnot
Cécile Fougeron
Robert Hagiwara
Bruce Hayes
Susan Hess
Sue Banner Inouye
Sun-Ah Jun
Pat Keating
Karn King
Paul Kirk
Jenny Ladefoged
Peter Ladefoged
Ian Maddieson
Joyce McDonough
Peggy McEachern
Benjamin Munson
Santa Samprath
Bonny Sands
Cheng Cheng Saw
Michael Shalev
Aaron Shryock
Caroline L. Smith
Siniša Spajić
Donca Steriade
Henry Teheranizadeh
Kimberly Thomas
Yuichi Todaka
Richard Wright

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:
USHHS grant 1 RO1 DC00642
USHSS grant 5 T32 DC00029
NSF grant BNS 9107004
NSF grant DBS 9213604
American Speech-Language-Hearing Association Foundation Award
and the University of California

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

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

This issue of UCLA Working Papers in Phonetics is edited by Ian Maddieson.

Associate Editors for UCLA Working Papers in Phonetics are Court Crowther and Susan Meyers Fosnot.

# UCLA WORKING PAPERS IN PHONETICS, VOL. 85

November, 1993

## Table of Contents

# Comments on privative versus binary features

**Patricia A. Keating**

*Presentation at a workshop on phonological features, UC Santa Cruz, July 1991*

This note is concerned with the implications for phonetic implementation of privative, or unary, phonological feature specifications. By phonetic implementation is meant the process of converting symbolic feature values into continuous values along phonetic (articulatory and/or acoustic) parameters. By privative or unary phonological feature specification is meant the limitation on features that they may have only one value (a positive, or plus, value). Privative features cannot take on a "minus" value, and thus there is no minus value that can be seen by the grammar or referred to by rules or constraints. Under privativity, then, phonetic implementation is limited to the implementation of positive feature specifications and cannot refer to negative values.

In at least some work on phonetic implementation, much use is made of traditional minus values of features to effect particular phonetic results.[1] The rise of privative theories presents a challenge to those phoneticians who have assumed and made use of full binary specifications. I will first comment generally on how phoneticians make use of feature values, and then consider whether and how privative features could be accommodated. All of this discussion will be couched in articulatory terms, since standard phonological features and most work in phonetic implementation are articulatory in nature.

The phonetic interpretation of a feature F can be viewed as generating continuous time functions along some set of phonetic parameters $P_1 \ldots P_n$. The feature specifications [+F] and [-F] each mean a specific, limited, range of possible values on P, one restricted so that it is clearly distinct from values seen for [+F] in the same context. The unspecified value, or [0F], means little or no restriction on possible values of F on P. In common practice, the plus value of a feature F is taken to indicate some specific articulation, and the minus value of that feature is taken to indicate the contrary gesture. For example, [+round] indicates lip protrusion, while [-round] may indicate lip spreading; [+nasal] indicates velum lowering, while [-nasal] indicates velum raising. Under this kind of phonetic interpretation, the minus value of the feature has as much meaning and impact as does the plus value.

In such a situation, there is no practical impediment to the use of privative features. Whatever interpretation results from [-F] in the binary system would pertain to [0F] in the unary system. There might be some objection that such an interpretation violates the spirit of privativity and its "no negative specifications" basis. In response it could be suggested that privative features are re-interpreted as binary before phonetic implementation.

The issue is not so simple, however. Some phoneticians rely on the distinction between [-F] and [0F], that is, the ternary power of underspecified binary features. While [+F] indicates a specific articulation and [-F] indicates an opposite, specific, articulation, [0F] indicates that there is no articulation specified. With [0F], the articulator in question

---

[1] Such work does not include Browman and Goldstein (e.g. 1989), whose gestural specifications are always unary; nor Pierrehumbert and Beckman (1988) and other work on intonation where the representations are two tonal units which can be construed as privative specifications of tonal features High and Low.

has no specific demands placed upon it by the segment in question. That articulator's activity is instead determined passively, by the context. Consider two examples of this from my own work.

First, in a brief discussion of vowel nasalization in English (Keating 1990a) I suggested that oral sonorants be unspecified for [nasal], while [-nasal] be reserved for oral obstruents, especially voiceless ones. (Nasal sonorants of course are [+nasal].) Whereas oral obstruents show velum raising, oral sonorants show context-dependent nasality; for example, a vowel between an obstruent and a nasal is transitionally nasalized. A crucial distinction is thus made between the phonetic interpretations of [-nasal] and [0nasal] within a single language, English. Second, I have suggested that languages differ in this same respect. This difference can be seen phonetically in the phenomenon of vowel-to-vowel coarticulation. (Vowel-to-vowel coarticulation refers to the mutual influence of one vowel on another, at least at their edges, across an intervening consonant; it can be thought of as a generalized phonetic precursor of harmony.) A language in which consonants bear contrastive secondary tongue body articulations will use the feature [back] for consonants, with repercussions on coarticulation. The degree of vowel-to-vowel coarticulation depends on how uninvolved the tongue is in making an intervening consonant. A [-back] or [+back] specification on a consonant will block or limit vowel-to-vowel coarticulation. In contrast, a language without any specification for [back] on consonants will show more extensive vowel-to-vowel coarticulation. Thus English shows much more vowel-to-vowel coarticulation than do various Slavic languages (Ohman 1966, Keating 1988, Choi and Keating 1991). The analysis of this phenomenon has been that a specified value ([+back] or [-back]) on consonants makes the values for [back] on vowels non-adjacent, and that coarticulation depends on adjacency of specifications. A similar analysis of consonant-to-consonant coarticulation in Marshallese is offered by John Choi in his 1992 dissertation. In her 1990 dissertation and her 1993 conference presentation, Abby Cohn develops these points with respect to the feature [nasal]. The point here is that [-back] and [0back] are taken to have very different phonetic consequences.

We have then a paradox. The phonological proposal for unary features is that [-F] is collapsed with [0F] because [-F] seems to be invisible. If it is invisible phonologically, it cannot then be visible later on to block phonetic interactions.

It seems to me that there is an alternative approach which can coexist with privative features, one in which [-F] is not referred to. This alternative comes from considering where the [-F] values in the examples above come from. The value [-nasal] comes from the values for other features ([sonorant], perhaps [voice]). The value [-back] for consonants comes from the system of consonant contrasts in a language: in a language with [+back] consonants, [-back] also carries information. Strictly speaking, then, [-F] specifications are a handy encapsulation of information which is available elsewhere in segment specifications. This information can be constructed in the phonetics as part of the phonetic implementation of [0F], rather than be handed over to the phonetics by the phonology. After all, if only the phonetics cares about this distinction, then the phonology should not encode it. Instead, phonetic interpretation can be made to depend more on the paradigmatic context.

In this light, let us revisit the examples given above. In the case of English sonorant nasalization, [+nasal] means a narrow range of open velum positions, while [0nasal] can depend on other features: for [-voice] segments (which will all be obstruents) it means a narrow range of closed velum positions, and otherwise it means a wide range of possible velum positions. Abby Cohn has shown that when the cross-language situation is considered, things are not so simple, but let us assume that the analysis could be developed more generally. In the case of blocking coarticulatory interactions, [+back] means a constriction behind the hard palate, while [0back] must be interpreted in a language-specific manner. In a language with [+back] consonants, [0back] means constriction in the hard palate region. In a language with no [back] specifications for consonants, there is no specific constriction or constriction location associated with the tongue body for consonants.

In sum, it would fall to the language-specific phonetic system to determine the phonetic interpretation of unspecified feature values.

A different sort of case is presented by the feature [voice]. The marked value [+voice] clearly refers to a set of articulations which allow vocal cord vibration. Unlike with [nasal], [round], and [back], however, the minus value of [voice] does not refer to any single specific gesture contrary to that for [+voice]. Instead, [-voice] indicates any gesture sufficient to prevent voicing (generally glottal spreading or constriction, but potentially perhaps certain supralaryngeal maneuvers as well). On its face, this makes [voice] look like a better candidate for a privative treatment, as proposed by e.g. Mester and Ito (1989, section 5), Cho (1990), and Lombardi (1991).

Nonetheless, a role for the distinction between [-voice] and [0voice] exists in the analysis of near-neutralization of voicing in final obstruents, a phenomenon discussed most convincingly by Port & Crawford (1989). The claim in this and other studies is that final obstruent devoicing in, for example, German, is not completely neutralizing; that is, derived [-voice] obstruents are not physically identical to underlying [-voice] obstruents. Rather, there are slight phonetic differences between them, differences which are highly variable across speakers and studies. As these and other researchers have noted, the lack of physical identity indicates that the output of a devoicing rule cannot be identical with an underlying [-voice] specification, at least not if that output is the sole input to phonetic interpretation of these different obstruent classes. In previous work (Keating 1984) I've hinted at an analysis of this phenomenon based on markedness that can be seen as relying on a difference between [-voice] and [0voice]. Let me spell this out here. First, as in current discussion of devoicing under privative theories, devoicing can be seen as the deletion or delinking of a [+voice] specification, resulting in [0voice]. Second, unlike in privative theories, there are underlying [-voice] obstruents. The output of devoicing thus remains distinct from underlying voicelessness. Phonetically, underlying [-voice] is interpreted by voicing-preventing gestures. [0voice] is interpreted as default or neutral settings, with variable results, as discussed by Westbury & Keating (1986).

This analysis, then, relies on ternary-powered voicing specifications in surface representation. It accounts for the phonetiic difference between devoiced and underlying-voiceless obstruents by giving them different representations. It accounts for the variability in devoiced obstruents by letting them be interpreted by a variable phonetic mechanism. In contrast, under privative voicing with no [-voice] value available, the [0voice] output of devoicing must be like the underlying voiceless specification. Here again is our paradox: the phonology may not distinguish between devoiced and voiceless obstruents, but if they are collapsed phonologically, then phonetic interpretation will see no difference between them. At the same time, the kind of solution offered for the cases above will not carry over to this one. Here, there is no other information from other features that could be used to distinguish the two kinds of voiceless obstruents.

Nonetheless, even in this case there are alternative approaches compatible with privative voicing. The more traditional one is to find a way to get extra information into a representation before devoicing, the deletion of a voicing specification, occurs. If there is other distinguishing information in the representations, then devoicing will not be neutralizing even phonologically, and phonetic implementation will preserve the indicated distinction. Such a way is the redundant feature values of Stevens et al.'s (1986) "enhancing" features, by which non-contrastive feature values are introduced into representations to bolster underlying contrasts. Suppose that enhancing feature values are introduced before devoicing, and that these values are not affected by devoicing. The generally-recognized enhancers of unspecified-voice would be [spread glottis] and [constricted glottis] (see also Keating 1990b). The analysis would be:

1) assign enhancing [spread glottis] to [0voice] obstruents (more precisely, assign it to all of them but subject to a co-occurrence restriction against [spread glottis] with [+voice];

2) delete [+voice].

The outputs of these rules are distinct laryngeally, but non-distinct re [voice], since both are [0voice].

The less traditional alternative is to regard devoicing as a phonetic, rather than a phonological, rule. Pierrehumbert (1990) outlines such an account, noting that devoiced obstruents are not referred to by later phonological rules; that is, nothing hinges on doing devoicing in the phonology. Putting her analysis into privative terms, Pierrehumbert suggests that phonetic implementation of voicing proceed as follows: [0voice] (underlying voiceless) obstruents are implemented with voicing-suppressing gestures, while [+voice] obstruents are implemented in a context-dependent fashion. Non-final [+voice] obstruents are implemented with voicing-permitting articulations, while final [+voice] obstruents are implemented weakly. Thus the phonological structural condition is preserved in the phonetic rule, but the phonological structural change is reformulated as a different phonetic implementation. In effect, it preserves the phonetic origin of devoicing process (Ohala 1983, Westbury & Keating 1986) as the synchronic devoicing mechanism, giving German simply a more extreme version of what we see phonetically in English (e.g. Veatch 1990).

In conclusion, I hope to have shown that there is quite a range of options for phonetic implementation, any or all of which could be used to pare down surface representations to privative feature specifications and yet still accomplish phonetic implementation with appropriate details.

## References

Browman, C. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201-252.

Cho, Y. Y. (1990). A typology of voicing assimilation. *WCCFL* 9.

Choi, J.-D. (1992) Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony. UCLA Ph.D. dissertation. [UCLA Working Papers in Phonetics 82]

Choi, J.-D. and Keating, P. (1991). Vowel-to-vowel coarticulation in three Slavic languages. *UCLA Working Papers in Phonetics* 81, 78-86.

Cohn, A. (1990). *Phonetic and Phonological Rules of Nasalization.* UCLA Ph.D. dissertation. [UCLA Working Papers in Phonetics 76]

Cohn, A. (1993). Privative features: consequences for an adequate theory of phonetic implementation. Presentation at the Phonetics-Phonology workshop, Ohio State U., August 1993.

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

Keating, P. (1988). Underspecification in phonetics. *Phonology* 5, 275-292.

Keating, P. (1990a). The window model of coarticulation: articulatory evidence. In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.*. Cambridge: Cambridge University Press, 451-470.

Keating, P. (1990b). Phonetic representations in a generative grammar. *J. of Phonetics* 18, 321-334.

Lombardi, L. (1991). *Laruyngeal features and laryngeal neutralization.* Ph. D. dissertation, U. Mass/Amherst.

Mester, R. A. and Itô, J. (1989). Feature predictability and underspecification: palatal prosody in Japanese Memetics. *Language* 65, 258-293.

Öhman, S. (1966). Coarticulation in VCV utterances: spectrographic measurements. *JASA* 39, 151-168.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *J. Phonetics* 18, 375-394.

Pierrehumbert, J. and Beckman, M. (1988). *Japanese Tone Structure.* Cambridge: MIT Press.

Port, R. and P. Crawford (1989). Incomplete neutralization and pragmatics in German. *J. Phonetics* 17, 257-282.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P.F. MacNeilage (ed.), *The Production of Speech.*. New York: Springer-Verlag.

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

Veatch, T. (1989). Word-final devoicing of fricatives in English. Annual meeting of the LSA, December 1989, Washington DC.

Westbury, J.R. and P. A. Keating (1986). On the naturalness of stop consonant voicing. *J. Linguistics* 22, 145-166.

# Phonetic representation of palatalization versus fronting

## Patricia Keating

From a talk given on May 4, 1991 at the University of Illinois conference
"The Organization of Phonology: Features and Domains"

### Introduction

The ways in which consonants and vowels can interact and share articulations provide important evidence about the identity and organization of features. In this note, I will describe the phonetic nature of two related -- though, as we will see, different -- kinds of interaction between primary consonants and vowel-like articulations: palatalization of consonants and contextual fronting of velars. I will suggest that in one sense palatalization is quite simple phonetically, although its repercussions can be complex, and that this phonetic simplicity places some limits on how we might want to represent it phonologically. Then I will argue that velar fronting in English is quite a different matter, one which shouldn't be represented phonologically at all.

### Palatalization

First, some terminology should be clarified. Following Ladefoged (1982:210), "Palatalization is the addition of a high front tongue position, like that of [i], to another articulation". This is called *secondary palatalization* because the palatalization is a secondary articulation added to a primary one. As Ladefoged also notes, "The terms palatalization and palatalized may also be...applied in describing a process in which the primary articulation is changed so that it becomes more palatal". This can be called *primary palatalization* because the primary articulation is affected and there is no separate secondary articulation. Examples of primary palatalization include changes of [s] to [ʃ] or of [k] to [tʃ]; the palato-alveolars [ʃ] and [tʃ], while not palatal, are more palatal than alveolars or velars. A variety of articulatory changes may count as primary palatalizations on this criterion. Bhat (1978) notes that palatalization can be viewed as three different articulatory processes that can occur seaparately or together: tongue (body) fronting, tongue (blade) raising, and spirantization (the addition of stridency).

Palatalization, both primary and secondary, has been of recent interest in feature theory, and a sort of consensus has been developing around the idea that it is a function of a coronal articulation. Two recent versions of this idea can be found in Mester and Ito (1989) and Clements (1991). Though these two proposals differ in overall organization and in details, they share the core idea that palatalization is not accomplished by the tongue body, as had been assumed, but by the tongue blade or front. The tongue blade and front together are taken to define the coronal primary articulator, and therefore palatalization is analyzed as the addition of a coronal articulation to a segment. Additionally, Mester and Ito (1989) incorporate Bhat's generalization that palatalization of coronals involves retraction of the primary constriction, by further specifying this coronal articulation as [-anterior]. Given Bhat's related generalization that palatalization of coronals involves lengthening of the primary constriction, the coronal articulation might be further specified as [+distributed].

Let us consider some phonetic evidence on the following questions. Is palatalization a function of the tongue blade or front, and not the body? When applied to coronal consonants, does palatalization definitionally involve retraction and/or lengthening of the consonant constriction? To address these questions, I have chosen to examine articulatory data from Russian, a language which has secondary palatalization, and which has the advantage of available articulatory documentation in the form of X-rays and palatograms.

Secondary palatalization, recall, is addition of an [i] articulation to a consonant, so palatalization of a labial should show us palatalization in a relatively pure form. In labials the palatalization is independent of the primary consonantal articulation. Figure 1 shows two pairs of palatalized and non-palatalized labials. This figure, like the others to follow, shows two superimposed tongue tracings. The tongue position for the non-palatalized consonant is shown as a solid line, and the tongue position for the palatalized consonant is shown as a dashed line. Some of the figures, including Figure 1, are accompanied by palatograms showing the area on the roof of the mouth contacted by the tongue. Increased contact on the roof of the mouth means a raised tongue. Note that the palatograms are not from the same tokens as the X-ray tracings. Figure 1 shows secondary palatalization with the whole tongue moved up and forward in the X-ray and with a large contact area in the palatogram.

Figure 2 shows secondary palatalization of an anterior coronal stop. In this and similar figures, the palatogram shows two different contact areas, the larger one (innermost line) corresponding to the palatalized consonant. The X-ray in this figure shows a similar movement by the tongue as a whole, but here it is constrained by the primary stop constriction. A sort of pivoting of the tongue occurs that rolls the primary constriction around up onto the alveolar ridge and back on the tongue blade. Thus the primary coronal articulation is retracted and made laminal by the secondary palatalization. As a result, the stop is generally also affricated.

Figure 3 shows a similar view of the anterior coronal fricatives. In contrast to /t/, secondary palatalization of /s/ has little effect on the primary constriction. The X-ray tracing shows the tip to be in the same place for the palatalized and non-palatalized consonants. Although the tracing suggests a lengthening of the constriction, the palatogram clearly shows virtually identical locations and lengths for the two constrictions. Finally, the X-ray tracing shows the bulk of the movement to be behind the blade, under the roof of the mouth. Figure 4 shows a different style of tracing from X-ray frames of the same sounds; it also shows that the primary constriction is essentially maintained under secondary palatalization. Here the dotted line indicates the tongue position for the palatalized fricative, while the dot-dash line indicates the tongue position for the non-palatalized fricative, as well as the rest of the vocal tract.

In Russian, the fricatives that in the IPA tradition are often transcribed as palato-alveolar [ʃ] etc. can be seen to be apical and slightly retroflex. I use [š], the traditional Slavic symbol, to mean such a weakly retroflex post-alveolar. There is a surface contrast in Russian between this fricative and a palatalized variant, primarily resulting from cluster simplification. Figure 5 shows X-ray tracings of these two fricatives in the same style as Figure 4. Both fricatives in Figure 5 show the clearly apical articulation, in essentially the same place, and the dotted palatalized fricative shows the tongue raised and fronted behind the primary constriction. The retroflex nature of the articulation is clearer still in Figure 6, where the shaded tongue is the non-palatalized fricative and the non-shaded tongue is the palatalized fricative. The tongue position for the non-palatalized fricative is notable in showing two simultaneous component articulations: the primary retroflexion and a secondary velarization. The palatalized fricative can be seen to have two component articulations also, but the tongue behind the blade is raised and fronted so that it is less clearly set off from the blade. Note how the movement of the tongue moves the blade forward as well, so that the primary constriction is slightly fronted under palatalization. This is not the retraction observed for the anterior stop in Figure 2, but it is to be expected if the

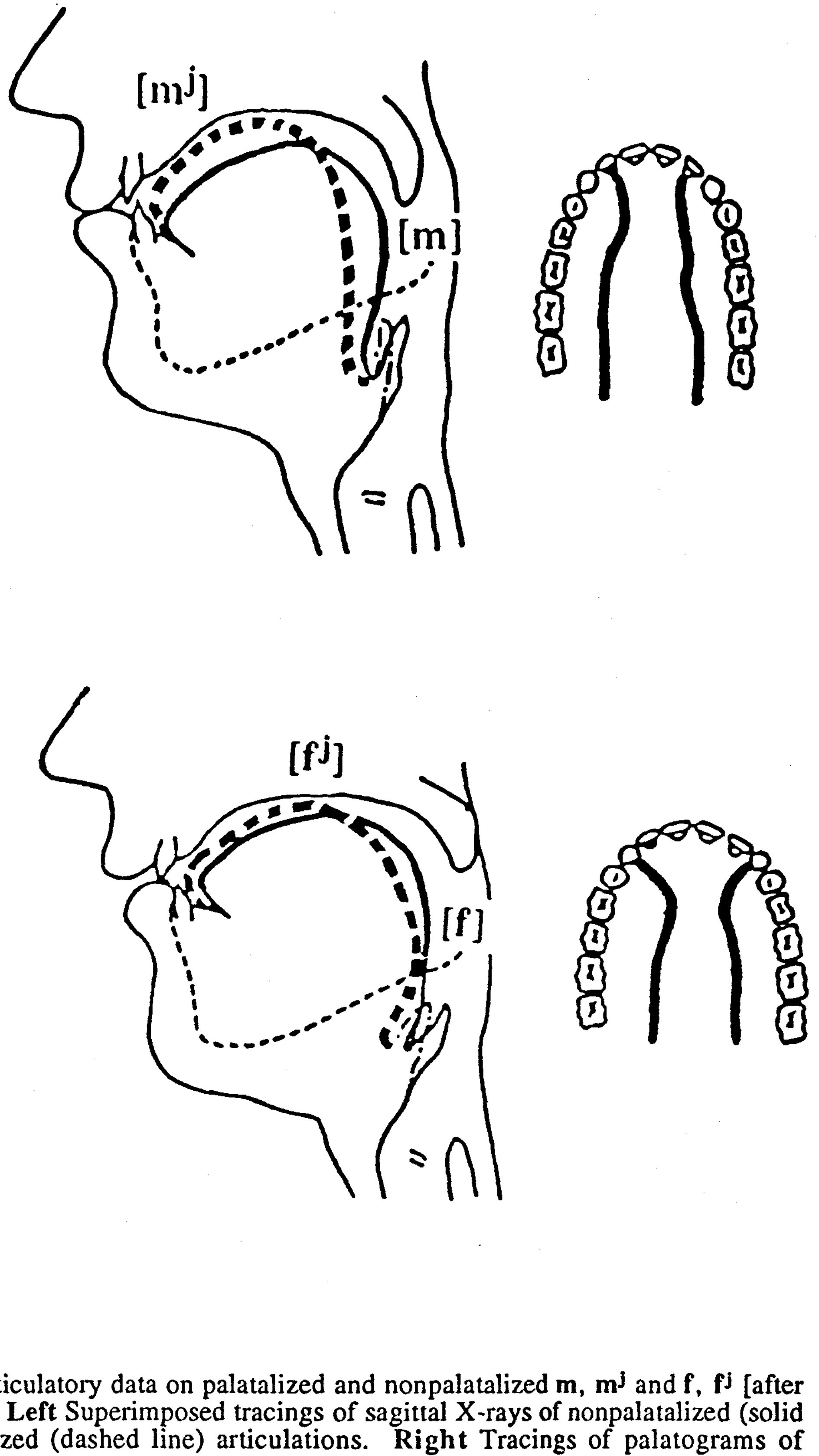


**Figure 1:** Articulatory data on palatalized and nonpalatalized **m**, **m**$^{j}$ and **f**, **f**$^{j}$ [after Oliverius, 1974]. **Left** Superimposed tracings of sagittal X-rays of nonpalatalized (solid line) and palatalized (dashed line) articulations. **Right** Tracings of palatograms of palatalized articulations.

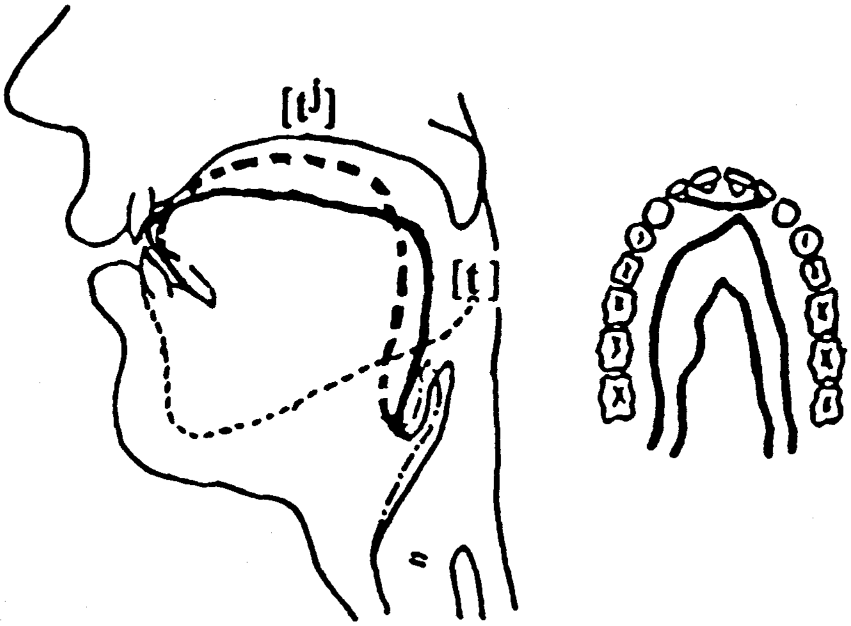


**Figure 2:** Articulatory data on palatalized and nonpalatalized **t**, $t^j$ [after Oliverius, 1974]. **Left** Superimposed tracings of sagittal X-rays of nonpalatalized (solid line) and palatalized (dashed line) articulations. **Right** Tracings of palatograms of nonpalatalized (outermost) and palatalized (innermost) articulations.

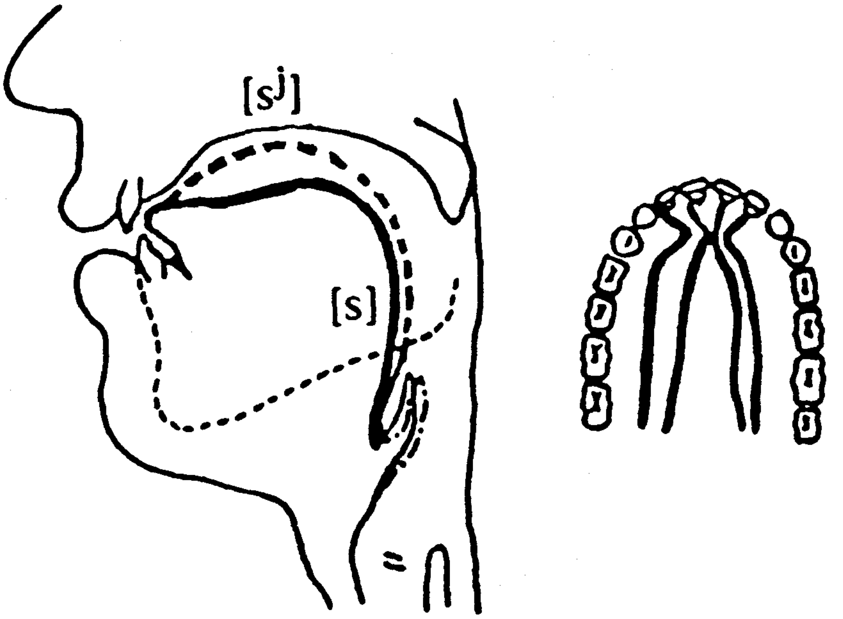


**Figure 3:** Articulatory data on palatalized and nonpalatalized s, $s^j$ [after Oliverius, 1974]. **Left** Superimposed tracings of sagittal X-rays of nonpalatalized (solid line) and palatalized (dashed line) articulations. **Right** Tracings of palatograms of nonpalatalized (outermost) and palatalized (innermost) articulations.

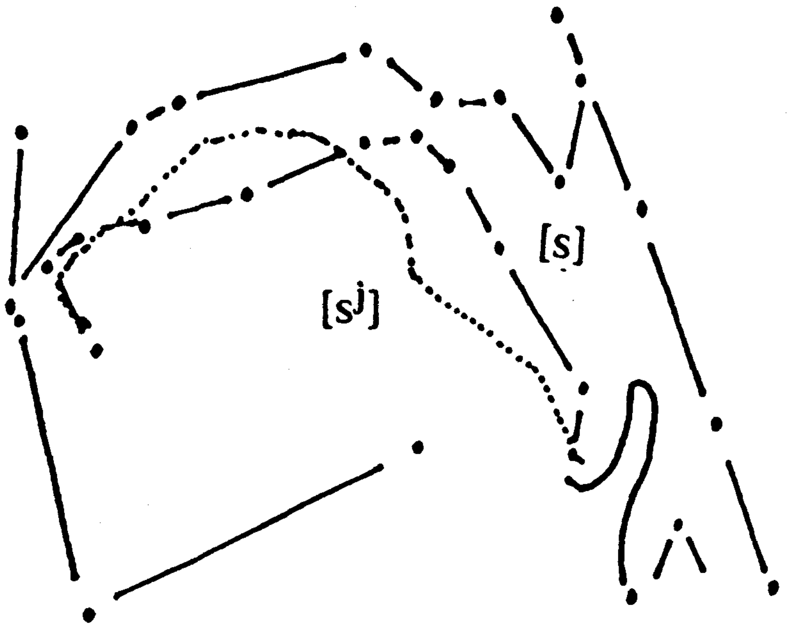


**Figure 4:** Articulatory data on palatalized and nonpalatalized s, s$^{j}$ [after Bolla, 1982]. Superimposed tracings of sagittal X-rays of nonpalatalized (dot-dash line) and palatalized (dotted line) articulations.

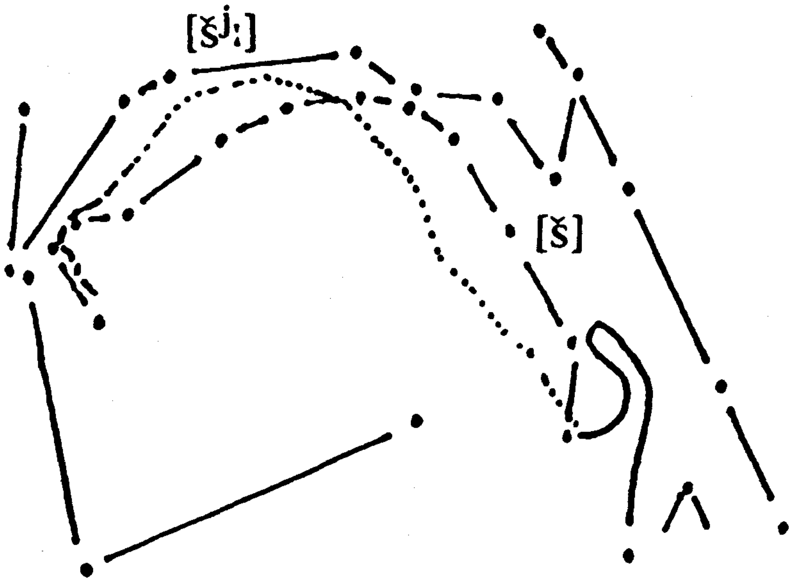


**Figure 5:** Articulatory data on palatalized and nonpalatalized š, š$^{j}$: [after Bolla, 1982]. Superimposed tracings of sagittal X-rays of nonpalatalized (dot-dash line) and palatalized (dotted line) articulations.

retroflexed tongue shape is maintained and the blade is pushed down. On the other hand, Figure 7 shows another set of these fricatives in which the palatalized variant is not retroflexed and its primary constriction is indeed retracted.

Finally, Figure 8 shows palatalized and non-palatalized velar stops. The shift of the tongue position is much as with the other consonant places, but here it is constrained in the back by the requirement of a dorsal stop articulation. The tongue as a whole moves forward (and for this speaker, up) along the roof of the mouth. The primary constriction moves forward from the soft to the hard palate without being lengthened. We might have expected the primary articulation to stay in the same place and a secondary constriction to be added in front of it along the palate, resulting in a single long constriction. Instead, the secondary palatalization is in fact a primary palatalization which is a simple shift in location of the primary articulation.

In one sense, because of the differing primary constrictions, these instances of palatalization vary. Yet there **is** an articulatory constant. The tongue body is bunched up and moved toward the hard palate. Imagine a string pulling a point on or in the tongue--a point fairly far back on it--towards the back edge of the hard palate. The approximation involved is indicated in Figure 9. Such a mechanism is modeled as a point attractor in task dynamics. What happens to the tongue overall as a result of such a force depends on the constraints imposed by the primary articulation. Looking across all the figures, the constant feature of palatalization appears to be the change in position of the part of the tongue under the roof of the mouth.

What constraints does this set of data place on the full surface representation of such secondary palatalization? To me as a phonetician, it makes little sense to attribute this palatalization to the tongue blade, even generously construed. The part of the tongue that appears to be attracted to the roof of the mouth is rather far back from the tongue tip, and it is a part of the tongue which is distinct from the part forming primary coronal articulations. If this is all blade, then it's a blade with two independent sub-parts, and we lose the whole notion of a single coronal articulator. I take **Dorsal** to be the expression of this articulator behind the tongue blade. Furthermore, these data also provide evidence that palatalization does not necessarily involve the retraction of a primary coronal articulation to a non-anterior position. We saw that palatalization can have no effect on constriction location for an anterior fricative, and that it can actually front the constriction location for an apical retroflex. In sum, the use of a Coronal articulator node, especially one specified as [-anterior], to indicate secondary palatalization finds little phonetic support in these data.

While it is not the phonetician's role to tell phonologists how to organize a feature geometry, on conservative lines the structure shown in (1) represents the phonetic situation. (Many details here are not crucial.)

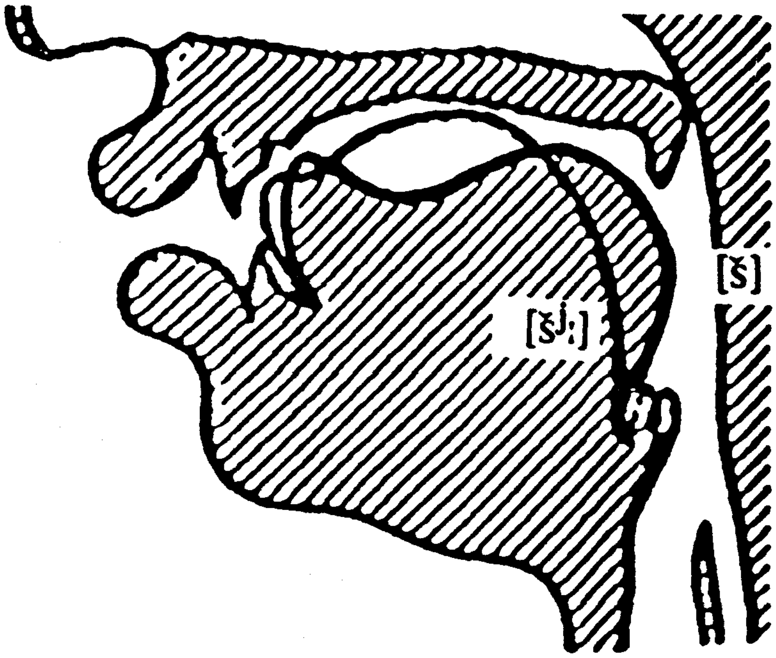


**Figure 6:** Articulatory data on palatalized and nonpalatalized š, šʲː [after Akishina and Baranovskaja, 1980]. Superimposed tracings of sagittal X-rays of nonpalatalized (shaded) and palatalized (unshaded) articulations.

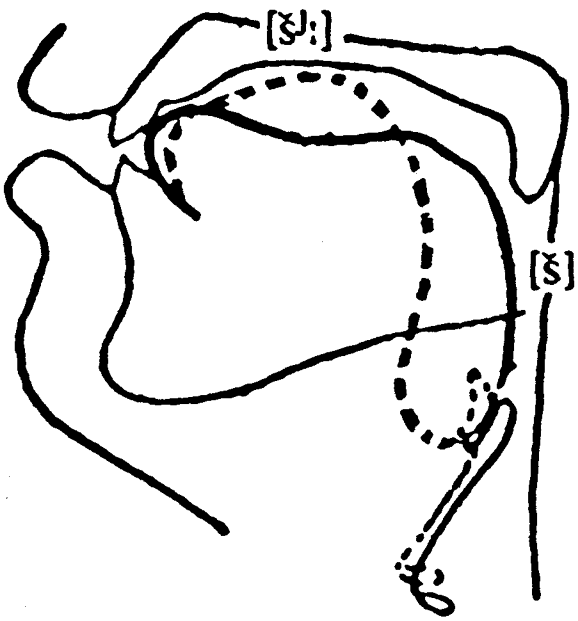
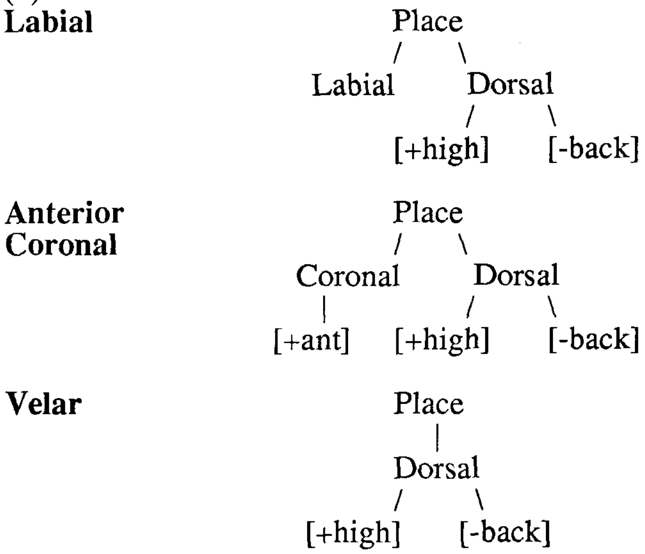


**Figure 7:** Articulatory data on palatalized and nonpalatalized š, šʲː [after Matusevič and Ljubimova, 1964]. Superimposed tracings of sagittal X-rays of nonpalatalized (solid line) and palatalized (dashed line) articulations.

(1)



All three places share [+high], [-back] specifications where [-back] means in front of the soft palate. All three indicate Dorsum as the executing articulator. Labial and Coronal as shown here preserve the primary articulator specification. But Velar, by virtue of its [-back] specification under Dorsal, is automatically fronted. The fact that velars are already Dorsal produces an asymmetry in the result of palatalization here. A unitary articulation, secondary palatalization, produces complex segments from labials and coronals, but not velars, which are already Dorsal. Palatalized velars are not articulatorily more complex. They do not add an articulator specification. Instead, they add a dependent feature under the articulator specification they already have.

Now let's return to the important observation motivating proposals to ascribe palatalization to a Coronal articulation -- namely primary palatalization. Sometimes secondary palatalization of anterior coronals results in a primary palatalization or a shift in the primary place of articulation; we saw this for /t/ in Russian. Sometimes secondary palatalization of velars results in much more extreme fronting, to a non-anterior coronal. This is the case in the lexical phonology of Russian, for example. If palatalization consists of an attraction of the tongue body to the rear of the hard palate, why should primary shifts in the coronal region result? Here I will consider only the kinds of retraction of coronals discussed by Mester and Ito (1989).

We have seen how the pivoting of the tongue as it raises and fronts for secondary palatalization can retract and lengthen the primary constriction for an anterior coronal stop in Russian. The phonetic characteristics of the resulting segment are summarized in (2). These characteristics describe a number of segment types: not only primary coronals with secondary palatalization, but also palatals and alveolopalatals (Keating, 1991). These latter segment types can be the result of primary palatalization. The question, then, is how and why the phonetic interpretation of this structure as secondary palatalization sometimes gives way to an interpretation as primary palatalization.

(2)

```
                    Place
                    /   \
             Coronal      Dorsal
             /    \        /     \
        [-ant] [+lam]  [+high] [-back]
```

Consider Figure 9 again. When the secondary articulation involves tongue backing (shaded tongue), then the tongue blade is relatively unperturbed in forming the primary constriction. The two articulations can be clearly seen. But when the secondary articulation involves tongue fronting (unshaded tongue), it crowds the primary coronal articulation; the tongue blade cannot rise as much. Palatalization will roll the tongue forward and perhaps force down the tongue tip unless some effort is made to stabilize its position. The palatalization will be secondary, but it will affect the primary articulation somewhat. In the limit, that effect will be large, and a single long laminal constriction will be produced. This articulatory configuration would still be represented as in (2) above, but is phonetically different. The palatalization is now primary. In Figure 10 I schematize how two independent articulations aimed at adjacent stretches of the palatal surface (top) might be more easily produced by a continuous stretch of the tongue (bottom). The top configuration is not impossible -- it seems to be maintained in the Russian derived contrast -- but more languages seem to prefer something like the bottom. In this I agree with Recasens (1990) that an articulation like that at the bottom of Figure 10 is likely to be inherently simpler to execute than that at the top, but it must be admitted that this is speculation.

Two kinds of coronals embody this simplied version of (2). One is the alveolopalatal [ɕ] (as in Japanese, Polish, and elsewhere). The other is the palatal [ç]. Both of these are different from two articulations that are even simpler: laminal palato-alveolar [ʃ] and apical retroflex [ṣ]. These latter are simple [-anterior] coronals, without the Dorsal component. These further simplifications (especially the laminal [ʃ]) would be the endpoint of the simplifications that could follow secondary palatalization. In sum, the continuum of coronal consonant types under palatalization would be as in (3):

**(3)** secondary palatalization, little effect on coronal articulation
secondary palatalization, features of coronal articulation affected
primary and secondary articulations appear merged into primary palatalization
loss of Dorsal component

## Fronting

Given the description of palatalized velars offered here, involving a fronting of the primary velar constriction, the question arises as to whether the representation for palatalized velars given above and repeated in (4) is also a description of fronted velars, as in the English word "key."

**(4)**

```
                Place
                  |
                Dorsal
                /     \
          [+high]    [-back]
```

Although in earlier work (Keating 1988a) I did indeed propose this description for English fronted velars, in subsequent work (Keating and Lahiri 1993) the contrary position was

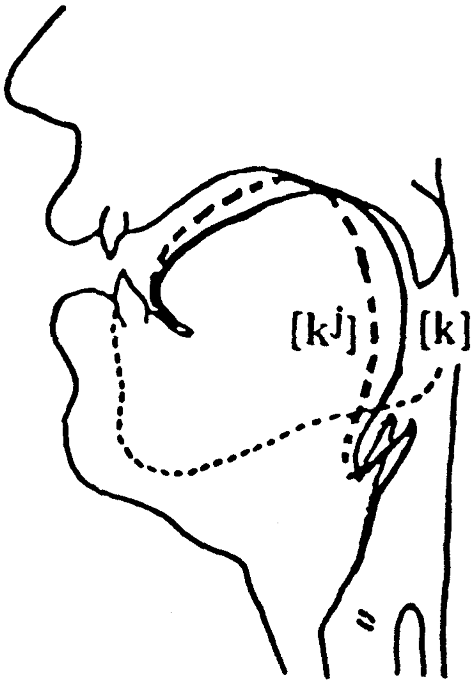


**Figure 8:** Articulatory data on palatalized and nonpalatalized k, kʲ [after Oliverius, 1974]. Superimposed tracings of sagittal X-rays of nonpalatalized (solid line) and palatalized (dashed line) articulations.

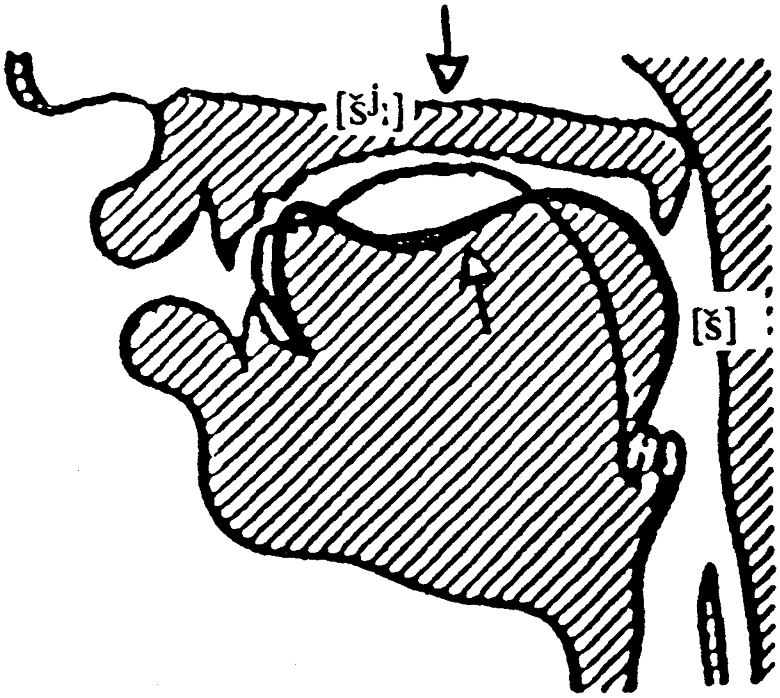


**Figure 9:** Articulatory data on palatalized and nonpalatalized š, šʲː [after Akishina and Baranovskaja, 1980]. Superimposed tracings of sagittal X-rays of nonpalatalized (shaded) and palatalized (unshaded) articulations.

**Figure 10:** Schema of how two separate articulations with blade and body (top) could simplify as a long continuous articulation (bottom).

taken. There were two reasons for concluding that palatalized and fronted velars are distinct phonetically. First, an acoustic comparison of English and Russian indicated that the fronting of Russian palatalized velars is more extreme than the fronting of English fronted velars. Second, existing articulatory studies of English velars suggest that they lack inherent specification for Back. Houde (1967) describes tongue body contact on the palate for velars as moving along the palate during closure. A velar before a front vowel is fronted by the moment of release, but not at the onset of the closure (similarly for Swedish, Ohman, 1966). Houde showed that the constriction location of a velar moves during the closure interval as a function of the contextual vowels. Figure 11 schematizes velar closure between back and front vowels: the velar constriction is first formed in a backed position because it follows a back vowel, then moves forward during the closure to a fronted position appropriate for the following front vowel.

This phenomenon is in fact familiar from spectrogram reading, where it often introduces errors of segmentation. A velar between two vowels which differ in backness will often be taken for a cluster of two consonants which differ in place of articulation. Figure 12 is a spectrogram of an utterance with two [g]s. For velars, the frequency at which the second and third formants converge, and the frequency of the major energy in the release burst, give an acoustic indication of backness; higher frequency values indicate a more fronted tongue position. In this utterance, as shown with arrows, the first [g] begins with a more fronted constriction but is released further back (and also more rounded). The second [g] begins with a more central constriction but is released further forward. The change in constriction location follows from the asymmetry of the vowel context in both cases. In Keating et al. (1992) we showed that the effects of a vowel are equally strong on the formant frequencies of preceding and following velar consonants.

Stated another way, velar fronting is something that happens gradually over the course of the velar. Such temporal/spatial variation, or phonetic gradience, can be interpreted as a transparency effect of the velar with respect to Backness. In earlier work (Keating 1988b), I have suggested that phonetic gradience indicates the lack of any phonetic feature value. Consider the similar behavior of /h/, as shown on the spectrogram in Figure 13. The tongue moves throughout; we can say that /h/ has no value for the feature Back of its own, its physical backness trajectory being determined by the vowel context. The velar case is exactly parallel, it seems. The schematic given in (5) shows how underspecification of Back, and the resulting transparency of velars, results in phonetic rules determining how the tongue moves in the backness dimension during velars. On this account we can conclude that English velar fronting should receive no featural representation at all, and in that respect is different from palatalization.

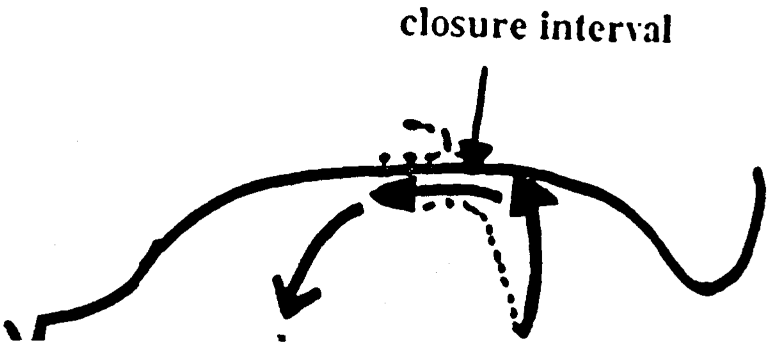


**Figure 11:** Schema of tongue movement during **k** closure and release in a back vowel-front vowel environment.

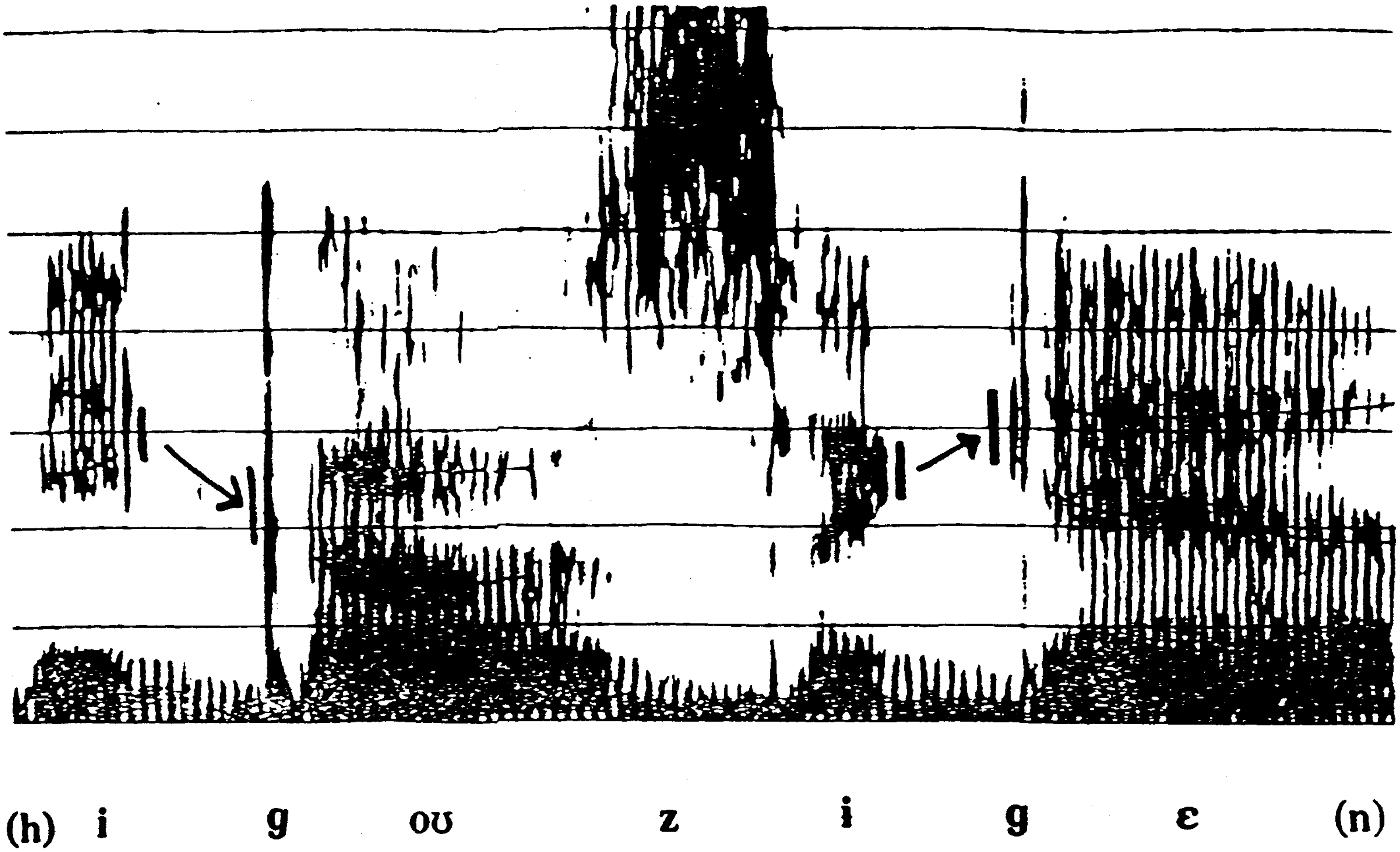


Figure 12: A spectrogram of the sentence "He goes again."

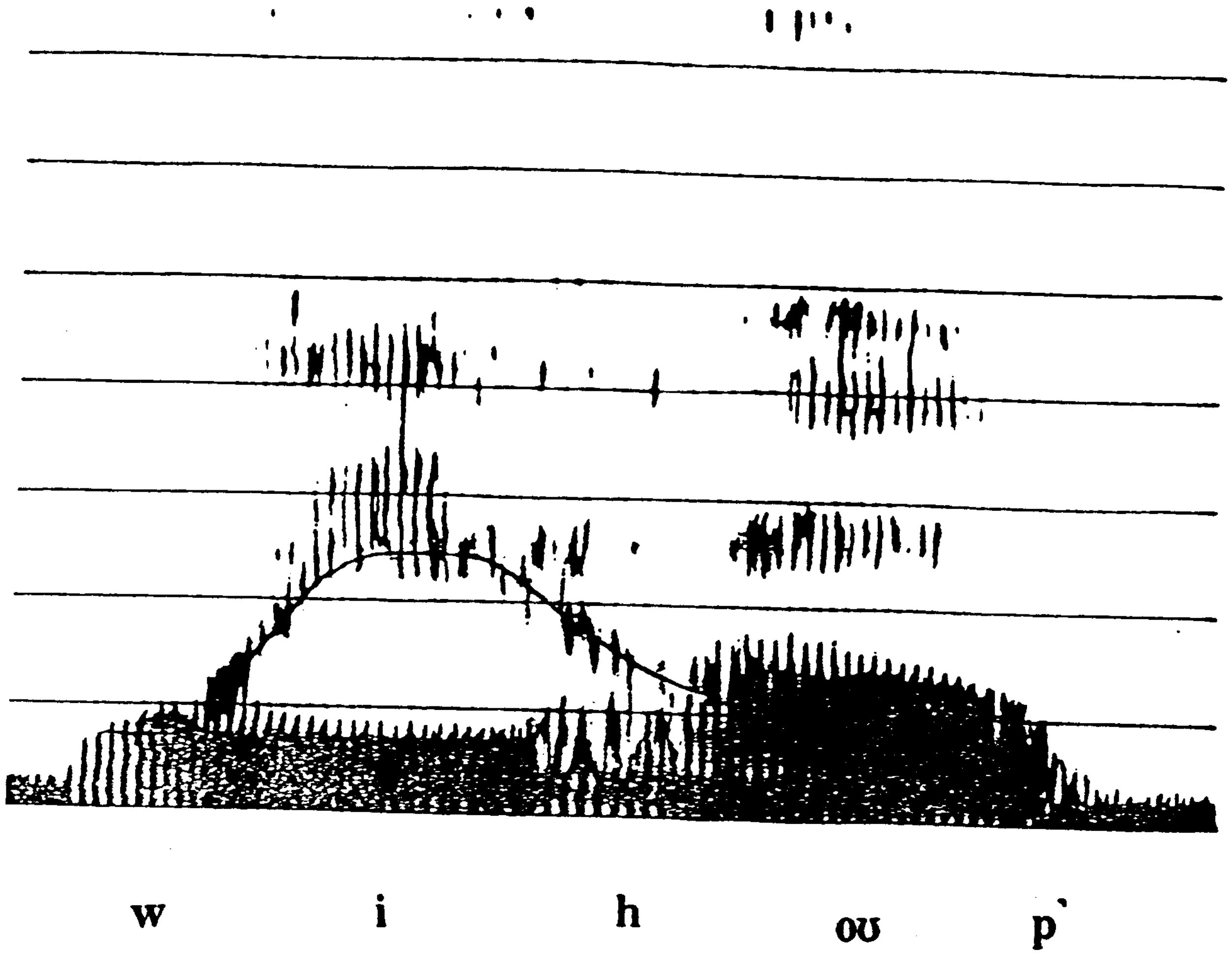
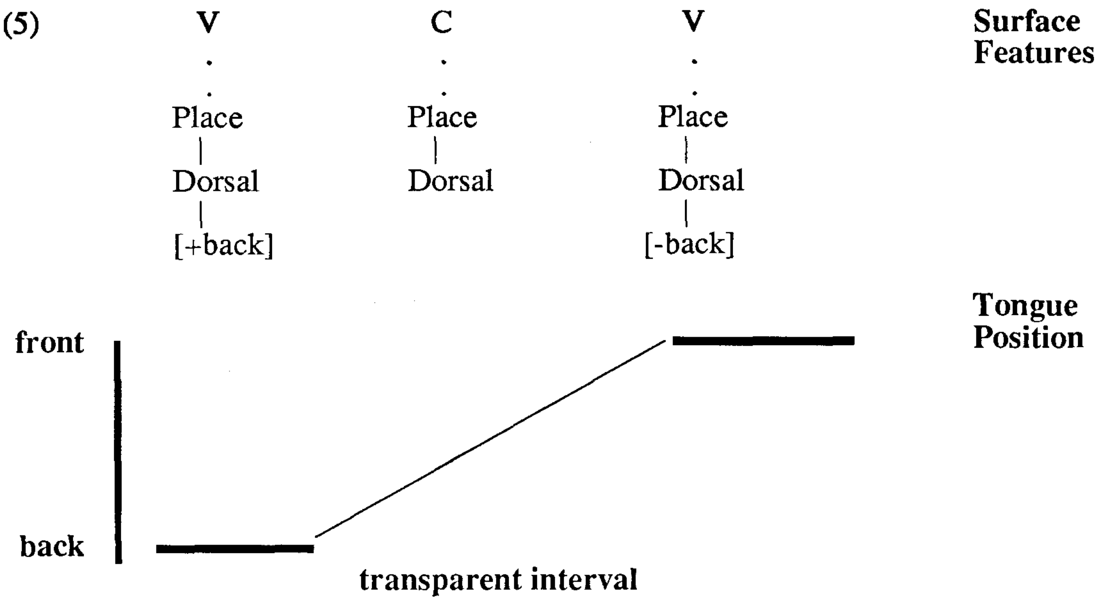


**Figure 13:** A spectrogram of the sentence "We hope."

(5)



## Conclusion

We have seen that secondary palatalization has a rather simple articulatory characterization that seems to refer to the tongue body as the active articulator. It is this kind of phonetic evidence that leaves me unenthusiastic for the proposal that palatalization is an articulation of the tongue blade, and more sympathetic towards the kind of tongue body accounts of e.g. Gorecka (1989) or Lahiri and Evers (1991). The important motivating observation of the tongue blade proposals, that secondary palatalization of coronals and velars often results in primary place shifts, certainly needs to be accounted for, but this observation must somehow be the consequence of the combination of secondary tongue body activity with a primary articulation, rather than being the mechanism of palatalization itself.

## References

Akishina, A. A. and S.A. Baranovskaja (1980) *Russkaja fonetika.* Izdatel'stvo Russkij Jazyk, Moscow.

Bhat, D.N.S. (1978) A general study of palatalization. In J.S. Greenberg, ed. *Universals of Human Language,* Vol. 2: *Phonology,* pp. 47-92, Standford: Stanford University Press.

Bolla, K. (1982) A phonetic conspectus of Russian: the articulatory and acoustic features of Russian speech sounds. *Magyar Fonetikai Fuzetek (Hungarian Papers in Phonetics)* 11. MTA Nyelvtudomanyi Intezete, Budapest.

Clements, G.N. (1991) Place of articulation in consonants and vowels: a unified theory. In B. Laks and A. Rialland, eds., *L'architecture et la geometrie des representations phonologiques.* Paris: Editions du CNRS.

Gorecka, A. (1989) *Phonology of articulation..* PhD dissertation, MIT.

Houde, R.A. (1967) *A study of tongue body motion during selected speech sounds*. University of Michigan PhD dissertation.

Keating, P. (1988a) Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics* 69: 77-91.

Keating, P. (1988b). Underspecification in phonetics. *Phonology* 5:275-292.

Keating, P. (1991) Coronal places of articulation. In Paradis and Prunet, eds. *The Special Status of Coronals. Phonetics and Phonology* 2, pp. 29-48, San Diego: Academic Press.

Keating, P. (1992) B. Blankenship, D. Byrd, E. Flemming, and Y. Todaka. Phonetic analyses of the TIMIT corpus of American English at UCLA. *UCLA Working Papers in Phonetics* 81: 1-16.

Keating, P. and A. Lahiri. (1993) Fronted velars, palatalized velars, and palatals. *Phonetica* 50: 73-101.

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

Lahiri, A. and V. Evers (1991) Palatalization and coronality. In Paradis and Prunet, eds. *The Special Status of Coronals. Phonetics and Phonology* 2, pp. 79-100, San Diego: Academic Press.

Matusevič, M.I. & N. A. Ljubimova. (1964) Artikuljatsija russkix zvukov pod udareniem na osnove rentgenografičeskix dannyx. *Serija filologičeskix nauk* 69: 37-44. (Leningrad University).

Mester, A. and J. Ito. (1989) Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language 65*: 258-293.

Öhman, S.E.G. (1966) Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am. 39*: 151-168.

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

Recasens, D. (1990) The articulatory characteristics of palatal consonants. *Journal of Phonetics* 18: 267-280.

# Investigating Ewe articulations with electromagnetic articulography

Ian Maddieson

*Paper presented at Workshop on Electromagnetic Articulography, Institut für Phonetik und Sprachliche Kommunication, Universität München, April 19-21 1993.*
*To appear in the Institut's* Forschungsberichte

## Introduction and issues

The majority of the participants in this symposium on electromagnetic articulography have been involved in the development or refinement of this important new technique, either in a very direct way or as a result of their attachment to one of the pioneering laboratories or companies creating the hardware and software. Through the generosity of Joe Perkell and with the essential help of Melanie Matthies and Mario Svirsky of the Speech Communication Group, Research Laboratory of Electronics at MIT, I have been one of the first "ordinary working phoneticians" to make use of the technique as a straightforward investigative tool. It is clear that it provides an extremely useful method of collecting substantial bodies of articulatory data in a comparatively short time. The results show not only that it provides an excellent method of testing existing hypotheses about particular aspects of speech production, but that it is capable of revealing unsuspected articulatory details that call for the development of further hypotheses. All this can be done with much less expense and with far less inconvenience to subjects than is involved with the nearest comparable technique, the x-ray microbeam.

The particular investigation to be reported here concerns two of the most interesting phonetic properties of the Ewe language. Ewe is spoken by about a million people principally in the southeastern part of Ghana and in southern Togo. It is closely related to Fɔ̃, Gɛ̃, and Gũ spoken in adjoining parts of Benin and Nigeria, and many scholars view this whole dialect chain as a single language (Capo 1991, Stewart 1989). The particular variety of the language examined is the Aŋlɔ dialect as spoken in the town of Kpando and neighboring villages in the Volta Region of Ghana. Among the more important descriptions of Ewe are those of Westermann (1930), Ansre (1966) and Stahlke (1970). The most comprehensive dictionary remains Westermann's (1954). The language is classified as a member of the Kwa branch of the Niger-Congo language family (Stewart 1989).

Ewe is one of the many West African languages with the interesting phonetic property of having doubly-articulated labial-velar stops, a type of sound only known to occur in Western and Central Africa and in a few languages of New Guinea. These segments are usually written /kp/ and /gb/ (or /k͡p/, /g͡b/) and analyzed straightforwardly as plosives with simultaneous bilabial and velar closures. In describing their production for European learners of these languages, Westermann and Ward (1933: 58) laid particular emphasis on the simultaneity of the two component articulations; for example their comment on /kp/ reads: "the two articulations must be simultaneous, i.e. when the sound occurs between two vowels, there must be no onglide to the **k** heard before the lips come together for the **p** position." Ladefoged (1968: 9), who provided detailed descriptions of labial-velars in a number of West African languages, also talks of their production as involving "the simultaneous articulation of the **k** and **p** or **g** and **b**".

Despite the stress that has been placed on the question of simultaneity of articulation, when these segments occur intervocalically the auditory impression is often that the transition from the preceding vowel to the consonant has a primarily velar character, while the release from the consonant into the following vowel has a primarily labial character. This is especially clear in our Ewe data. The asymmetry of the consonant transitions can often be seen quite clearly in acoustic records, such as the spectrogram shown in Figure 1. A rising F2 transition and a lowering F3 transition can be seen at the consonant onset, as is typical of velars, whereas the release of the stop is followed by a rising movement of both these formants from a low origin, as is more characteristic of bilabials. We have confirmed in an informal experiment that this asymmetry has perceptual significance: When just the preceding vowel and the onset transition are presented to listeners, they have no hesitation in identifying the upcoming consonant as velar. When just the release and the following vowel are played, they have no hesitation in identifying the consonant as bilabial.

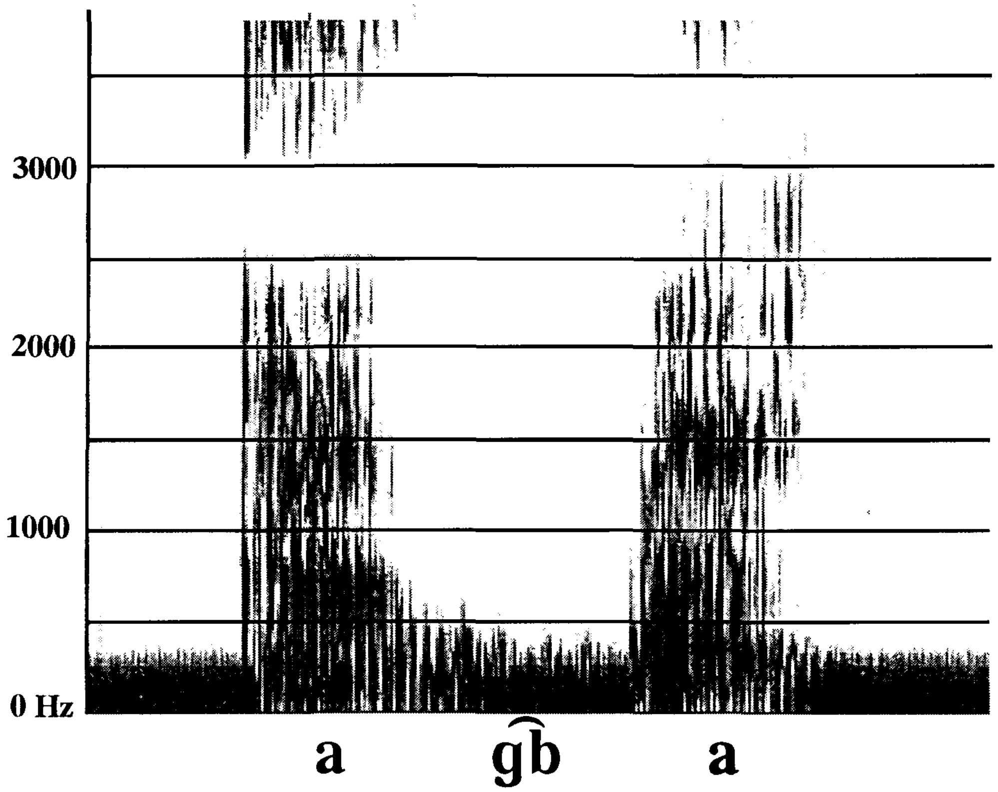


Figure 1. Spectrogram of the Ewe word /ag͡ba/ “load, trouble”.

Maddieson and Ladefoged (1989) suggested that the acoustically and auditorily distinct transitions into and out of labial-velar stops result from planned articulatory non-simultaneity (rather than, say, being a consequence in some way of the fact that the outward air-flow meets the velar closure first and encounters the labial one last). They proposed that the velar and labial articulatory gestures that constitute the doubly-articulated stops are typically not completely simultaneous, but are timed so that the velar closure is formed slightly before the bilabial one, and the bilabial one is released slightly after the velar one. This timing would achieve a clearer auditory distinction between the labial-velar stops and the singly-articulated velar (or bilabial) stops with which they might be confused if the articulations were in fact entirely simultaneous.

Although Ladefoged (1968) and Connell (1987) have studied the articulation of labial-velars by the indirect method of examining aerodynamic conditions surrounding their production, no previous studies have looked directly at their articulatory control. Electromagnetic articulography makes the direct examination of the relative timing of articulatory gestures relatively easy.

Ladefoged observed that the air in the oral cavity between the labial and velar closures underwent some rarefaction in many tokens he recorded. Based on these observations, he inferred that "during the labial and velar closures the back of the tongue moves slightly further back, creating a slight suction effect as in a click" (Ladefoged 1993: 164). The appropriateness of this interpretation is also open to direct confirmation.

Ewe is also well-known as one of a comparatively small number of languages which include both labiodental and bilabial fricatives in their repertoire of consonants (Ladefoged 1968, 1990). Ladefoged has suggested that because these two articulations need to be distinguished, a more extreme gesture is used to produce the labiodental fricative than is found in languages (such as English) which lack bilabial fricatives (Ladefoged 1993: 269). Specifically Ladefoged suggests that the labiodental fricatives involve a raising and fronting of the upper lip to enhance the differentiation between them and the bilabial fricatives. A rather different positioning of the upper lip for bilabials and labio-dentals is apparent in the photographs of a speaker of Logba (a neighbouring language to Ewe) in Ladefoged (1968) but as these are still photographs, it is not possible to determine if there is any active movement of the upper lip in labio-dentals. Thus, in addition to the study of labial-velar stops, the articulation of bilabial and labio-dental fricatives was also examined in the present experiment.

## Equipment and Data

Data was collected from two speakers of Ewe from Kpando using the electromagnetic articulography system at MIT. This system, known as EMMA for Electro-Magnetic Midsagittal Articulometer, is described in Perkell et al (1992) as well as in a contribution to this volume [i.e. the Munich conference report] and is illustrated in a videotape prepared for the Acoustical Society of America (Perkell et al 1993). Discussion of the important issues involved in obtaining reliable data from systems of this sort, and some description of the post-processing of data are contained in these sources. Despite the availability of these descriptions of the system, a brief summary will be repeated here for the sake of self-sufficiency of this paper. The EMMA system uses three transmitter coils positioned around the head of a subject in a helmet as shown in Figure 2. This figure is also included here to emphasize the fact that for all results presented in this paper the speaker is assumed to be looking in the direction illustrated; that is, the front is to the left of the page, as is traditional in phonetic vowel and consonant charts.

The transmitter coils generate alternating magnetic fields at three different carrier frequencies (all relatively high). Two or more receiver coils are positioned in locations (typically the bridge of the nose and the upper incisors) that are immobile with respect to the skull. These provide reference positions for correcting for movement of the head during recording. Further receiver coils are placed on the moveable articulators of interest. In the present experiment five coils were used, placed as close as possible to the midsagittal line and held in position by a surgical adhesive. The locations were on the outer surface of the upper lip and the lower lip near the vermilion borders to monitor lip movement, near the base of the lower incisors to record jaw movement, on the front of the tongue at the back of the blade region, and on the back of the tongue at a location judged to be in the area involved in contact for velar consonants. These locations are sketched in Figure 3.

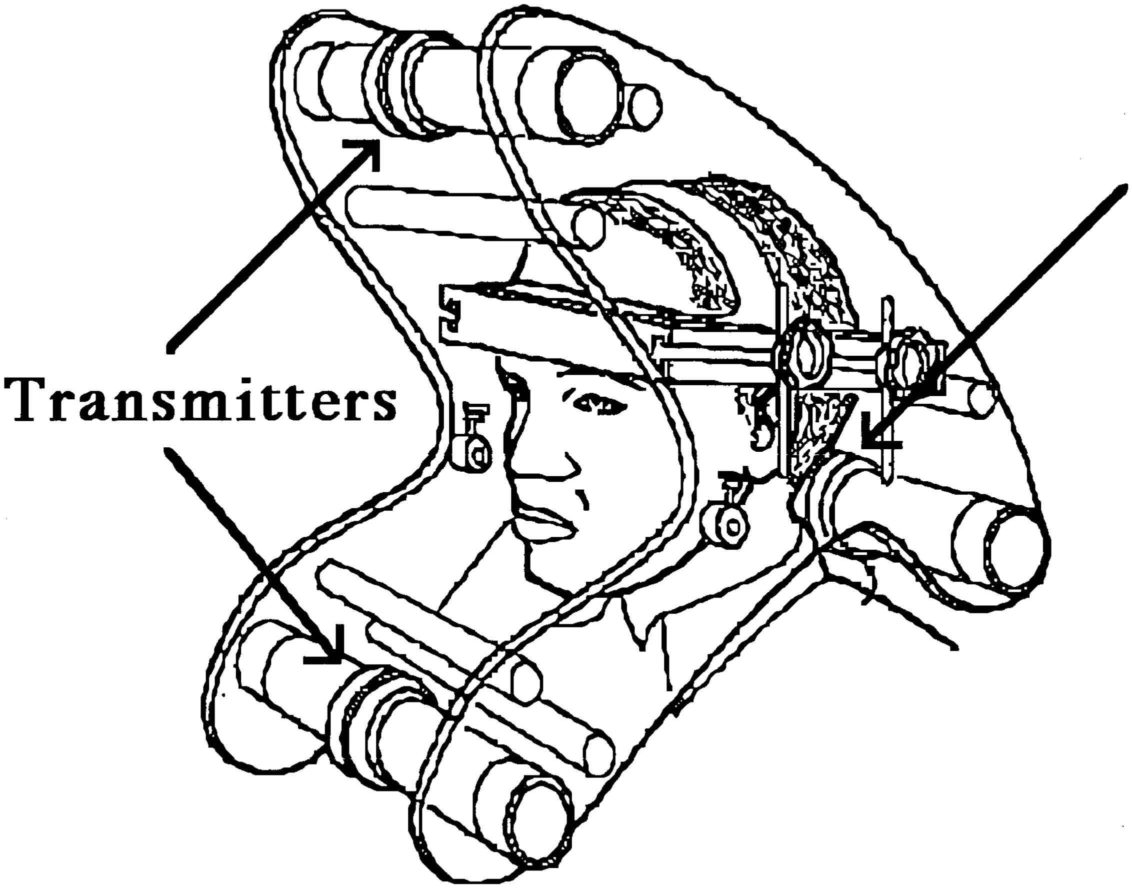


Figure 2. Schematic illustration of the EMMA helmet around a subject's head, showing position of the three transmitter coils.

The signals received from the transmitters by each of the receivers are separated into components at the different alternating frequencies and the low-frequency amplitude changes in these components are recorded direct to computer at 312.5 Hz. The amplitudes of the signals are related to the distance from the transmitters, and by triangulation can be converted into a spatial location. Post-processing rotates the axes of the location space into horizontal and vertical components (x and y co-ordinates) with respect to the occlusal plane, using the reference receivers as 'anchors'. Simultaneously with the recording of the articulatory data, a high-quality audio signal is digitized at 10KHz, so that acoustic and articulatory events can be related to each other. Output from the system is in the form of two data channels for each receiver of interest giving displacement in x and y directions respectively, plus a third channel which gives an estimate of error magnitude. Interpretation of this error signal is discussed in Perkell et al. (1992). The error signal remained within acceptably small range during collection of all the data discussed in this paper.

Ten repetitions of the three sets of words shown in Table 1 spoken in two different carrier phrases were recorded. (In the table the bilabial fricatives are represented by the symbols '*ƒ*' '*v*' which approximate the shape of the letters used in the Ewe orthography.)

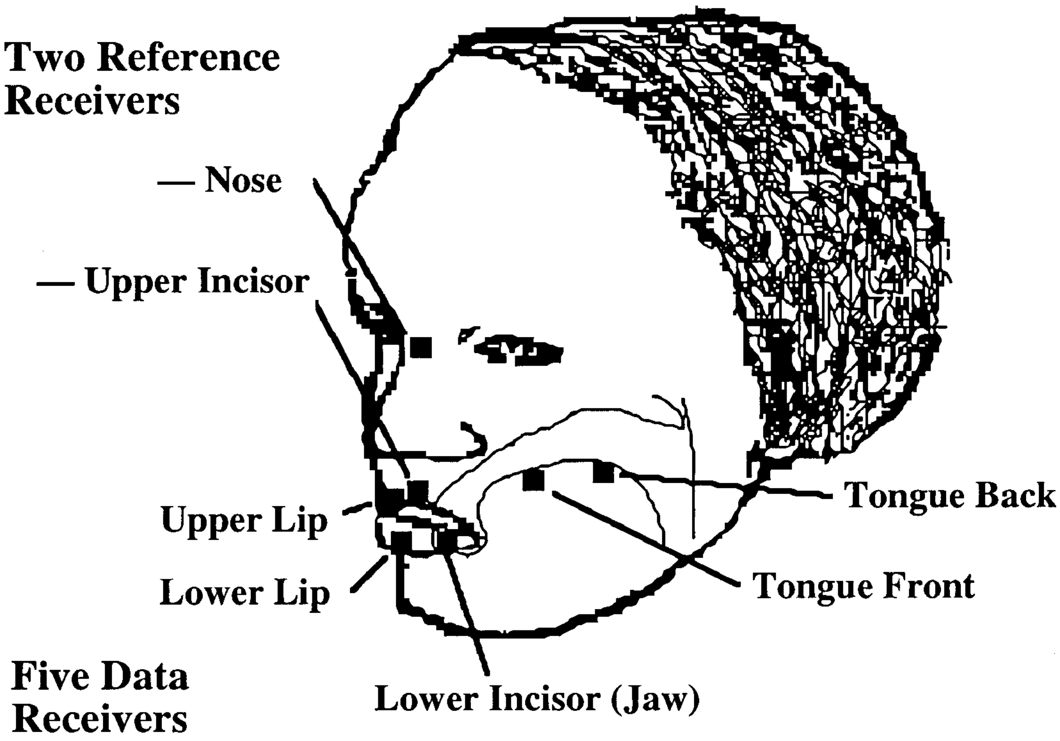


Figure 3. Schematic representation of the location of the two reference receiver coils and the five coils placed on moveable articulators.

Table 1. Wordlist used in Ewe recording.

| ***Labial and velar articulations*** | |
|---|---|
| apaá | "job, contract labor" |
| abaá | "mat" |
| amaá | "green vegetables" |
| aká | "charcoal" |
| aga | "slope, stairs" |
| ákpá | "too much" |
| àgbà | "load, trouble" |

| ***Velar coarticulation*** | |
|---|---|
| eké | "sand" |
| ekú | "seed" |
| ege | "beard" |
| e gu | "it collapsed, failed" |
| ekpé | "stone" |
| ekpo | "log" |
| egbe | "language" |
| gbugbo | "suck" |

| ***Labial fricatives*** | |
|---|---|
| efe | "nail;debt" |
| efú | "pregnancy" |
| éfá | "it's cold" |
| fufu | "fufu" |
| eve | "two" |
| evu | "fight, argument" |
| vavá | "coming, arrival" |
| eƒe | "year" |
| ƒúƒú | "dryness" |
| eƒú | "bone" |
| aƒá | "shout" |
| e*v*e | "Ewe people" |
| e*v*u | "stomach" |
| a*v*a | "war; penis" |

Speakers read a series of randomized lists of phrases containing words from one of the sets in one of the carrier phrases. Several such blocks of data were printed in large type (ten to a page) on sheets of paper and suspended at eye level in front of the speaker. Pages were turned by one of the experimenters, and the subject read the page number before reading the phrases in order to facilitate place-finding in the recorded data. Although data was obtained from two speakers only the words spoken in the carrier phrase /___ e mi/ "There's a ___" from one speaker have been analyzed so far.

Following completion of the recording and basic post-processing, a number of derivative measures, such as the distance between the upper and lower lip receivers, and the velocities and accelerations were also calculated, and added as additional tracks to the main data file.

Individual tokens were then examined using a specially written MITSYN program. Onset and release of closure were determined for each of the words containing stops and the center of the friction duration was marked for words containing fricatives on the basis of examination of the acoustic records, using displays such as that seen in Figure 4. The top panel shows an amplitude envelope over a long interval (containing ten phrases and a page number in this example). Below is a waveform of the selected interval outlined in black in the amplitude window (a token of /akpa/). Precise location of the cursor in the waveform display is shown in a smaller expanded display to the right (it is placed at the release burst - brief prevoicing is apparent). Several of the movement tracks are shown in the lower part of the display.

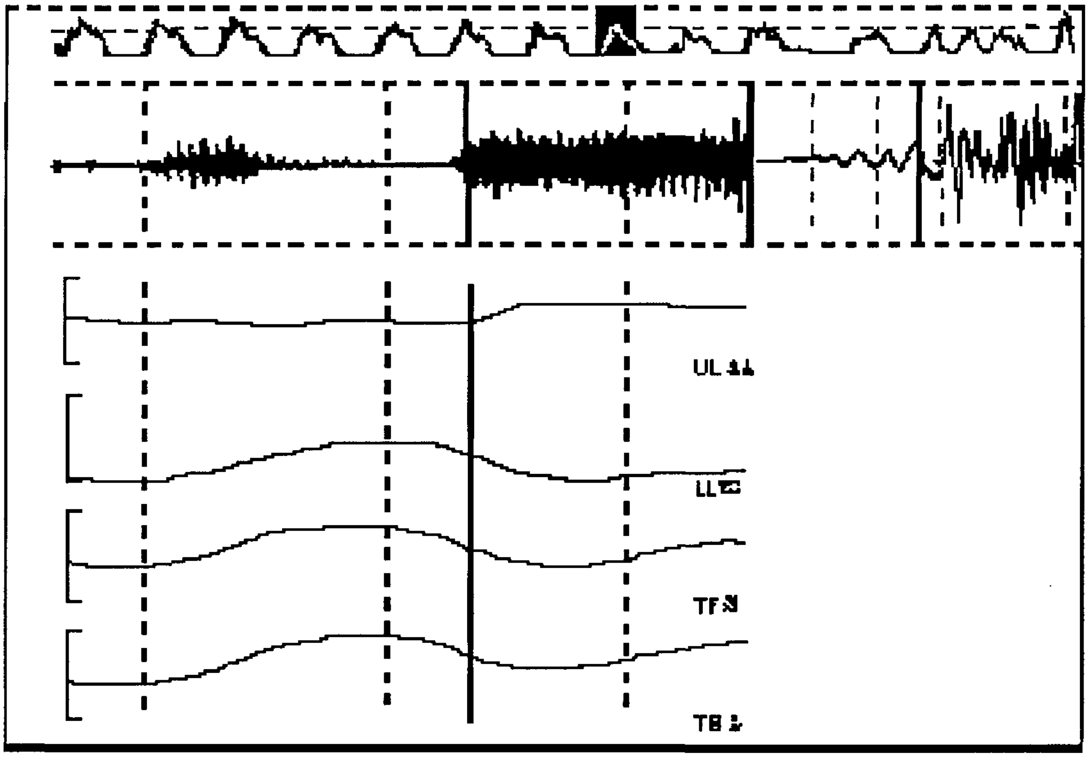


Figure 4. Data display used for locating time-points for alignment of repetitions. Continuous cursor line marks consonant release.

Following the marking of time points in all tokens the movement tracks were averaged for ten repetitions of each item, aligned at the instant of acoustic release for stops, and at the center of friction duration for fricatives. Aligned traces for the inter-lip distance measure for five repetitions of /apaa/ are shown in Figure 5. The constancy of movement traces across repetitions of the same utterance can be seen in this figure. Ensemble averages of the movement tracks for upper lip, lower lip, jaw, tongue front and tongue back were computed, and it is these ensemble averages that will form the primary basis of the discussion of results below.

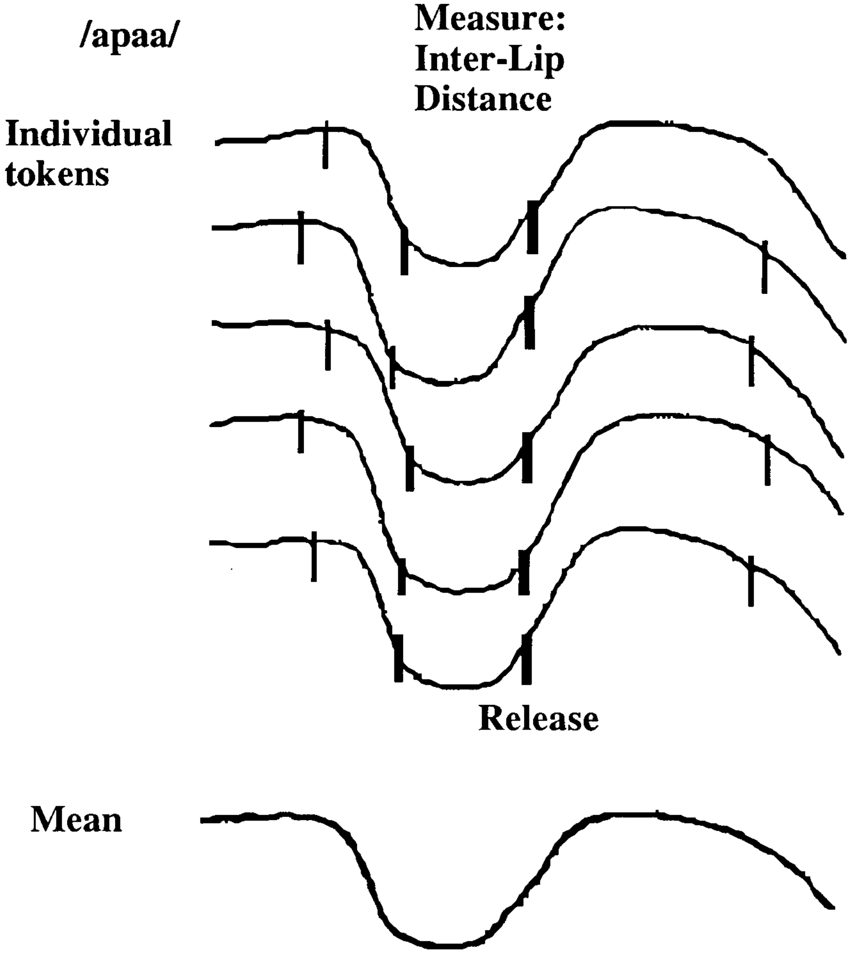


Figure 5. Alignment and averaging procedure for movement tracks across repetitions.

### Labial-velar stops

With respect to the labial-velar stops, the basic hypothesis investigated is an extreme statement of the Maddieson and Ladefoged claim that inter-articulator timing is crucial in the production of labial-velars. It can be expressed as saying that there is nothing special about the articulations involved; they are the same as those used in singly-articulated bilabial and velar stops, but they must be crucially coordinated in a particular way. This strong general hypothesis was broken down into four more specific hypotheses as follows:-

Hypothesis 1. The (slightly) longer duration of the doubly-articulated stops occurs because the velar gesture leads the bilabial gesture by the amount of the added duration, not because the individual gestures are lengthened.

Hypothesis 2. The component gestures of a doubly-articulated labial-velar stop are (essentially) identical in time-course and magnitude to the single gestures of the corresponding simple bilabial and velar stops. (This reference to gestures should not be taken to mean that movement traces will necessarily be identical, as the movement traces can reflect the influence of several gestures at the same time. Differences in concurrent conditions may also affect movement traces in predictable ways.)

Hypothesis 3. Any rarefaction of air between the labial and velar closures is due to the tongue movement required to create and break the velar closure (i.e. is a by-product of the overlap of the articulations, not a separately planned characteristic).

Hypothesis 4. Coarticulation of the velar component of a labial-velar stop with the surrounding vowels is similar to that observed with simple velar stops.

**Timing of movements in labial-velars**

The temporal asymmetry of the two articulations postulated in hypothesis 1 can be seen in Figures 6 and 7. These figures show the vertical displacements over time of the tongue back and the lower lip for /akpa/ and /agba/ respectively. The vertical scale is normalized so that comparisons can be made more easily between the two movements. For both /akpa/ and /agba/ the raising of the tongue back occurs faster and peaks earlier than the raising of the lower lip. The downward movement of the lip is faster than its upward movement.

The closure durations of bilabial, velar and labial-velar stops, measured from the acoustic waveform displays of ten tokens each, are shown in Table 2. Velar stops are shorter than bilabial ones, and labial-velars are 25-30 ms longer than bilabials.

Table 2. Consonant closure durations in /aCa/ context (10 repetitions).

| | Voiceless | | Voiced | |
|---|---|---|---|---|
| Velar | k | 142 | g | 133 |
| Bilabial | p | 158 | b | 150 |
| Labial-Velar | k͡p | 174 | g͡b | 179 |

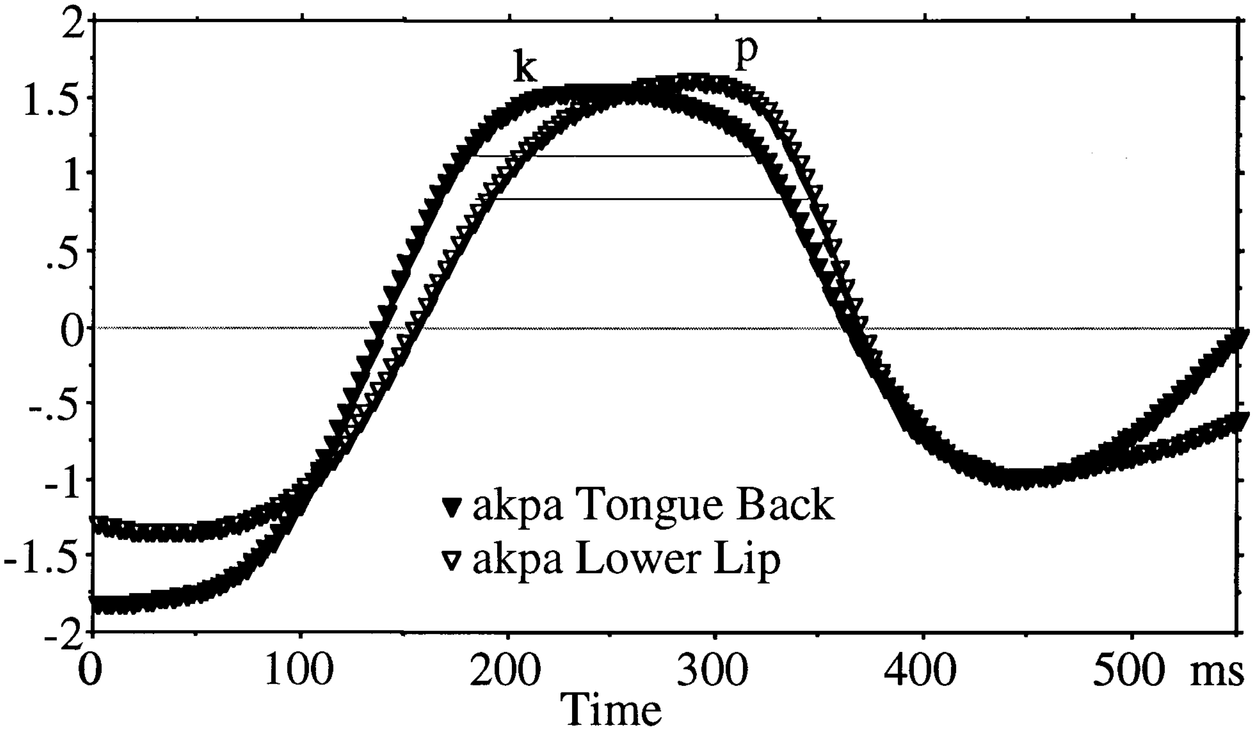


Figure 6. Coordination of lower lip and tongue back movements in /akpa/ (vertical displacement, normalized scale, mean of 10 tokens aligned at release).

In both Figures 6 and 7, the consonant release is at 350 ms. Two horizontal lines have been drawn on these figures. One connects the lower lip height at the time of release on the labial movement trajectory to the same height on the closing phase of the lip movement. A similar line connects the tongue back height at the mean time of onset of closure (derived from Table 2) to the same height on the downward trajectory of the tongue back. Using these lines to visualize the likely durations of contact, several conclusions are suggested. A greater proportion of the lip movement than of the tongue back movement involves tissue compression (the labial line is lower than the velar one). The duration of velar and labial contact are more equal in /akpa/ than in /agba/. The releases of the two articulations are more temporally separated in /agba/ than in /akpa/. Most obviously, however, the velar closure leads the labial one.

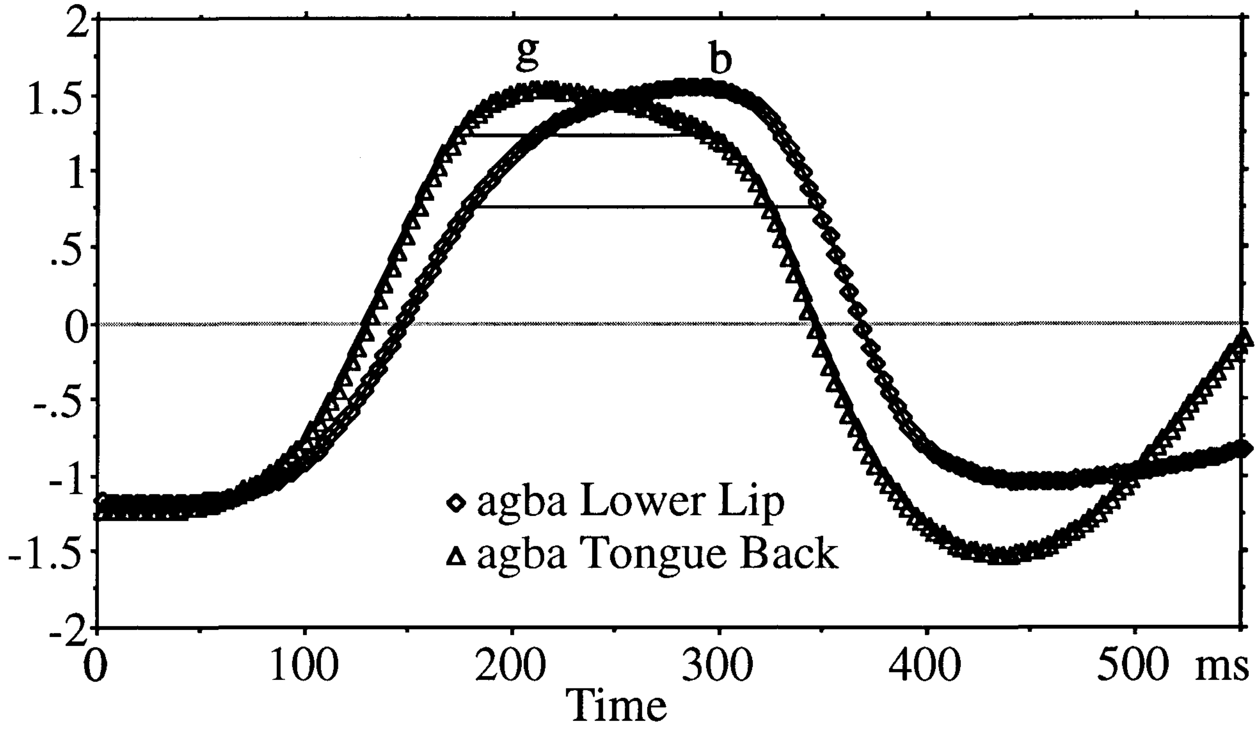


Figure 7. Coordination of lower lip and tongue back movements in /agba/ (vertical displacement, normalized scale, mean of 10 tokens aligned at release).

Estimates of the closure durations of the velar and labial components of the labial-velar stops are given in Table 3. These were obtained in the following way. For the labial closure durations, the time point during the upward lower lip movement with the nearest articulator height to the height at labial release was subtracted from the release time (350 ms). For the velar closure the mean time at closure was subtracted from the time point with the nearest height to the height at closure on the downward movement of the tongue back. From these times the temporal offsets between the two closures and the two releases, and the net amount of added duration that might be attributed to these offsets were also calculated.

These estimates of course have a number of sources of error, including the initial assumption that closure and release occur at similar heights of the articulators. The plausibility of this assumption is supported by comparison between the height at closure onset and at release of simple bilabial and velar stops, where the time of both onset and release of closure are known from the acoustic records. For each of the four comparisons between onset and release height in the set /apaa/, /abaa/, /aka/ and /aga/ there is less than 1 mm difference; the difference between one sample point and the next in these phases of the movement tracks is typically over 0.5 mm (Recall that the sample rate provides a sample about every 3 seconds.) Comparing the data in Tables 2 and 3 shows that three of the four estimated closure durations for the component articulations in labial-velars are very similar to the durations of simple bilabial and velar stops of the same voicing category. The differences are 6 ms or less. The exception is the considerably longer duration of the estimated labial closure in /gb/ compared with the measured duration of /b/ (174 vs 150 ms). This extends the offset between the velar and labial releases in /gb/. Apart from this detail, the results of the above analyses provide general support for the hypothesis concerning the timing of labial-velars put forward above.

Table 3. Estimates of closure duration and timing offsets in labial-velars.

| | ak͡pa | ag͡ba |
|---|---|---|
| Estimated duration of velar closure | 148 | 130 |
| Estimated duration of labial closure | 164 | 174 |
| Offset between closures | 10 | 5 |
| Offset between releases | 26 | 49 |
| **Net offset** | **16** | **44** |

**Movement patterns in labial-velars and simple labial and velar stops**

Hypothesis 2, that no special articulatory maneuvers other than the correct temporal coordination is required, was investigated by comparing the production of the simple stops with the component articulations in the labial-velars. The articulations of the simple stops will be described and illustrated first in the following sections, including differences between stops at the same place that are associated with voicing. Subsequently, the articulations of the labial-velars will be described and compared with that seen in the simple stops. Two views will be presented of each consonant articulation; in one the vertical displacement of the five receiver coils over time will be plotted, in the other a trajectory of the receiver movement in the x-y plane will be shown. The major displacements are in the vertical direction, hence the first view provides a good way to look at the temporal coordination of the different moving structures, while the second view show the overall spatial path of the movements.

The articulatory patterns for bilabial stops and the bilabial nasal will be described first. Figures 8 and 9 illustrate the articulation of /apaa/. In these and following figures the movements during an interval of 550 ms around the consonant are shown. Consonant release occurs at 350 ms. The beginning of transitions to the following vowel /e/ in the carrier phrase can often be seen toward the end of the time interval shown. In Figure 8 the top trace shows the vertical movement of the upper lip (ULY), the next two traces are the tongue back (TBY) and tongue front (TFY) with the tongue front being marginally higher during the consonant. The fourth trace is the lower lip (LLY), and the lowest trace is the lower incisor (LIY).

Not surprisingly, the two tongue positions monitored show very little movement in the /apaa/ sequence; they remain low. The principal movement in producing the consonant is a large and relatively symmetrical 'bell-shaped' excursion of the lower lip upwards, assisted by a lesser movement of the jaw (shown by the lower incisor track). The upper lip shows relatively rapid movements that coincide with the formation and release of the labial closure with a relatively stable hold in between, rather than the bell-shaped movement shown for the lower lip. (Note that the two lip traces do not meet since the receivers are not placed on the inner surfaces of the lips, but rather on the outside.)

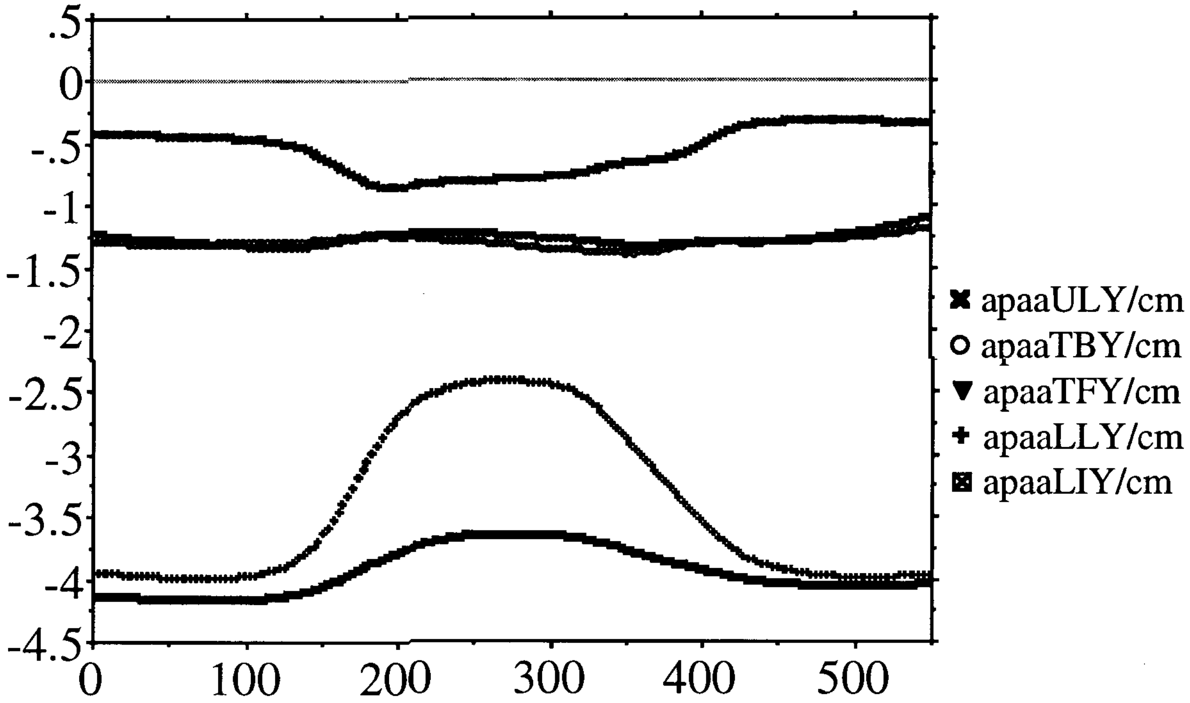


Figure 8. Mean vertical displacement over time of the five receiver coils during the production of ten repetitions of /apaa/. Vertical scale is in cm from an arbitrary but fixed origin. Time scale is in milliseconds. Acoustic release of the stop occurs at 350 ms.

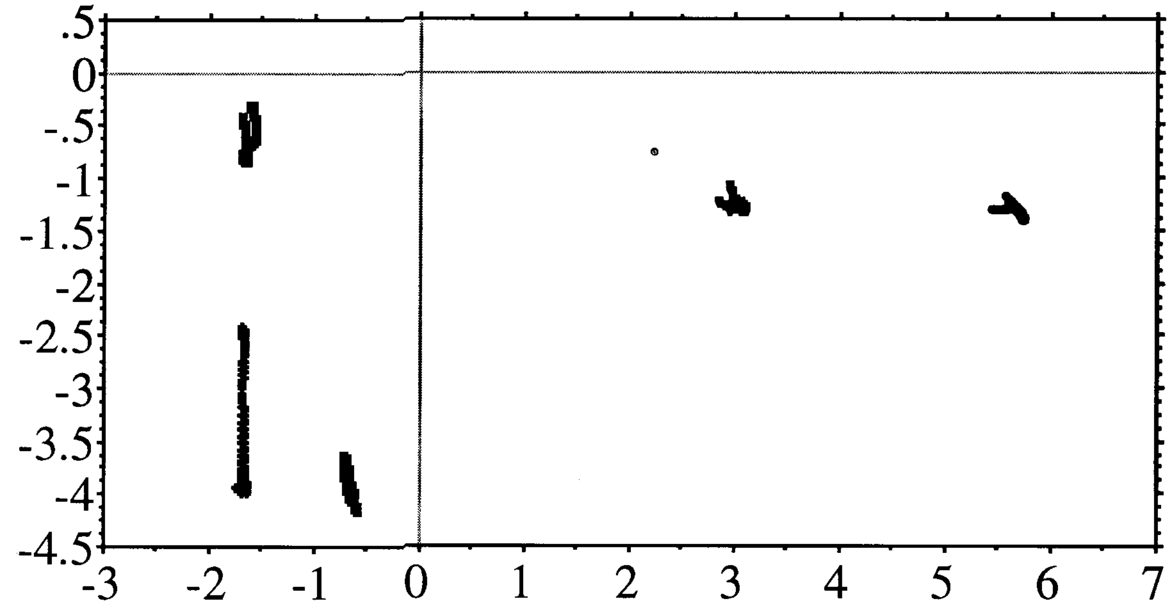


Figure 9. Displacement in two-dimensional space of the five articulatory points tracked, during /apaa/. Axes are in cm, to the same scale in both dimensions. The time interval covered is 550 ms as in Figure 8. The lips are to the left, tongue points to the right.

Figure 9 shows the spatial pattern for these movements. Following international convention, the front of the mouth is at the left. The small movements of the tongue result in two relatively tight clusters of points. The movements of the two lips towards each other have a very vertical orientation, whereas the jaw movement can be seen to involve a rotational component.

The corresponding movement traces for /abaa/ are shown in Figures 10 and 11. The labial movements for /b/ are extremely similar to those seen for /p/ except that the upper lip is lowered less at closure onset, resulting in an overall smaller excursion of the upper lip. The most salient difference between /p/ and /b/ lies in the lowering of the tongue back that occurs during the production of /b/. This movement begins at about the instant of closure formation and reaches its lowest point at about the moment of release. The rather smaller movement of the tongue front follows the tongue back trajectory and is presumably simply the result of the physical linkage between these two points. Since Ewe voiced stops are usually voiced throughout their closure, it seems likely that the presence of tongue lowering in /b/ but its absence is /p/ is part of a maneuver to enlarge the size of the supralaryngeal cavity by pressing the tongue still lower in the mouth so that transglottal air-flow can be sustained to maintain voicing. If so, then it would predict the absence of lowering in the production of /m/, since the nasal escape channel provides a route for the air and supraglottal cavity expansion is unnecessary.

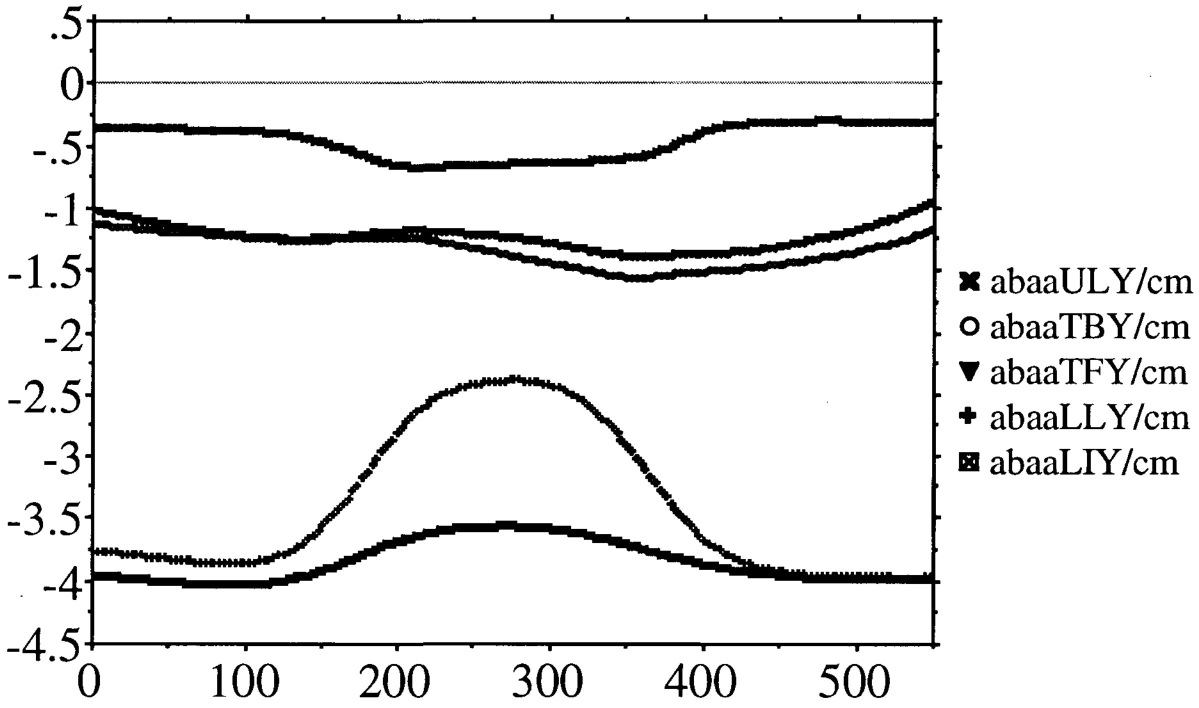


Figure 10. Mean vertical displacements over time of the five receiver coils during the production of ten repetitions of /abaa/. Consonant release occurs at 350 ms

The correctness of this prediction is shown by the movement traces for production of /amaa/ shown in Figure 12 since there is very little movement of either of the tongue points during the consonant. Interestingly, a more sharply defined upward movement of the upper lip occurs at the release than is seen for either /p/ or /b/.

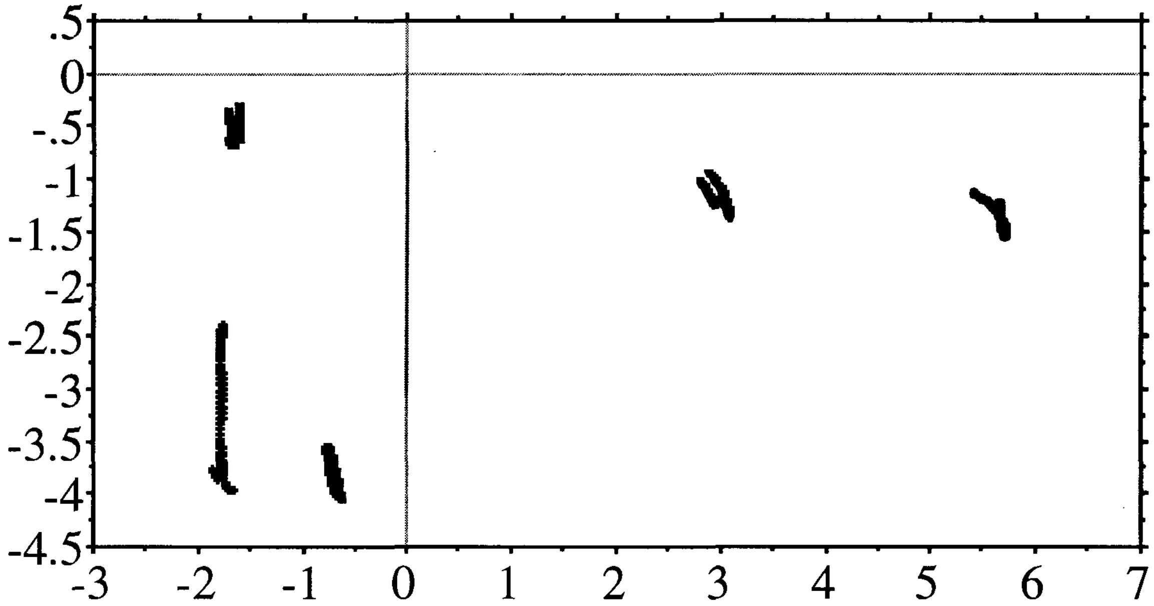


Figure 11. Mean displacements in two-dimensional space of the five articulatory points tracked during the production of /abaa/.

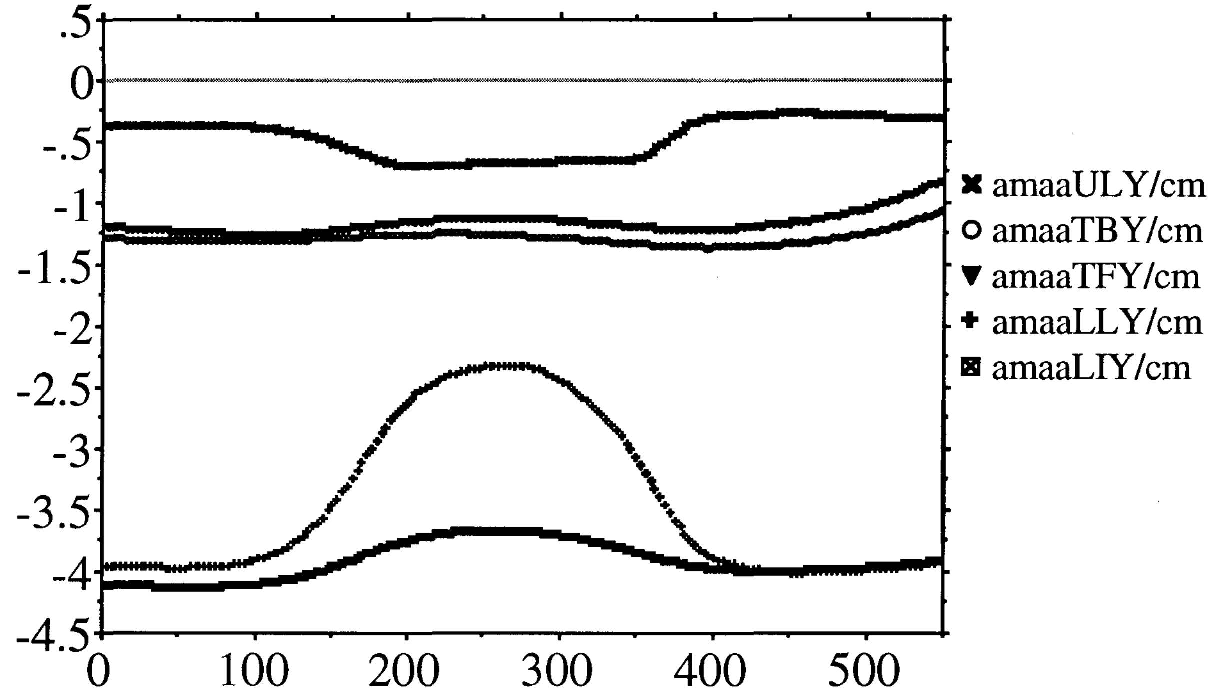


Figure 12. Mean vertical displacements over time of the five receiver coils during the production of ten repetitions of /amaa/.

In summary, the bilabial consonants involve large bell-shaped movements of the lower lip, and smaller movements of the upper lip that are closely associated with the onset and offset of closure. There is greater lowering of the upper lip at closure onset for /p/ than for /b/. The main difference between /p/ and /b/ lies in the presence of significant lowering of the tongue back during the closure for /b/ which is probably attributable to cavity expansion to sustain voicing.

The production of simple velar stops is illustrated in the next two figures. Figures 13 and 14 show the articulatory trajectories in the production of the sequence /aka/. The major articulatory movement, as expected, involves raising the back of the tongue. The vertical component of this movement is less symmetrical than that seen for the lower lip in bilabial consonants: the upward movement is more rapid than the lowering movement. It seems probable that this asymmetry is due to anticipation of the upcoming /e/ vowel in the carrier phrase. The low point for the following /a/ vowel is not as low as for the preceding vowel. The asymmetry between closing and opening movements is even more apparent in the trajectory followed by the tongue front. The tongue front is raised quite rapidly in synchrony with the tongue back as the closure is formed, but the front begins to lower earlier than the tongue back and lowers more slowly. Tongue raising is assisted by a small vertical movement of the jaw, which lifts the lower lip by an equal amount. The jaw movement trajectory is more closely correlated with the movement pattern of the tongue front than that of the tongue back. The upper lip remains essentially still.

An important rotational component to the tongue back movement can be clearly seen in Figure 14. The tongue retracts and raises to form the closure, with the releasing movement being forward, beginning during the hold phase of the closure. Similar forward releasing movements in velar stops have been observed in other languages, including English (Perkell 1969, Kent & Moll 1972, Coker 1976, Munhall, Ostry & Flanagan 1991) and German (Mooshammer 1992). Two proposed explanations for this forward movement are a) functional differences in the muscles used for raising the back of the tongue versus lowering it (styloglossus can pull the tongue upward but also pulls it backwards, genioglossus pulls the tongue down and forward), and b) aerodynamic consequences of the air pressure behind the closure (suggested by Kent & Moll 1972). (The latter account suggests that the forward movement should be absent, or at least reduced, in velar nasals. Mooshammer found very little movement with /ŋ/ in German.)

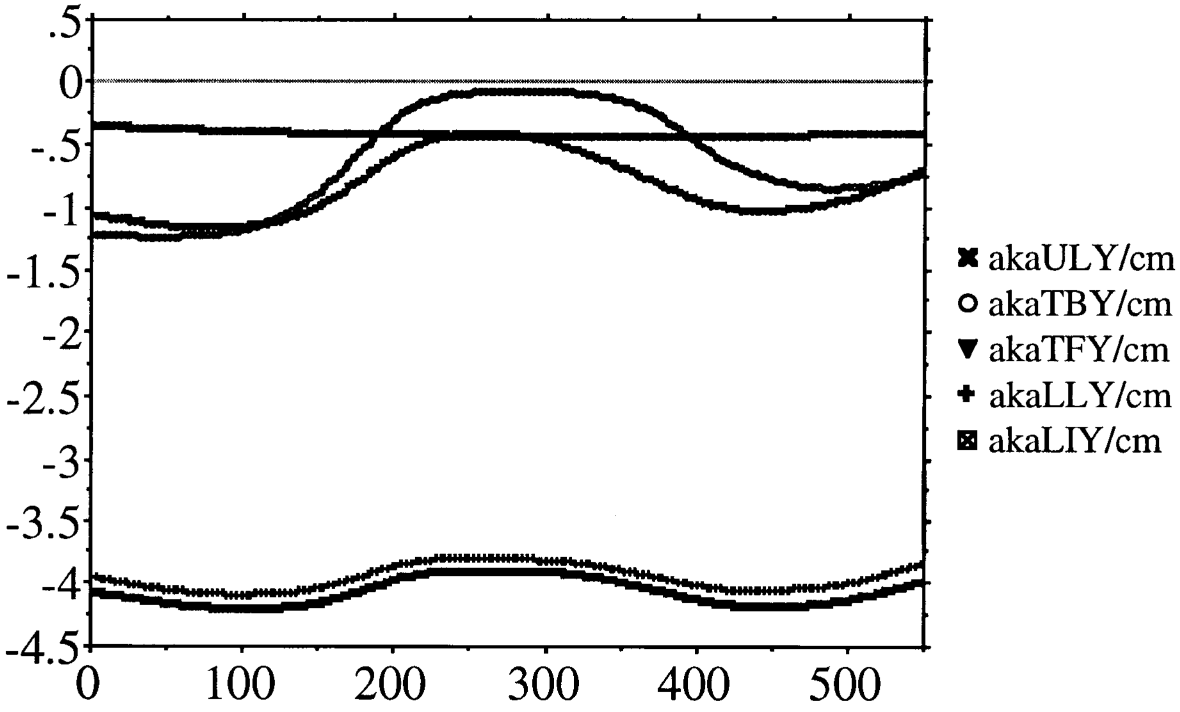


Figure 14. Mean vertical displacement over time of the five receiver coils during the production of ten repetitions of /aka/. Consonant release at 350 ms.

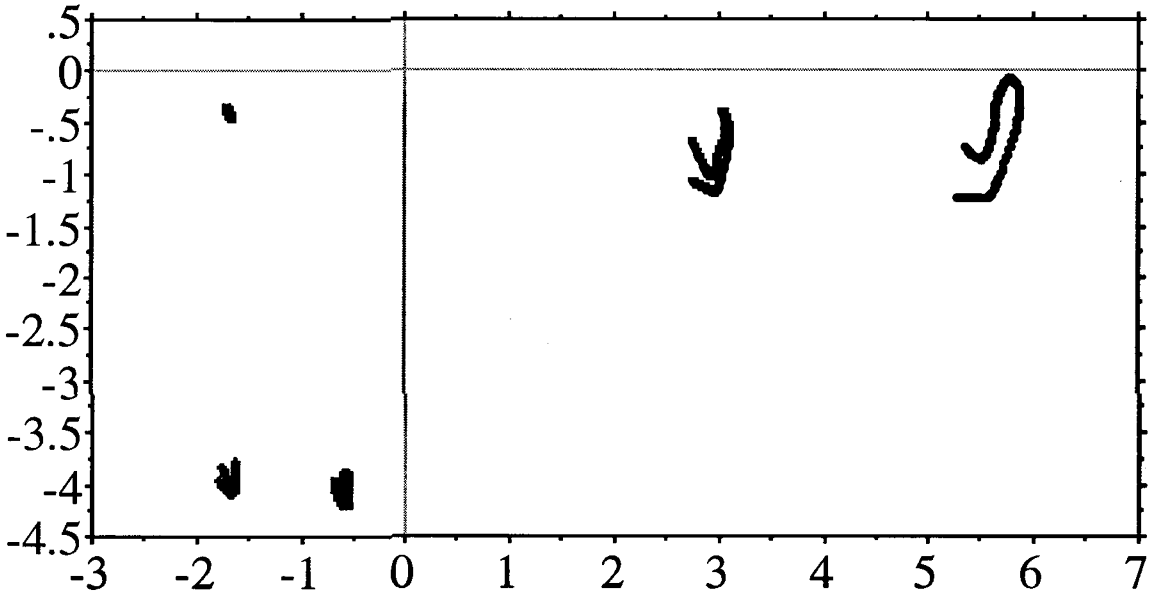


Figure 15. Mean displacements in two-dimensional space of the five articulatory points tracked during the production of /aka/.

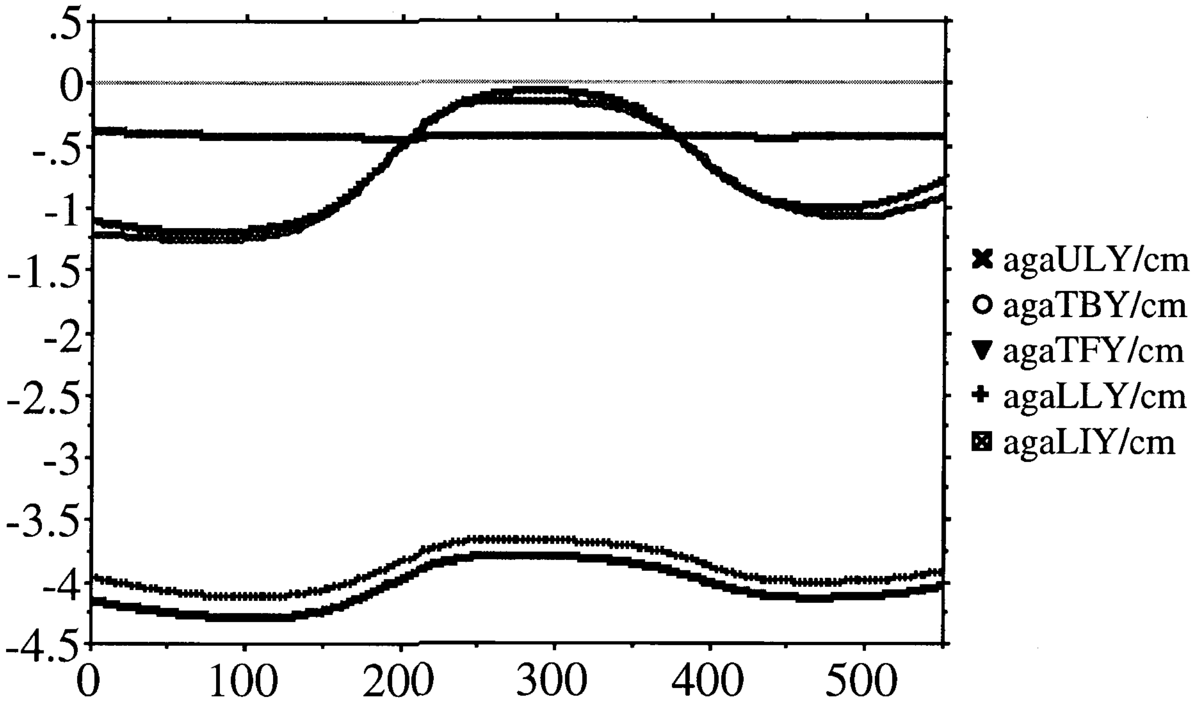


Figure 16. Mean vertical displacement over time of the five receiver coils during the production of ten repetitions of /aga/.

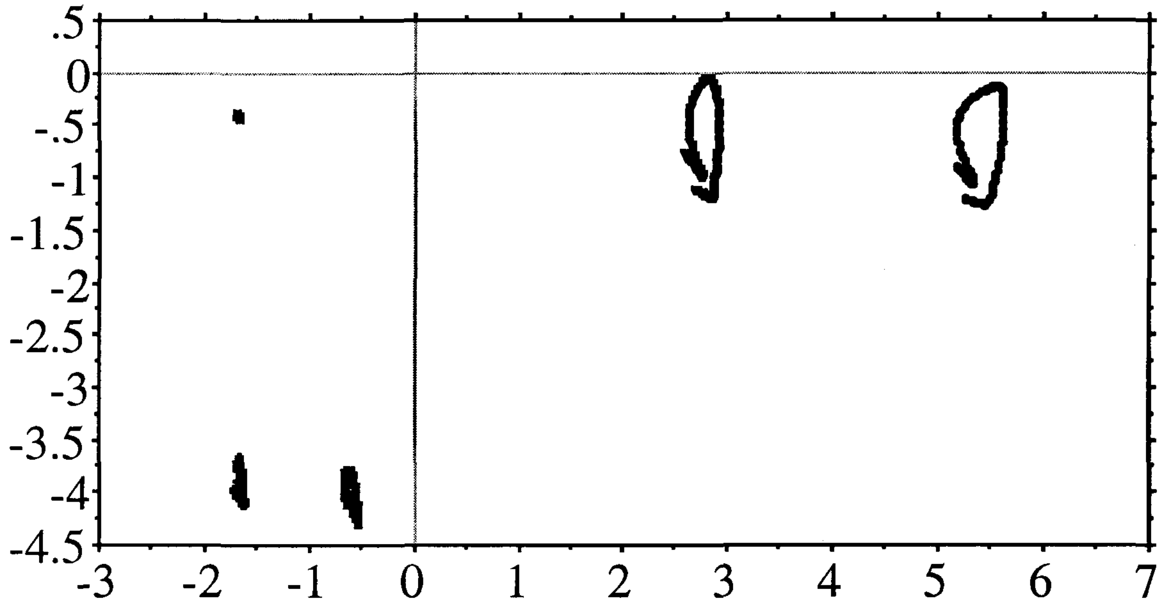


Figure 17. Mean displacements in two-dimensional space of the five articulatory points tracked during the production of /aga/.

The movement patterns in the sequence /aga/ are shown in Figures 16 and 17. In /g/ the tongue front and back move in very close synchrony. The vertical components of both gestures are asymmetrical in that the release is slower than the formation of the closure. The tongue is not as retracted as in /k/ and it moves markedly further forward during the closure and at release than in production of /k/ (compare Figures 15 and 17), so much so that a backward movement follows to produce the vowel /a/. As with the labials, a likely explanation of this difference associated with voicing is that there is some active expansion of the supraglottal cavity to assist the maintenance of voicing. Since this is a velar consonant, lowering of the tongue back would conflict with its articulation, but the location of the closure can be shifted forward. The higher tongue front position for /g/ than for /k/ may be an aspect of this maneuver. The closure for /g/ is also held for a shorter duration than for /k/. No independent lip activity occurs, but a small jaw raising movement raises the lower lip.

In summary, velars involve a large vertical movement of the tongue back which is faster in closing than in releasing and a forward releasing movement. A small jaw raising assists the tongue raising, but no independent movement of the lips occurs. There is much greater vertical tongue front movement in /g/ than in /k/, a shorter closure for /g/ than for /k/ and a more forward position and greater forward movement during closure and release for /g/ than for /k/. These differences may all be associated with maintenance of voicing, particularly through cavity expansion.

The articulatory patterns during the production of /akpa/ are illustrated in Figures 18 and 19. In the vertical direction the tongue back and tongue front move in close synchrony (as in /g/ rather than as in /k/). The asymmetry of closing and release gesture speeds is as in simple velar stops, and is assumed to be related to a higher tongue position for following than for preceding /a/ due to anticipation of the upcoming /e/ vowel.

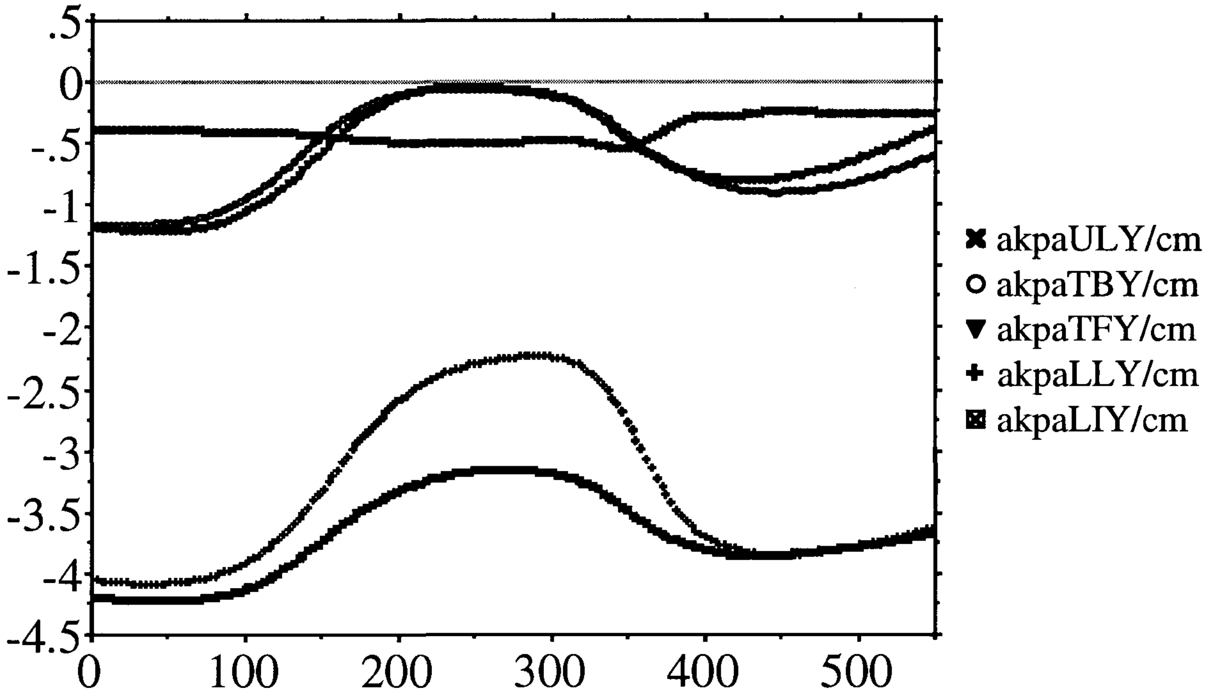


Figure 18. Mean vertical displacement over time of the five receiver coils during the production of ten repetitions of /akpa/.

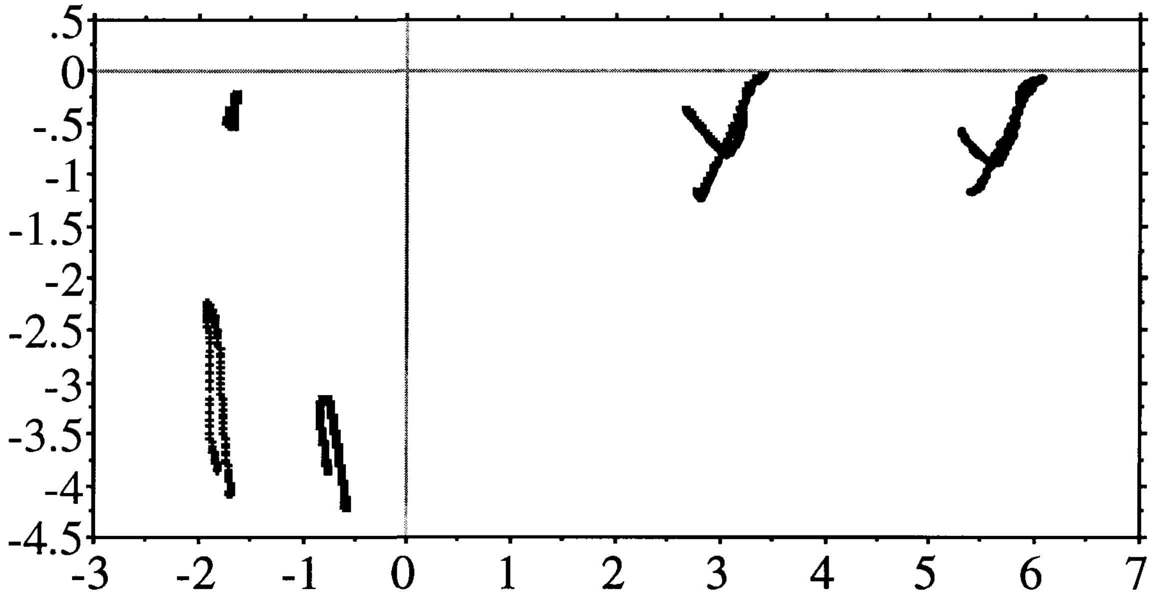


Figure 19. Mean displacements in two-dimensional space of the five articulatory points tracked during the production of /akpa/.

The lower lip movement in /akpa/ is similar in height to that seen in /apaa/ but has a marked asymmetry between closing and releasing phases that is not present in /apaa/. There is a much larger jaw movement in /akpa/ than is seen either /apaa/ or /aka/. The vertical component of this movement correlates more closely with the lower lip movement than with the tongue movement. It is possible to view this pattern as the summation of a jaw movement that assists the lower lip raising (the larger component) with a jaw movement that assists the tongue raising (the smaller component — compare Figures 8 and 14). Nonetheless it is clear that the labial closing movement differs in /akpa/ from that in /apaa/ in the less active role of the upper lip and the slower approach of the lower lip to the closure position.

A very salient difference between the tongue movements in the plain voiceless velar and the doubly-articulated stop is apparent in the spatial configuration shown in Figure 19. The tongue movement in /akpa/ moves more sharply backwards as it moves up to closure and further retraction occurs during the closure hold. The eventual releasing movement is forward but the trajectory of the lowering following release lies approximately in the same spatial region as the raising movement, rather than lying well in front of it, as is the case with /aka/. A further difference between /akpa/ and /apaa/ is the small forward displacement of the lips in /akpa/ that is not present in /apaa/.

The movement trajectories during the production of /agba/ are illustrated in Figures 20 and 21. The labial movements during /agba/ are very similar to those seen in /akpa/, showing the same slower vertical movement to closure with very little upper lip involvement in the closure formation, and a small forward movement of the lips. The jaw movement is similar and can again be regarded as a summation of movements assisting lower lip raising and tongue raising. As for the tongue, Figure 20 shows that the tongue back does not reach the same height as the tongue front and, moreover, it lowers during the closure. Figure 21 shows that there is considerable backward movement of the tongue as it raises toward closure, in contrast to the largely vertical movement in /aga/. This is followed by a small retraction movement during the closure and a substantial forward movement as the release is made. The postconsonantal /a/ has a tongue position well behind that for the preconsonantal vowel so that the downward trajectory crosses the upward path (a much smaller effect of the same kind occurs in /akpa/).

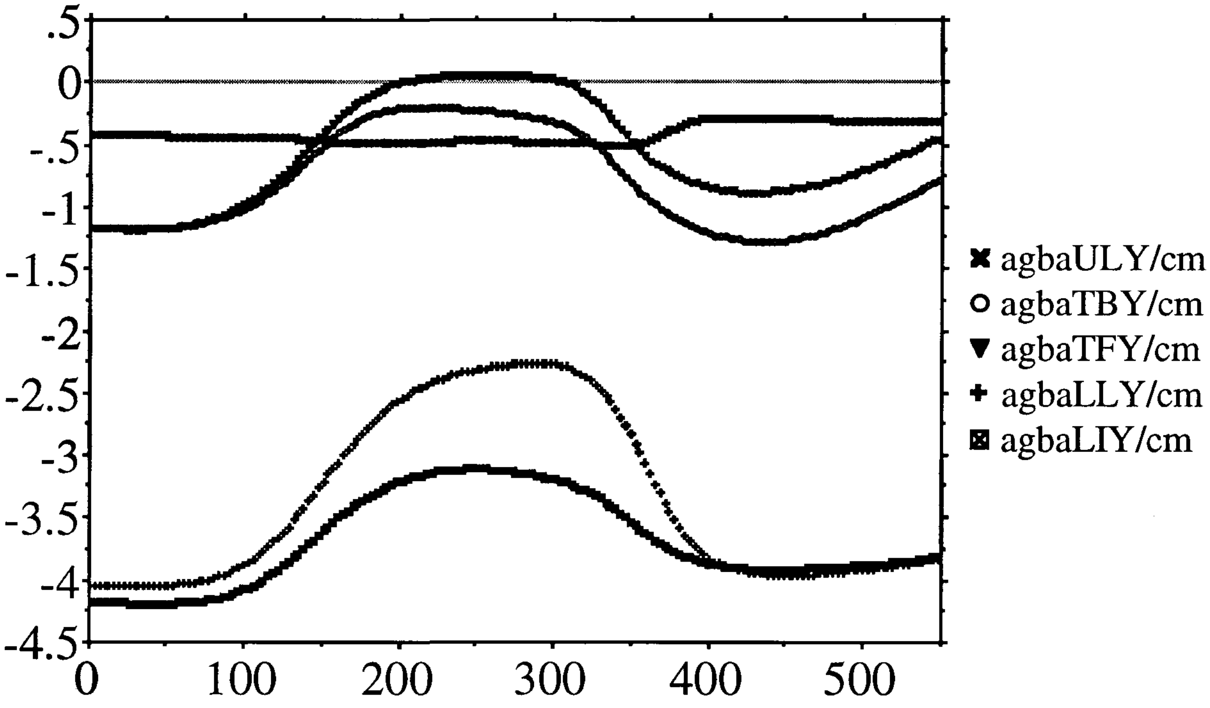


Figure 20. Mean vertical displacement over time of the five receiver coils during the production of ten repetitions of /agba/.

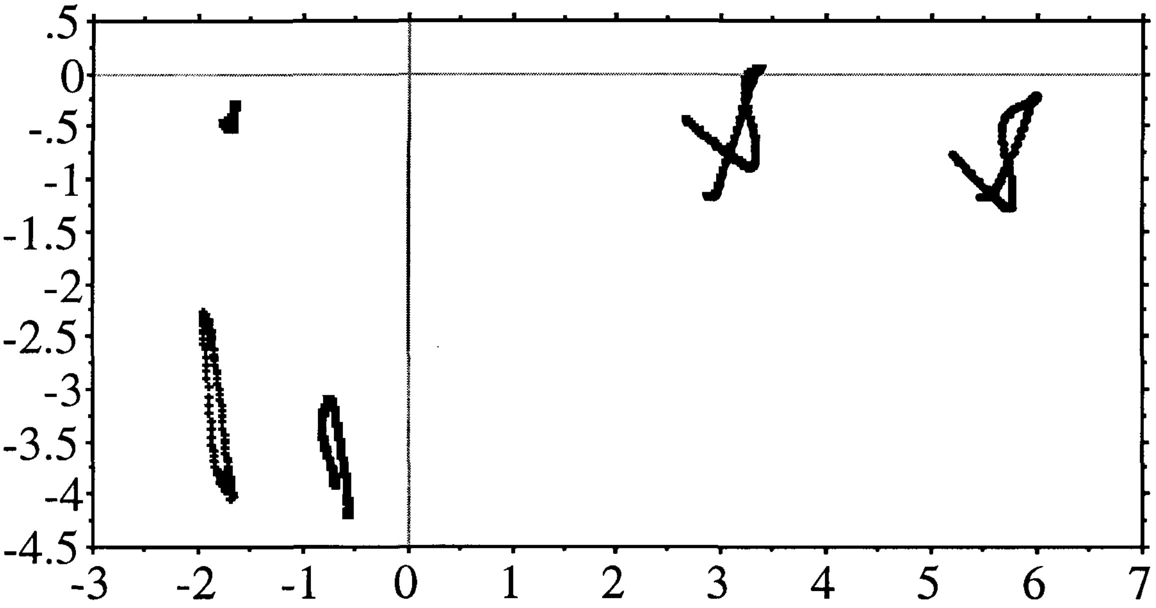


Figure 21. Mean displacements in two-dimensional space of the five articulatory points tracked during the production of /agba/

Thus the lip and jaw movements are almost identical in /akpa/ and /agba/ and both labial-velar stops have a backward movement of the tongue toward closure but a forward release. They differ in that both the tongue back and the tongue front are high during /akpa/, but the back is markedly lower that the front in /agba/ and lowers further during the closure. The tongue back movement lasts for a shorter time in /g͡b/ than in /k͡p/. There is also greater forward movement of the tongue at release in /g͡b/ than in /k͡p/. This greater forward movement can again be interpreted as assisting with voicing maintenance. As the closure is made further back for the labial-velars than for plain velars and consequently the supraglottal cavity is even smaller, the lower tongue back position and the lowering during closure observed in /agba/ may be additional strategies to assist in maintaining transglottal flow. The lowering movement of the tongue back in /abaa/ is interpreted as having a similar function, but in that case the tongue is low throughout, whereas the tongue back must be raised to make the velar closure in /agba/. One possibility is that the tongue back lowering in /agba/ occurs as the velum itself is being lowered, permitting some small volume of air to pass out through the nasal escape channel.

Some of the principal differences between labial-velar and simple bilabial and velar stops apply to both voiced and voiceless stops. Most significant is the retraction of the tongue in both /k͡p/ and /g͡b/ compared with /k/ and /g/. The tongue moves backward as it is raising and retracts a little further during the hold of the closure. The significance of this backward displacement of the tongue will be discussed following an analysis of coarticulation between vowels and velar consonant gestures. Additionally the closing gestures of the lips are somewhat different in simple labials and in labial-velars. We suggest that since in labial-velars the velar closure is formed first, the labial closing movement can be less vigorous as it has no airflow to resist.

A specific difference between the tongue movements in voiceless /k͡p/ and /k/ concerns the tongue front, which is relatively low in /k/ but high in /k͡p/. The tongue front can be regarded as only passively involved in both productions, and its greater height in /k͡p/ may therefore simply be the result of riding on the larger jaw excursion that occurs in /k͡p/ because of its accompanying labial gesture. This explanation entails having a different account of the high tongue front position in the simple voiced velar /g/. No fully persuasive hypothesis has been developed but as noted above it may be connected with the greater forward displacement of the tongue body as /g/ is released, resulting in more bunching up of the tongue front. Note that the tongue front moves forward at release in /g/ almost as much as the tongue back, but in /k/ the tongue front shows very little forward movement at release. The main difference between the tongue movement in the voiced stops /g/ and /g͡b/ is the lowered tongue back position in /g͡b/, which is interpreted as related to different conditions required to maintain voicing.

The small forward lip movement in labial-velars also can be related to the jaw raising associated with tongue raising. Since jaw movement usually has a rotational component, the lips are moved forward in a way that does not occur when lips alone are moved.

Hypothesis 3 referred to the possibility of rarefaction of the air enclosed between the labial and velar closures occurring as a result of the relative timing of the velar and labial gestures. It is not known if any rarefaction occurred in these particular productions, as no aerodynamic data was collected. However, since the tongue back begins to lower well before the labial release in both /akpa/ and /agba/, it is probable that some rarefaction is produced as the contact area for the tongue reduces. The backward movement of the tongue during the closure in labial-velars might also contribute to expanding the size of the cavity between the labial and velar closures. Recall Ladefoged's inference that in labial-velars "the back of the tongue moves slightly further back, creating a slight suction effect as in a click". This issue will also be discussed more fully following a review of the coarticulation between the tongue back and surrounding vowels.

### Coarticulation

Hypothesis 4 proposes that coarticulation between the tongue back movement in labial-velars and surrounding vowels will not be more constrained than between the tongue back movement in simple velars and the vowels surrounding them. This is in contrast to clicks, another class of complex consonants involving a velar articulation together with a second oral closure. Sands (1991) reports strong perceptual and acoustic evidence that the back closure of Xhosa clicks is unaffected by the tongue position of adjacent vowels. Sands suggests that this is because "both the front and the back of the tongue have to be in particular positions to produce [clicks]".

Coarticulation in simple velars and labial-velars is compared in Figures 22-25, which show two-dimensional tongue back trajectories on a larger scale than the earlier figures. A symmetrical front vowel environment, /eCe/, is compared with the central vowel environment /aCa/, and for 3 of the 4 consonants concerned also with an asymmetrical environment that involves movement from the front vowel /e/ to a back vowel (either /o/ or /u/). The three trajectories for /k/ are shown in Figure 22. The looping movement in /aka/ is already familiar, and /eke/ also shows a forward movement at the release. The closure in /eke/ is formed about 7 mm further forward than in /aka/. Closure in /eku/ is formed in between these two locations, and a large backward movement occurs during the hold so that the release occurs at about the same location as in /aka/. The releasing movement in this case is down and backward as the tongue continues to move back for the back vowel position for /u/.

Figure 23 shows the tongue back movement trajectories for /kp/. As with /k/, the closure in the front vowel environment is about 7 mm in front of the location for the central vowel environment in /akpa/, although both locations are shifted about 3 mm backwards compared with the simple velars. A similar retraction and forward releasing gesture occurs in /ekpe/ and /akpa/.

Of particular interest is the trajectory seen in /ekpo/. The general backward tongue movement required to move from the front to the back vowel has superimposed on it a sharp backward then forward excursion that is quite unlike the continuous backward movement seen in /eku/. It is as if some force exerts a brief but strong pull backwards, from which the tongue rebounds when it ends. Seeing this superimposed movement makes it easier to envisage the trajectories in the symmetrical vowel environments as containing a similar superimposed backward tug from which there is a rebound.

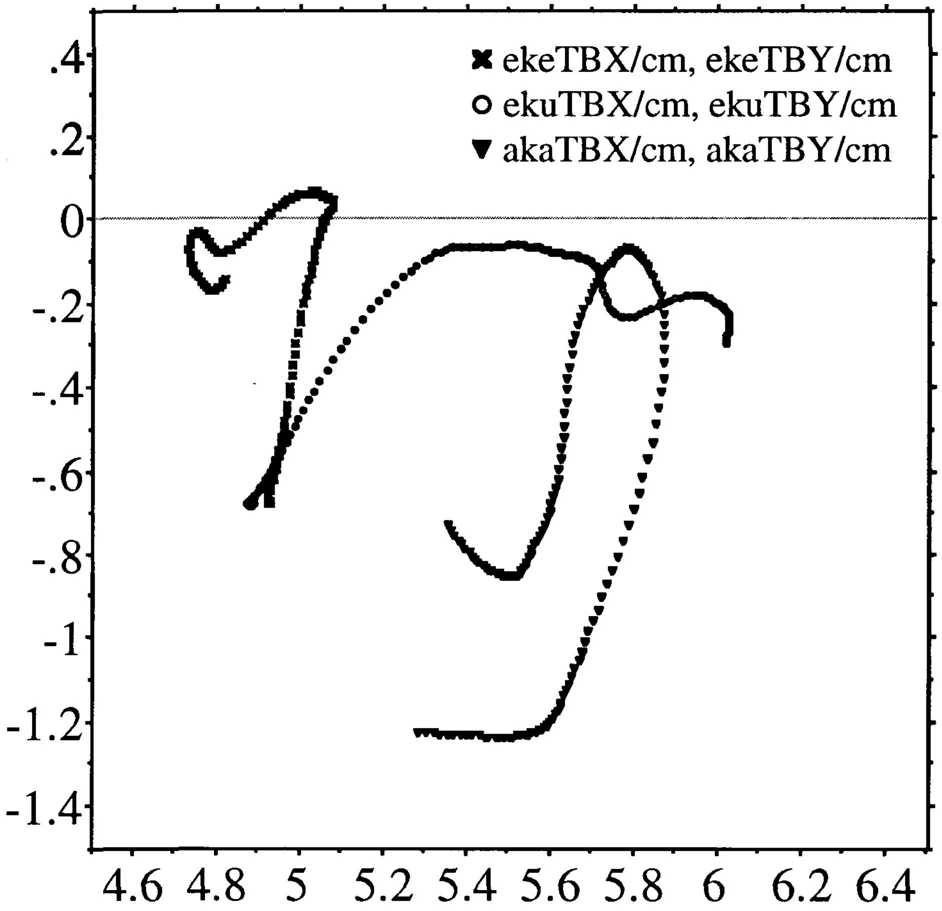


Figure 22. Mean tongue back movement in two-dimensional space for ten repetitions of /eke, eku, aka/. Measurements are in cm, with the axes on the same scale. The time interval covered is 350 ms. The front is to the left.

The tongue back movement trajectories for /ege, egu, aga/ are shown in Figure 24. There is a smaller difference in the position of the tongue back at the gestural peak of /ege/ and /aga/ than in the voiceless counterparts; /ege/ is further back than /eke/, and /aga/ is further forward than /aka/. The difference is about 4 mm. There is a large backward movement during the /egu/ sequence, with a very small forward movement at the release before backward movement resumes for the /u/ vowel. The trajectories for /egbe/ and /agba/ are shown in Figure 25. The gestural peak for /egbe/ is about 4.5 mm in front of that for /agba/. The backward translation of the movement in labial-velars results in the peak for /egbe/ lying at about the same location as that in /aga/.

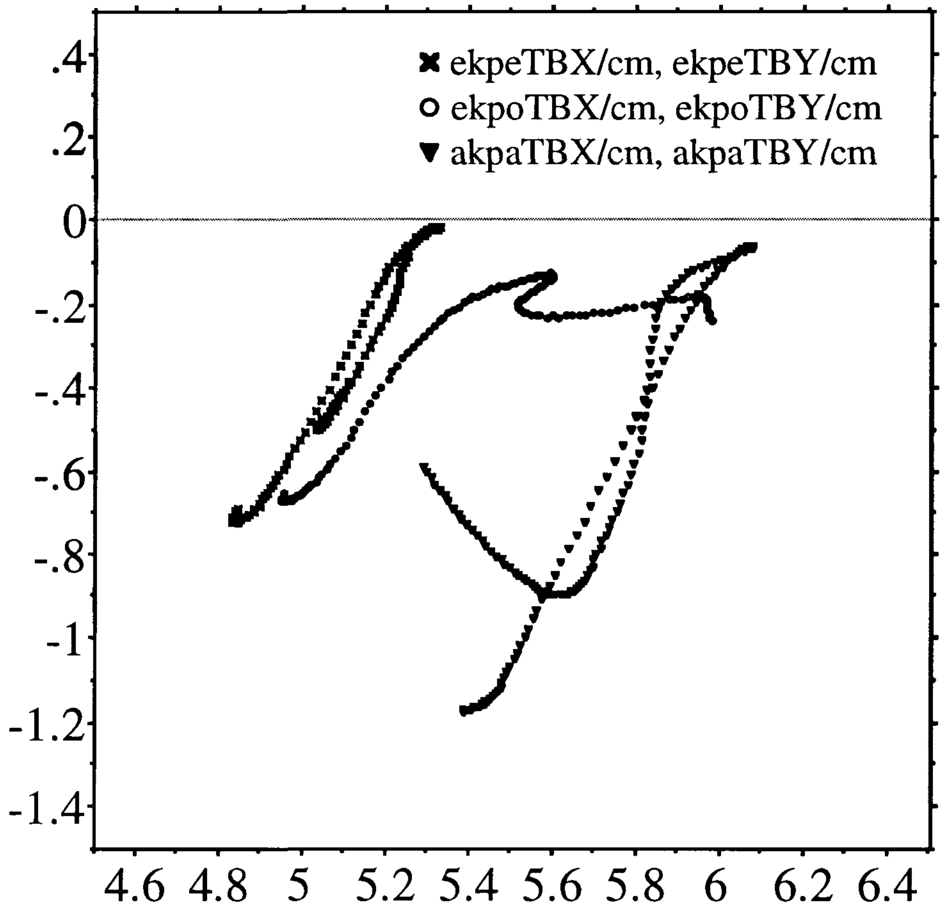


Figure 23. Tongue back movement in two-dimensional space for /ekpe, ekpo, akpa/.

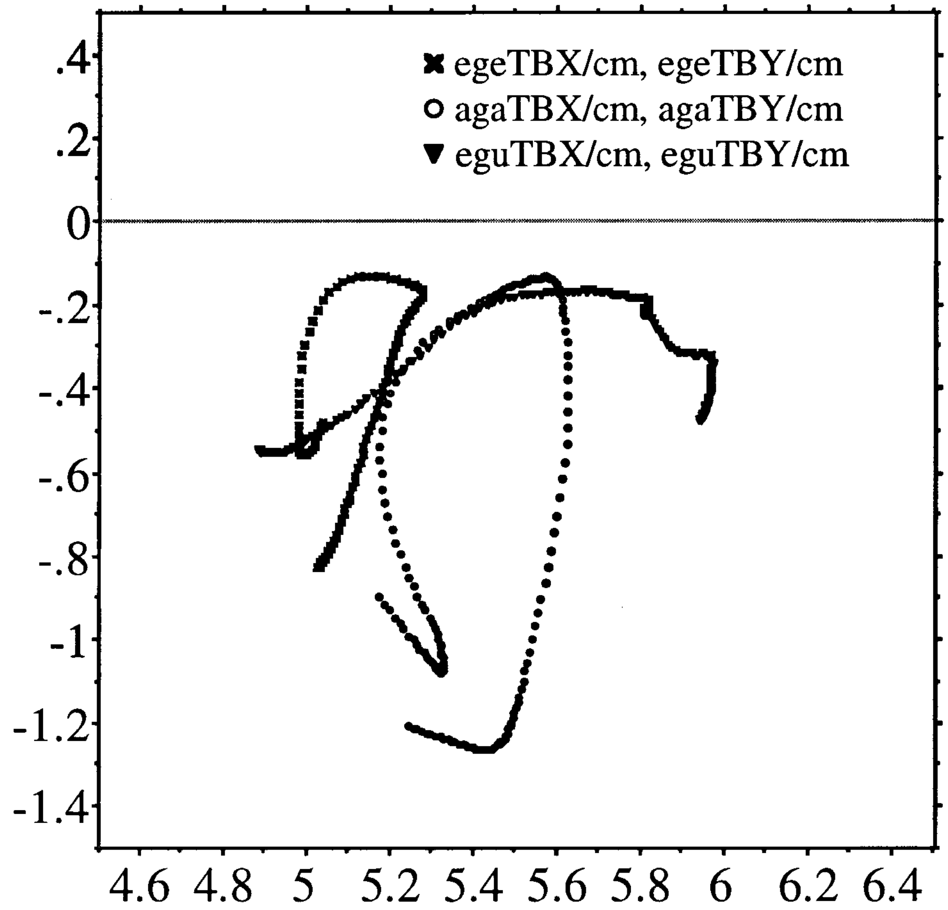


Figure 24. Tongue back movement in two-dimensional space for /ege, egu, aga/.

This analysis shows that the degree of coarticulatory adjustment of the tongue back to surrounding vowels is equal in simple velar stops and labial-velars. Labial-velars are in this respect unlike clicks. But, because of the backward movement in labial-velars, the trajectories of the tongue back in velars and labial-velars are not similar in matched vowel environments.

## The backward movement in labial-velars

Broad similarities between gestures in labial-velar stops and gestures in simple velar and labial stops have generally been found. Several of the differences in the movement tracks have very natural explanations, in terms of such factors as the interaction of jaw and lip movements, or efforts to sustain voicing. However, the backward movement of the tongue in labial-velars has no such explanation. In order to get a view of the two-dimensional movements over time tangential velocity was computed (Tangential velocity is the square root of the the sum of the squared velocities in the x and y planes. A seven-point triangular smoothing window was applied.) Figures 26-30 compare several matched pairs of words containing velars and labial-velars. The traces are lined up at the acoustic consonant release. Since this is the labial release in labial-velars, comparable time points in the movements for simple velars occur later. Voiceless labial-velars reach a stationary position at their turn-around point (0 tangential velocity), whereas velars do not. The effect is a little less dramatic for the comparison of voiced velar and labial-velars, but a distinction between the patterns is nonetheless apparent.

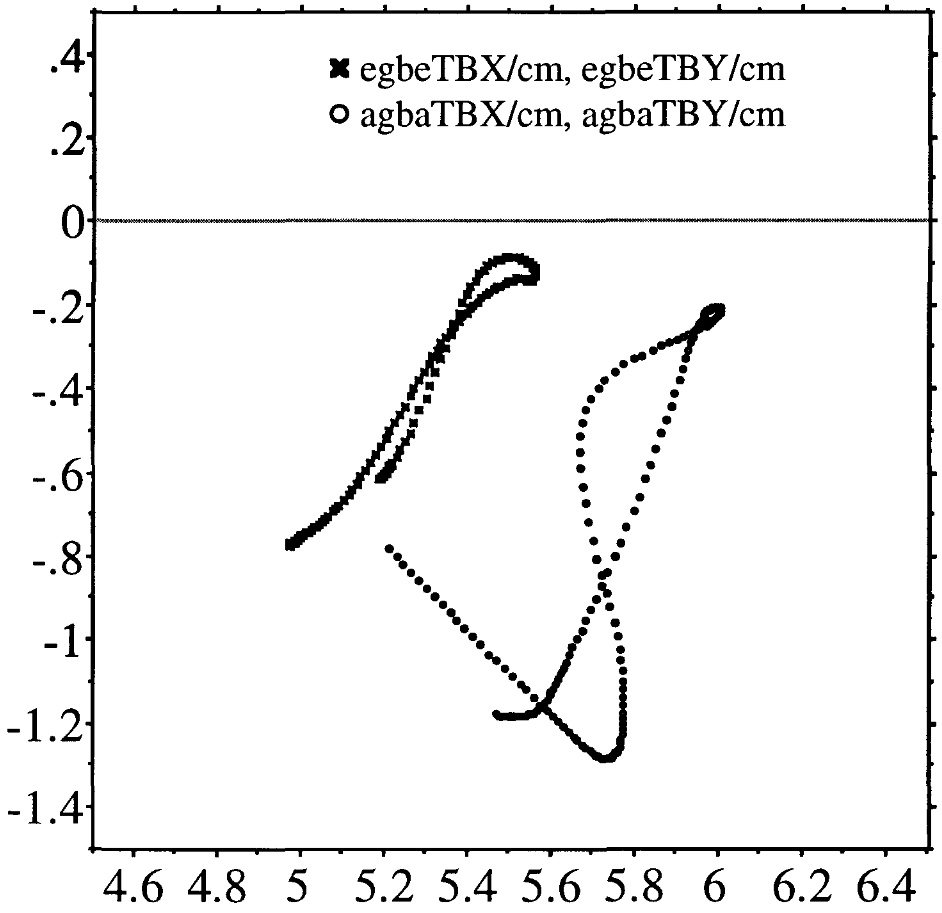


Figure 25. Tongue back movement in two-dimensional space for /egbe, agba/.

The minimum in the tangential velocity of the labial-velars is the time-point at which the backward movement ends and the forward movement begins. Figures 26-30 show that this moment occurs 100 ms or more before the labial release (which occurs generally near a velocity maximum in the tongue movement). The sharp backward movement thus occurs early in the consonant.

Several hypotheses concerning the reason for the retraction in labial-velars were considered. These will be briefly outlined and discussed in turn.

Hypothesis A: Retraction results from movements intended to create *suction* in the front cavity. Some labial-velars in Edo and other languages examined by Ladefoged (1968) showed *rarefaction* of the air in the cavity between the labial and velar closures. If retraction were part of a strategy to expand the size of the front cavity, so as to create a distinctive release, then the timing of the movement would be expected to fall much later in the consonant than is seen in Figures 26-30. The Ewe labial-velars do have backward movement of the tongue, but the timing of this movement makes it unlikely that this movement is designed to create a click-like release.

Hypothesis B: Retraction results from *increased air pressure* in the front cavity pushing the tongue backwards. If front cavity *pressure* produces the backward movement, it is most unlikely that simply closing the lips could generate a sufficient pressure; some additional movement to create that pressure should be apparent, for example greater tongue front raising or an early, large jaw raising movement. No evidence of such gestures that could contract the size of the front cavity is apparent.

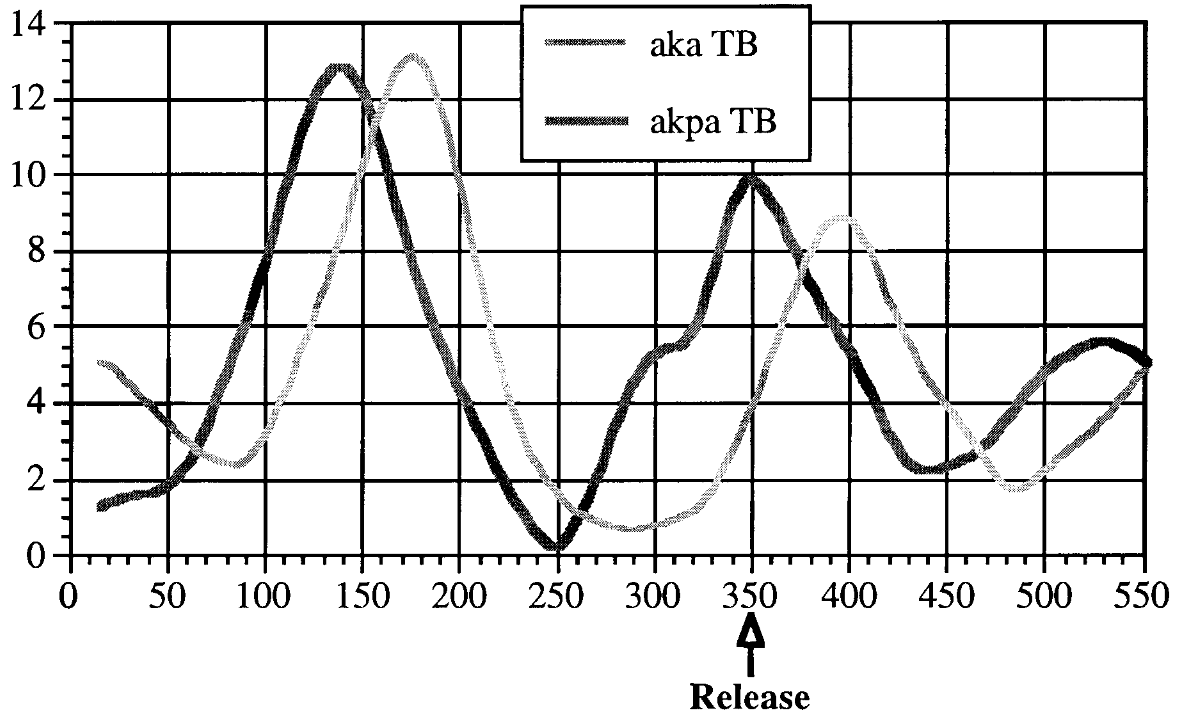


Figure 26. Tangential velocity of the tongue back in /aka/ and /akpa/. The vertical scale is in cm/s.

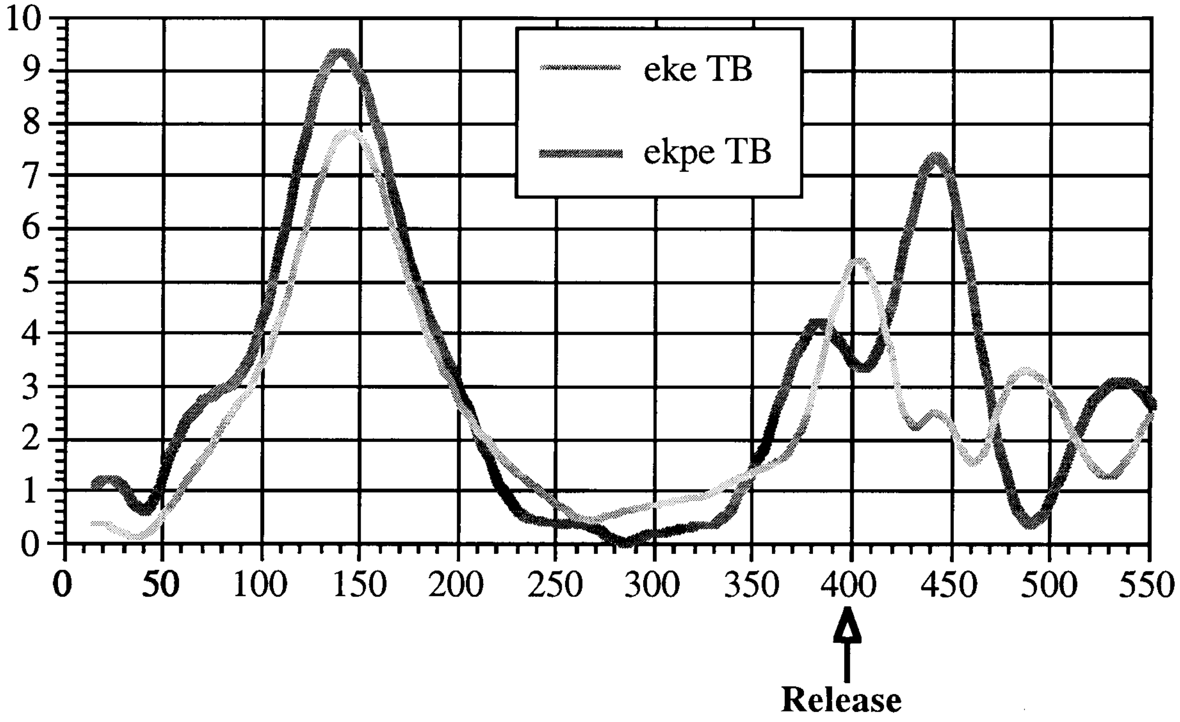


Figure 27. Tangential velocity of the tongue back in /eke/ and /ekpe/.

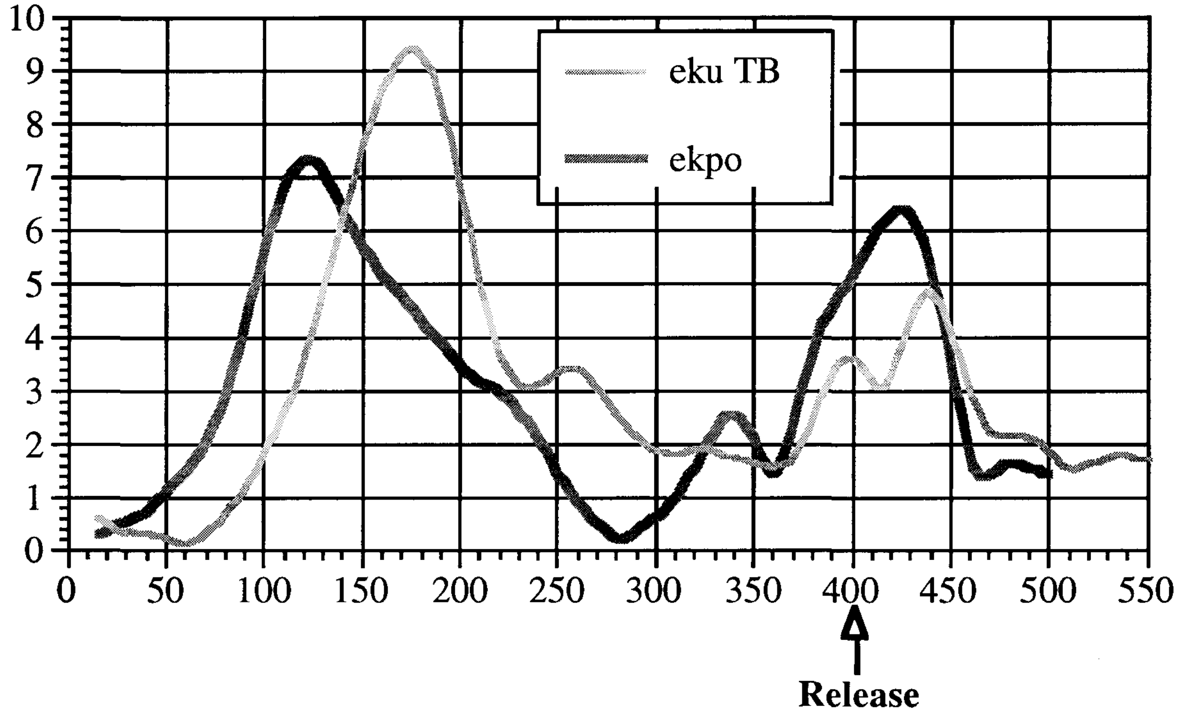


Figure 28. Tangential velocity of the tongue back in /eku/ and /ekpo/.

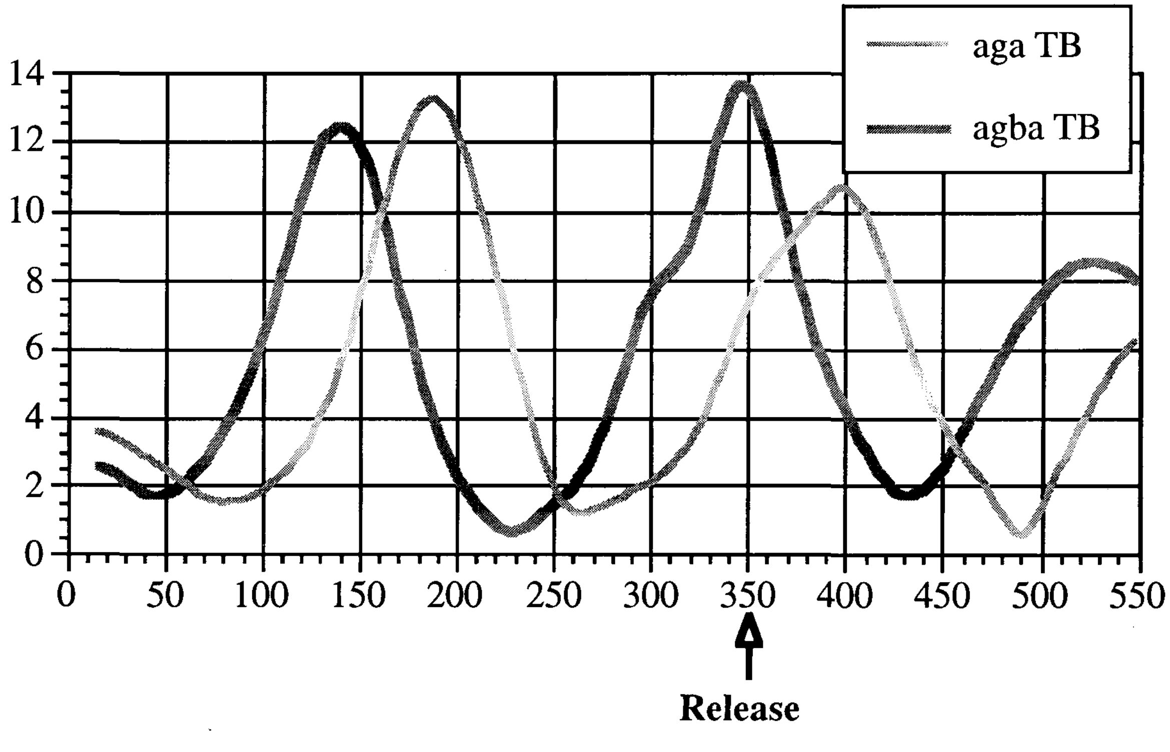


Figure 29. Tangential velocity of the tongue back in /aga/ and /agba/.

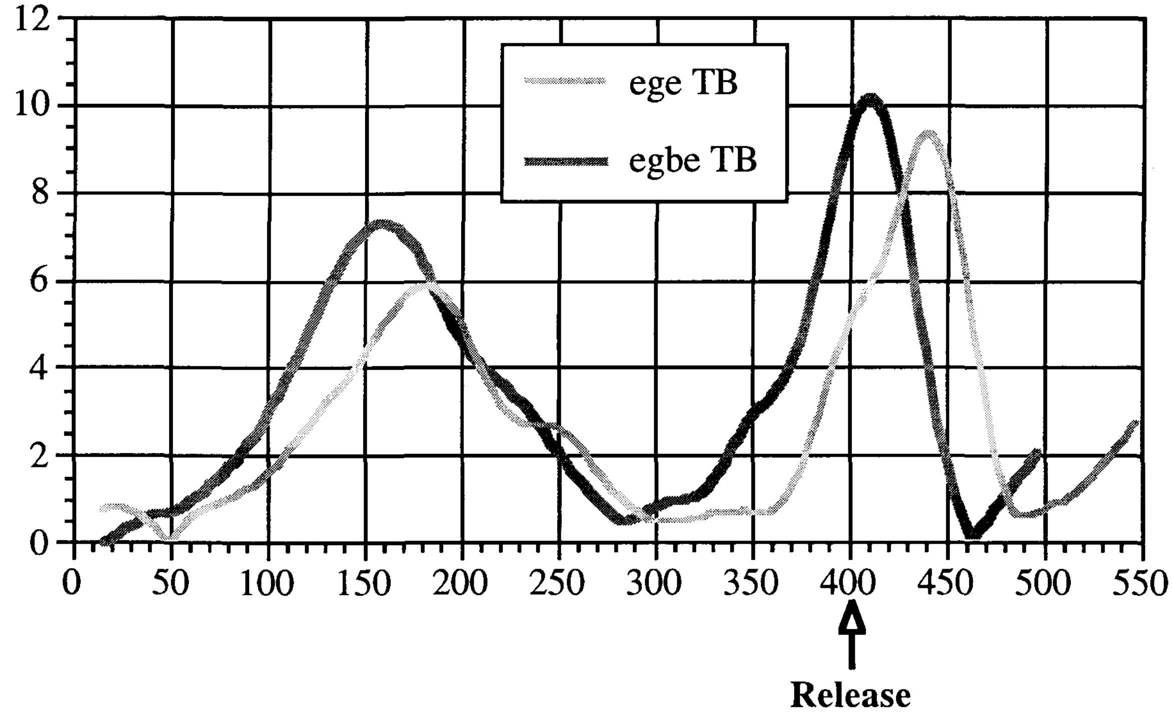


Figure 30. Tangential velocity of the tongue back in /ege/ and /egbe/.

Hypothesis C: Retraction results from greater *fortition* of the velar contact, resulting in more compression of the tongue, leading to rearward displacement of the point where the receiver coil is attached. If greater pressure of the tongue against the velum causes flattening of the tongue and backward displacement of the point where the transducer is attached, then a higher peak displacement in labial-velars than in velars might be expected due to both tongue compression and compliance of the velum. This does not occur, as can be seen by comparing the height of the tongue back in Figures such as 22 and 23.

Hypothesis D: Retraction is part of a set of cordinated movements to create *pressure* in the pharyngeal cavity. Again, some of the labial-velars examined by Ladefoged (1968) showed high pressure in the pharyngeal cavity as for an ejective stop. Rearward movement of the tongue (especially the root) could be part of a set of coordinated movements operating to produce increased pressure by reducing pharynx volume. Again, if the intent were to create a distinctive type of release, the timing of the retraction would be expected to be later in the consonant.

Hypothesis E: Retraction results from larynx lowering, creating *suction* in the pharyngeal cavity, pulling the tongue dorsum back. A volume change in the pharyngeal cavity behind the velar closure could be created by lowering the larynx with the glottis fully or partially closed. Once again, the timing seems problematical.

None of these hypotheses at present seems satisfactory. Availability of aerodynamic data from the same speaker would provide significant assistance in deciding between them, or in suggesting additional possibilities. We note that simple labials and velars and labial-velars differ in voicing features measured in the acoustic signal, suggesting perhaps that some difference in laryngeal activity occurs. Figure 31 shows mean measurements of VOT for the voiceless stops. Mean voice onset time is even shorter for labial-velars than for labials, and on occasion /kp/ is prevoiced, as in Figure 4. Pharyngeal cavity expansion or some other action of the larynx may therefore be involved in producing labial-velars. Although no answer can be given to this riddle at present, the most important fact is that without the articulatory tracking capabilities of EMMA, we would not know that there is a question to be posed.

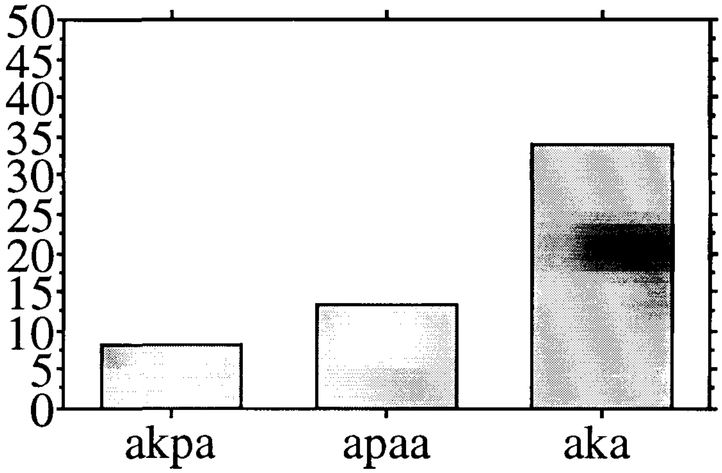


Figure 31. Voice onset time in voiceless labial-velar, bilabial and velar stops.

**Labial Fricatives**

The final issue that will be addressed in this paper concerns the production of the labio-dental fricatives /f/ and /v/ and the bilabial fricatives /ɸ/ and /β/. The idea of an enhanced articulatory distinction in languages in which these contrast has been discussed in several publications by Ladefoged. In a recent book he writes:

> "In English there is no phonemic distinction between [ɸ] and [f] or between [β] and [v]. Consequently, there is no need to maintain a perceptual distinction between these two possibilities. In saying a word such as "fin", the upper lip is seldom lifted out of the way. The friction is often formed between the lower lip and both the upper teeth and the upper lip. But in the West African language Ewe, where there is a contrast between words such as [éɸá] (he polished) and [éfá] (he was cold), the distinctions between bilabial and labiodental fricatives are maximized. In making the labiodental fricatives, the upper lip is actively raised so that friction can occur only between the upper teeth and the lower lip." (Ladefoged 1993: 269).

In order to investigate the spatial pattern of the contrast between bilabial and labio-dental articulations the third set of words in Table 2 was recorded in the carrier phrase /ɖo be ___ ni mama ni se/ "Say ___ for grandma to hear". The original selection of words was designed to examine how the hypothesized lip raising in labio-dentals would interact with the protrusion gesture of rounded vowels. There is, however, no lip raising.

The upper and lower lip and the jaw movement tracks in two dimensional space for bilabial and labio-dental fricatives surrounded by unrounded vowels are given in Figures 32-35. Figures 32 and 33 compare /eve/ "two" and /eβe/ "Ewe people."

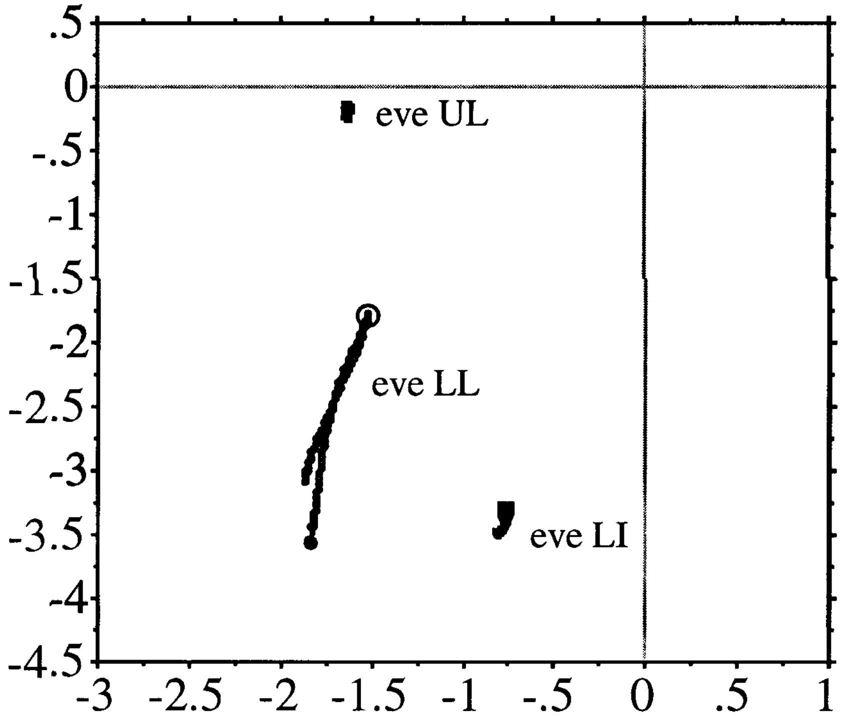


Figure 32. Mean lip and jaw movement trajectories in two-dimensional space in ten repetitions of /eve/. The time interval covered is 550 ms. Front is to the left.

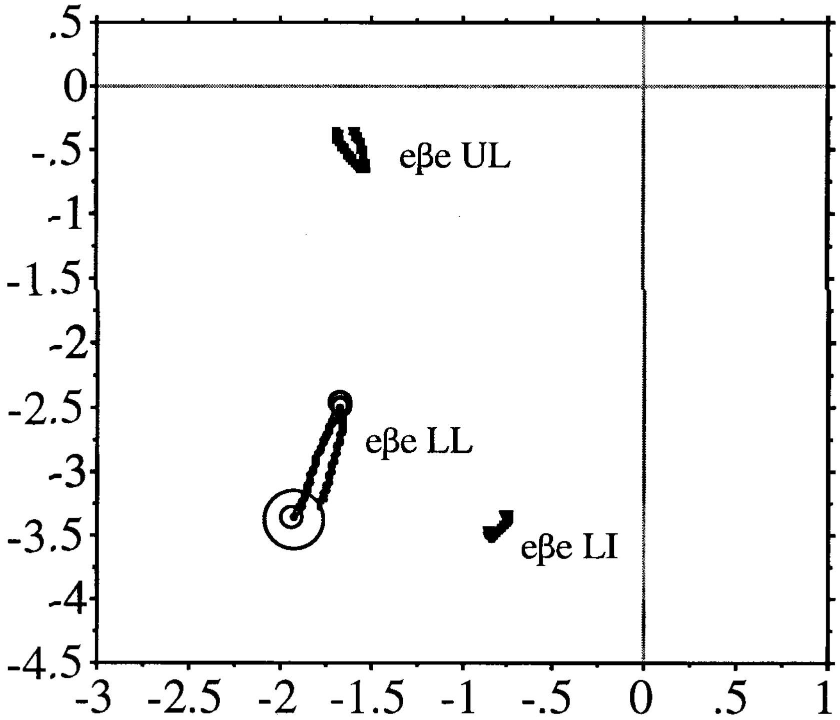


Figure 33. Mean lip and jaw movement trajectories in ten repetitions of /eβe/.

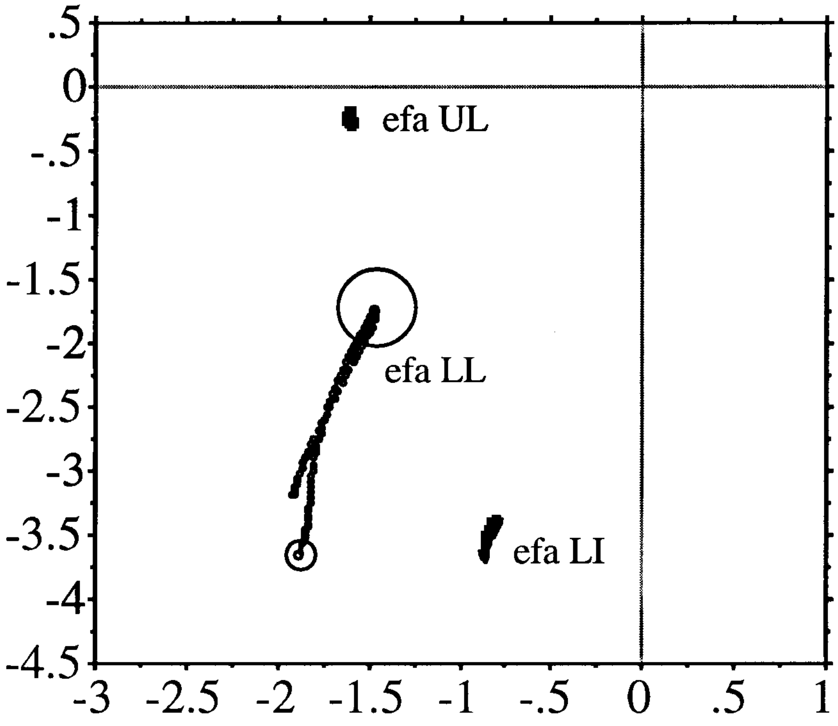


Figure 34. Mean lip and jaw movement trajectories in ten repetitions of /efa/.

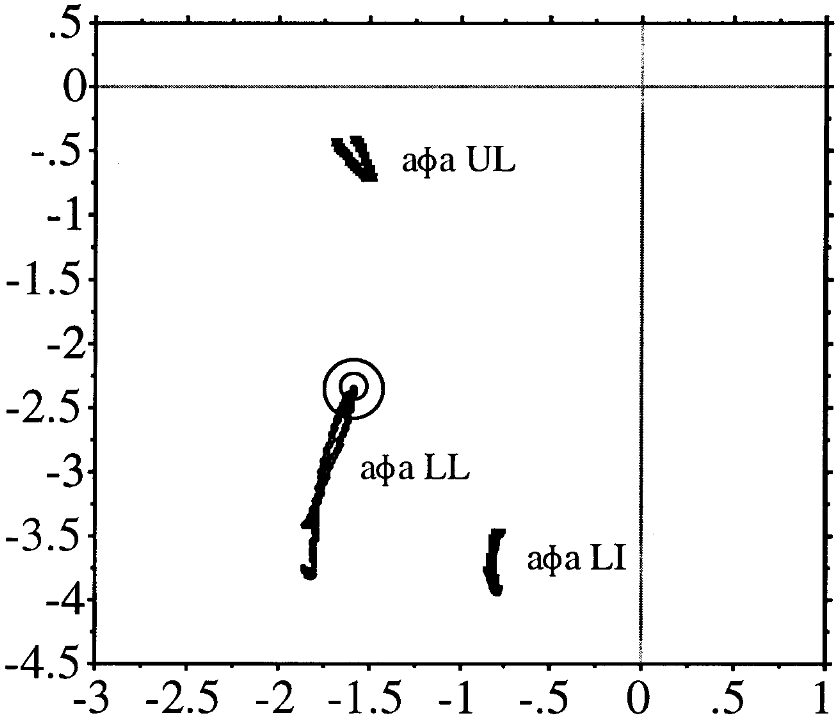


Figure 35. Mean lip and jaw movement trajectories in ten repetitions of /aϕa/.

In the voiced labio-dental, the upper lip makes no significant movement in any direction; all the work of forming the fricative constriction is done by raising and retracting the lower lip. During the consonant the jaw also remains stationary — a small movement occurs in the first vowel. In the bilabial, the lower lip also raises and retracts, but to a smaller degree in both directions than in the labio-dental. A retraction of the jaw assists the lower lip movement, especially in its rearward direction. The upper lip is actively lowered, and also retracted so as to meet the lower lip. The lower lip receiver coil reaches a maximum height in /β/ that is about 1 cm lower than in /v/, while the upper lip moves about 7 mm down. The maximum excursion for /β/ is about 2 mm in front of that for /v/. The releasing movement of both lips is down and rearwards, so that the rotational component of the lip movements in Figure 33 is clockwise.

Figures 34 and 35 show the movements in the voiceless labio-dental and bilabial fricatives in the words /e fa/ "it's cold" and /aɸa/ "shout" respectively. As with the voiced labio-dental, the upper lip remains completely stationary during the production of /f/. All the labial movement is made by the lower lip alone. The bilabial /ɸ/ is produced similarly to its voiced counterpart, with a raising and somewhat rearward movement of the lower lip and a lowering and backward movement of the upper lip. The release of /ɸ/ has a smaller rearward displacement component than /β/. There is a greater vertical movement of the jaw in the voiceless cases than in the voiced cases.

In summary, bilabial fricatives are truly BI-labial, labio-dental fricatives involve only the raising and retraction of the lower lip to meet the fixed upper teeth (or perhaps the lower inside surface of the upper lip). Although it is possible that other speakers use an 'enhancing' movement of the upper lip in labio-dentals, this speaker does not, even in the rather formal 'laboratory speech' style that inevitably results in experiments of this type. The lips are, of course, not in the same position for labio-dentals and bilabials, but this is because they move to produce a bilabial.

## Conclusion

Rich data on articulatory movements is now more easily obtainable than ever before. Despite the formal experimental situation in which data is collected with the EMMA system, it is evident that quite naturalistic speech can be obtained. Dynamic views of such movements enable hypotheses that depend on timing information to be critically examined. Confirmation of hypothesized timing patterns inferred from acoustic records was provided for labial-velar stops. However, the movement patterns — especially viewed in two dimensions — may provide unexpected surprises and interpreting the significance of such data remains challenging. This is demonstrated by the retraction oberved in labial-velars. In this case, collection of aerodynamic records from the same subject would facilitate the interpretation of the movement data. The analysis of labial fricatives suggests that dynamic views of articulations can change interpretations made on the basis of static images.

## References

Ansre, Gilbert. (1966). The grammatical units of Ewe. Ph. D. dissertation, University of London.

Bendor-Samuel, John (ed.). (1989). *The Niger-Congo Languages.* University Press of America, Lanham, MD.

Capo, Hounkpati B. C. (1991). *Comparative Phonology of Gbe.* Foris, Berlin.

Coker, Cecil. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE* 64: 452-460.

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

Kent, R. D. & K. Moll. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research* 15: 453-473.

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

Ladefoged, Peter. (1990). What do we symbolize? Thoughts prompted by bilabial and labiodental fricatives. *Journal of the International Phonetic Association* 20.2:33-36.

Ladefoged, Peter. (1993). *A Course in Phonetics. Third Edition.* HBJ, New York.

Maddieson, Ian. & Peter. Ladefoged. (1989). Multiply-articulated segments and the feature hierarchy. Paper presented at the Annual Meeting of the LSA, Washington, D.C. Printed in *UCLA Working Papers in Phonetics* 72: 116-138.

Mooshammer, Christine. (1992). Artikulatorische Untersuchung mit EMA — die Zungenbewegung bei der Produktion von VCV-Sequenzen mit velarer Konsonanz und langem und kurzem Erstvokal. Hausarbeit zur Erlangerung des Magistergrades an Ludwig-Maximiliens-Universität, München.

Munhall, K., D. Ostry & J. Flanagan. (1991). Coordinate spaces in speech planning. *Journal of Phonetics* 19: 293-307.

Perkell, Joseph S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. MIT Press, Cambridge, MA.

Perkell, J. S., M Cohen, M. Svirsky, M. Matthies, I. Garabieta & M. Jackson. (1992). Electromagnetic Midsagittal Articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America.*

Perkell, J. S., M. A. Svirsky, M. L. Matthies & J. Mazella. (1993). Measuring articulatory movements with an electro-magnetic midsagittal articulometer (EMMA) system. *Journal of the Acoustical Society of America* 93: 2415.

Sands, Bonny. (1991). Evidence for click features: acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 80: 6-37.

Stahlke, Herbert. (1970). *Topics in Ewe Phonology.* Ph. D. dissertation, University of Illinois, Champaign-Urbana.

Stewart, John. (1989). Kwa. In J. Bendor-Samuel (ed.). (1989).

Westermann, Diedrich. (1930). *A Study of the Ewe Language.* Oxford University Press, London. (Translation by A. L. Bickford).

Westermann, Diedrich. (1954). *Wörterbuch der Ewesprache (revised edition).* Akademie Verlag, Berlin.

Westermann, Diedrich and Ida C. Ward. (1933). *Practical Phonetics for Students of African Languages.* Oxford University Press for International African Institute, London.

# Acoustic and Articulatory Correlates of P-center Adjustments in Tokens of Matched Overall Durations.

**Kenneth de Jong**

**Abstract:**
To evaluate articulatory models of p-center location, listeners performed perceptual adjustments on stimuli which were obtained with accompanying articulatory records. Adjustments were submitted to a regression analysis against the timing of various articulatory and acoustic events. The experiments find various articulatory predictors of p-center location, many of which are likely indexes of vowel gestures. However, predictors were not consistent across words, speakers, or experiments. Thus, p-center locations do not correspond to any particular kinematic articulatory event. In the present experiments, in which stimuli were matched in overall duration, voicing onset was the most effective predictor of p-center adjustments. These results are discussed in terms of their relevance to a model of p-centers as indexes of underlying gestural timing.

## 1. Introduction.

In an earlier experiment, de Jong (1992) presented subjects with words extracted from a corpus of microbeam data in a p-center adjustment paradigm. Listeners were asked to adjust alternations of the words to obtain a perceptually regular rhythm. Twelve repetitions of each of two words, *toast* and *totes*, which varied in the amount of stress placed on them, were used. In order to investigate which aspects of the signal the listeners were using, their adjustments were correlated with the anisochrony of various events in the stimuli syllables. Results varied, depending on which word was being adjusted.

The results were somewhat different for the two words. Listener responses for the set of *toast* tokens suggested that the listeners were aligning an articulatory event to occur at a regular rhythm. Anisochronies in the timing of the peak velocity of movement of the tongue dorsum toward the posterior constriction for the [ow] vowel best predicted the listener's adjustments. Anisochronies in the tongue tip minima and in acoustic events better predicted adjustments of the *totes* . The best predictor was the release of the initial consonant. Dorsal predictors performed rather more poorly for *totes*. For both words in this experiment, all of the acoustic events and tongue tip and jaw events which occurred in the first half of the vowel performed quite well, each predicting more than 70 percent of the variance in listener adjustments.

These results were additionally complicated by the fact that different factors explained different portions of the variance in the listeners' adjustments. In a second analysis of the data, average adjustments of the word *toast* were subtracted from the anisochrony of dorsal velocity maxima to get a deviation from predicted response. These deviations then were subjected to four stepwise regression analyses -- against the intervals of time between the events used in the initial analysis of the data. The multiple regression analysis found that differences in the duration of portions of the later half of the tokens accounted for a significant portion of the variance of the listener adjustments from predicted response. That is, tokens with coda consonants and the second half of the vowel having longer durations tended to have p-centers later in the syllable. This result replicated results of Cooper, Whalen and Fowler (1988) and Marcus (1981), which showed that longer durations in the second half of the syllable tend to pull the p-center later into the syllable. This effect seems to be in addition to the listeners' location of articulatory events near the beginning of the syllable.

Two attributes of the stimuli in this earlier experiment complicated the interpretation of the results. First, the various predictors of p-center location studied were strongly collinear. This was apparent in the similar results for many of the predictors which occurred near the beginning of the tokens. The second difficulty lay in the different overall durations of the tokens. The tendency of tokens with codas of longer durations to have later p-centers could have been due to a tendency for listeners to avoid extreme re-adjustment of the tokens. Extremely long and short tokens demand that the listeners shift the tokens a larger amount to achieve event to event isochrony. If they undershoot this amount, factors most closely correlated with the overall duration of the tokens, such as coda consonant duration, will be correlated with the amount of undershoot.

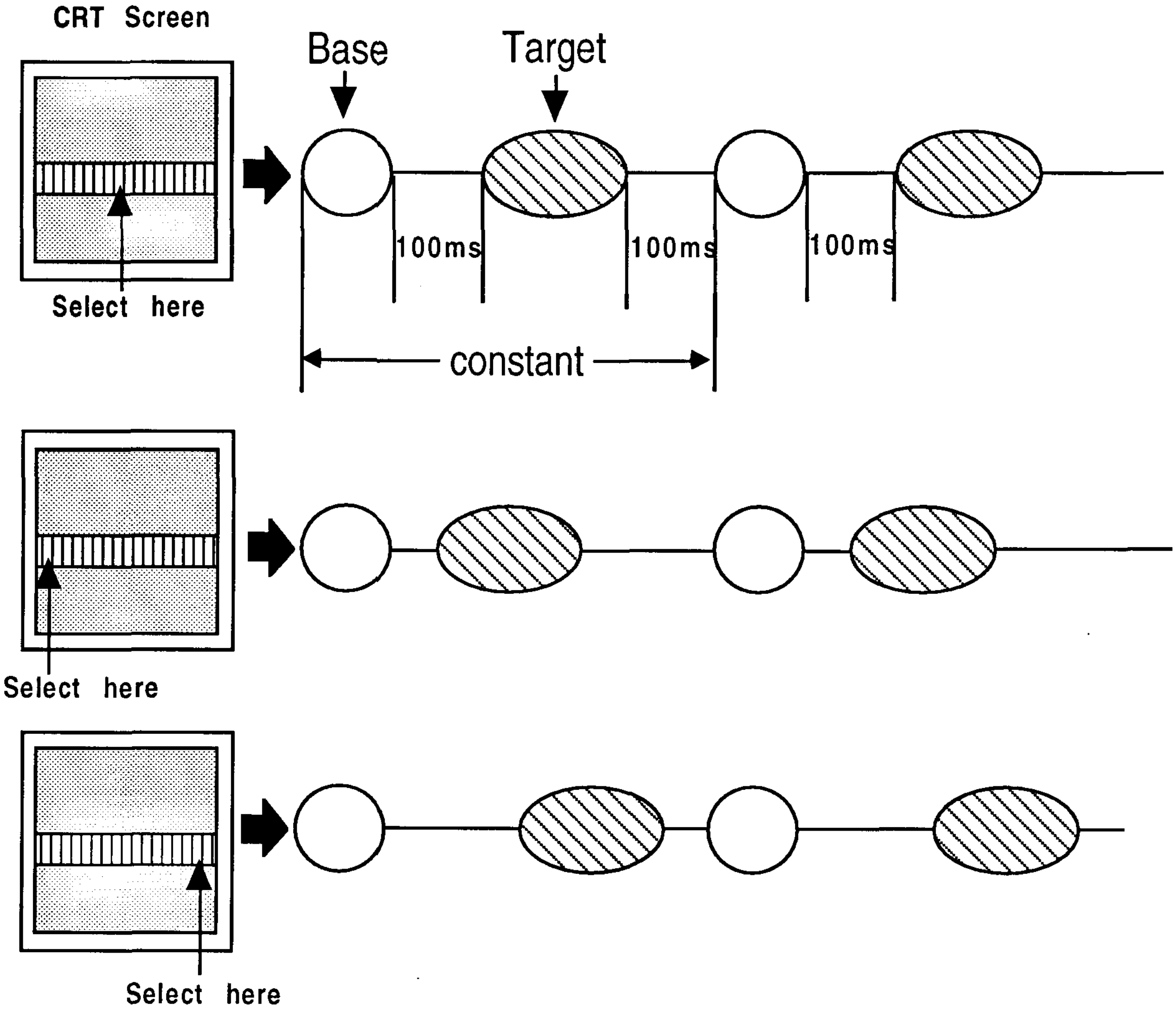


**Figure 1.** An illustration of the perceptual adjustment method. Subjects are presented with an alternating base and target. They then use a mouse to adjust the relative timing of the target with respect to the repeating base and replay the series until they are satisfied with the perceptual isochrony of the series

## 2. Experiment 1.

To eliminate potential problems with longer and shorter overall tokens, this experiment was re-run using stimuli with more similar overall durations. The same stimuli were used as in de Jong (1992); however, the temporal structure of the tokens was varied by matching tokens of *toast* with tokens of *totes*. In addition, tokens of the word *toasts* were also extracted from the same corpus of microbeam data, and matched with stimuli from the earlier experiment.

### *Methods.*

Experiment 1 uses a perceptual adjustment paradigm in which listeners were presented with a base token alternating with a target token as is illustrated in Figure 1. During each trial, the subjects were presented with six syllables, three alternations of a base and a target token. The subjects then could move the targets with respect to the base and replay the six syllables until they were satisfied that the syllables occurred in a regular rhythm ("like soldiers marching"). The subjects used a mouse and a horizontal series of boxes to select from a range of rhythmic possibilities which differed from one another in 3 ms increments. By clicking on boxes closer to the edge of the screen, the subjects chose presentations with a greater interval added or subtracted between the base and target. The same amount of time was either subtracted or added to the duration of time between the target and the base (which had an initial value equal to that between the base and the target), so that the interval of time between each presentation of the base was fixed within each trial. The final intervals between base and target were recorded as the subject responses. The original interval (100ms) was subtracted from each response to give the amount of adjustment performed.

The target stimuli consisted of renditions of the words *toast, totes, and toasts* by a northern midwestern male speaker who was an undergraduate at the University of Wisconsin. The tokens occurred in the frame sentence, *"I said, 'Put the toast/totes/toasts on the table.'"* The speaker had been given a miniature discourse in which the utterance was to appear. In these discourses, the speaker responded to an imaginary hearer who misheard a portion of the utterance by placing nuclear (sentence) accent on the item which the imaginary hearer misheard. Nuclear accent occurred on the target word, on the following preposition, *on*, (all cases in which the target bore a prenuclear accent), and on *put*, (cases in which accent is precluded from being placed on the target word). For a more complete description of this database and the elicitation procedures, see de Jong (1991a).

Tokens of the words, *toast, totes,* and *toasts* were chosen which had roughly equal overall duration. There were four sets of triplets; two sets consisted of nuclear accented tokens; two consisted of unaccented tokens. Figure 2 illustrates the durations of the tokens. Each token was used as a base and as a target, and tokens never occurred as both base and target in the same trial. Thus, for each triplet, each of the three words occurred in combination with one of the other two words, for a total of six pairings. For each listener there were four triplets by six pairings by three trial repetitions, making a total of 72 trials (except that one block of six trials was inadvertently omitted for Subject 2). Listeners were presented with three alternations of the base and the token, and then given as many opportunities as they wished to adjust the base-to-token interval and replay the sequence. Each subject started with one set of six trials as a practice.

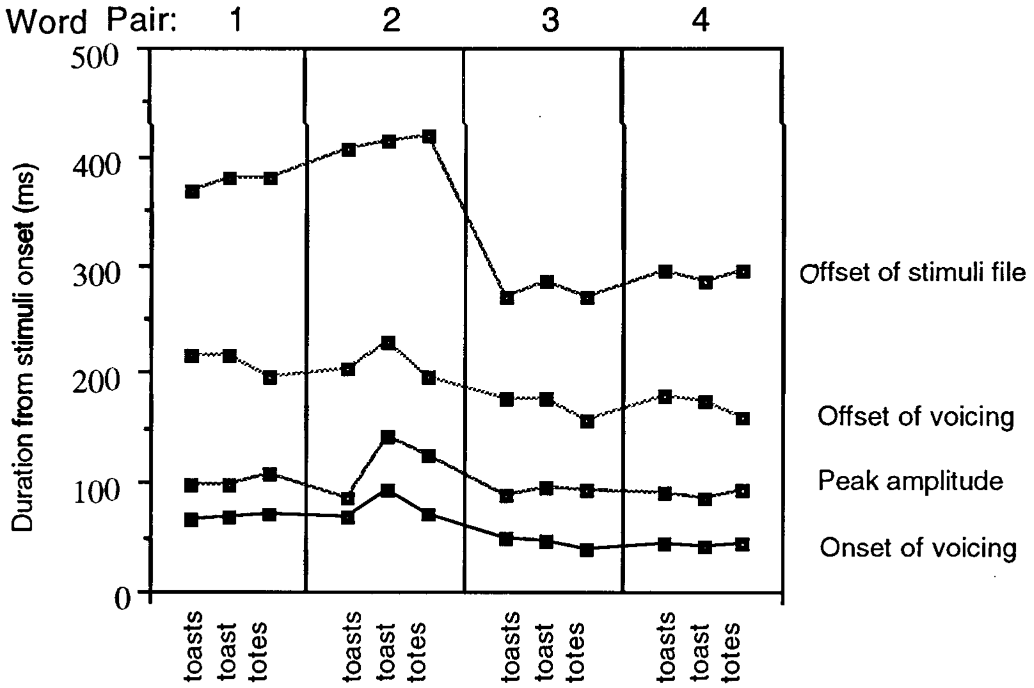


**Figure 2**. The duration of stimuli used in Experiment 1. Shown here are the four sets of three stimuli. The two sets on the left are accented tokens, the two on the right are unaccented.

There were five subjects. Subjects 1 and 2 were the same as in de Jong (1992, the author and one graduate student, respectively). Three additional subjects (S3 - S5) were taken from an undergraduate subject pool, and were paid for their participation in the experiment.

*Results*

Since there was no consistent base across trials, but rather each word was crossed with the other two as a base, the response data was obtained by pairing responses of each word pair with a response for the same word pair presented in the opposite order. The response for one order was subtracted from the response for the opposite order.

To test for consistency in listener responses to each pair of tokens, their responses were subjected to a 3-way ANOVA, with subject, word pair, and token set as factors. The results showed less consistency in listener response than in de Jong (1992). There was a marginal main effect of the token set ($F(3,117) = 3.136$, $0.05 > p > 0.01$). Other main effects were not significant (subject, $F(4,117) = 0.253$; word pair, $F(2,117) = 1.856$; data set, $F(2,117) = 1.867$; all $p > 0.05$). There was a strongly significant interaction between word pair and token list ($F(6,117) = 8.461$, $p < 0.001$). The other two way interactions were not significant (subject X word pair, $F(8,117) = 0.702$; token set X subject, $F(12,117) = 1.263$; both $p > 0.05$). The three way interaction was also marginally significant ($F(24,117) = 1.637$, $0.05 > p > 0.01$). Mean responses for each subject are given in Table 1.

Table 1. Experiment 1. Mean adjustment differences performed by each listener for each target word pair.

**Toast - Totes**

| Word | Listener | | | | | |
|---|---|---|---|---|---|---|
| Set | S1 | S2 | S3 | S4 | S5 | Avg. |
| 1 | 20.0 (10.0) | 3.0 (12.7) | 8.0 (12.2) | 4.0 (19.3) | 12.0 (12.2) | 9.8 (13.6) |
| 2 | 18.0 ( 4.0) | -3.3 ( 5.8) | 1.3 ( 5.8) | -19.3 (13.0) | -14.7 (10.1) | -3.6 (15.3) |
| 3 | -33.3 ( 9.5) | -35.3 (19.2) | -34.0 (32.7) | -13.3 (17.5) | -12.7 (30.3) | -25.7 (22.6) |
| 4 | -6.0 ( 7.2) | 0.0 ( 7.2) | 4.0 (11.1) | 12.0 (13.1) | 9.3 (26.9) | 3.9 (14.3) |
| Avg. | -0.3 (23.6) | -10.0 (19.4) | -5.2 (24.0) | -4.2 (19.0) | -1.5 (22.5) | |

**Toast - Toasts**

| Word | Listener | | | | | |
|---|---|---|---|---|---|---|
| Set | S1 | S2 | S3 | S4 | S5 | Avg. |
| 1 | 2.7 (24.4) | 5.0 ( 7.1) | 5.3 ( 6.1) | 3.3 ( 4.6) | 4.0 (27.5) | 4.0 (14.9) |
| 2 | -6.0 ( 6.9) | 5.3 ( 9.2) | -26.0 (33.0) | 2.0 (30.3) | -17.3 (17.9) | -3.6 (15.3) |
| 3 | -8.7 (13.0) | 3.3 ( 9.0) | 14.0 (27.7) | 5.3 (28.0) | -13.3 (18.1) | -0.1 (20.2) |
| 4 | -0.7 (14.0) | 4.0 ( 8.7) | 4.7 ( 5.8) | 11.3 (13.3) | 28.2 (16.3) | 4.7 (14.1) |
| Avg. | -3.2 (14.3) | 4.4 ( 7.4) | -0.5 (24.5) | 5.5 (19.0) | -5.7 (22.5) | |

**Totes - Toasts**

| Word | Listener | | | | | |
|---|---|---|---|---|---|---|
| Set | S1 | S2 | S3 | S4 | S5 | Avg. |
| 1 | 2.7 ( 2.3) | -12.0 (33.9) | 0.7 ( 1.2) | -2.7 ( 4.1) | -2.0 (30.2) | -2.0 (16.0) |
| 2 | -26.0 ( 9.2) | 6.0 ( 3.5) | -14.0 (22.3) | 10.0 (15.1) | -22.7 (23.4) | -9.3 (20.7) |
| 3 | 20.7 ( 8.1) | 40.7 (20.8) | 46.7 (35.2) | -14.0 (17.8) | 15.3 (31.0) | 21.9 (30.4) |
| 4 | 10.7 ( 9.9) | -4.7 ( 6.4) | -14.7 (27.6) | 6.7 ( 7.6) | 1.3 (18.9) | -0.3 (16.5) |
| Avg. | 2.0 (19.4) | 9.3 (25.7) | 4.7 (33.7) | 0.0 (14.4) | -2.0 (26.6) | |
| Avg. Listener | -5.0 (19.0) | 1.2 (20.2) | -0.3 (27.3) | 0.4 (17.5) | -3.1 (23.3) | |

(Figures in parentheses indicate standard deviations.)

To tease apart the three-way interaction, separate two-way ANOVA's were performed for each subject. Since the different listeners had different amounts of experience with the p-center paradigm, Subjects 1 and 2 having more experience, it is possible that they performed differently in the p-center task. Subjects 4 and 5 made no reliable distinctions, as was evident from the lack of any significant effect on subject response. Subject 1 had a marginally significant ($p > 0.01$) main effect of data set. Subject 4 had a marginal main effect of word pair. The only effect which reached significance at the 0.01 level was the interaction for subjects 1, 2, and 3. Since the other two subjects showed no consistency in their responses to the stimuli, the three way ANOVA was performed again on the remaining three subjects (S1, S2, and S3). In this analysis, there was a strong interaction between word pair and token list ($F(6,69) = 14.000$, $p < 0.001$). None of the other effects were significant at the 0.05 level (data set, $F(3,69) = 1.347$; word pair, $F(2,69) = 2.996$; subject, $F(2,69) = 0.083$; set X subject, $F(16,69) = 0.164$; pair X subject, $F(4,69) = 0.520$; three-way interaction, $F(4,69) = 1.240$). None of the effects involving subject were significant, indicating that these three subjects were responding in the same manner to the stimuli. Thus, the remaining analyses for this experiment will involve the responses of these three subjects only.

As in de Jong (1992), the anisochrony of various events in the initial condition were used to predict subject responses. The same predictors were used, except that the jaw movement predictors were eliminated. The results of these correlations are illustrated in Figure 3. Three things are noteworthy. First, onset-to-onset isochrony (labeled "burst of [t]" in Figure 3) performed very poorly as a predictor, as is expected given that the tokens were matched in overall duration, and there was variation in the subject responses from zero (as evidenced in the significant effect in the ANOVA reported above). Second, the articulatory predictors perform much more poorly than in de Jong (1992). Only one of the predictors explained more than half of the variance in listener adjustment. Third, the other acoustic predictors performed much better than the articulatory predictors.

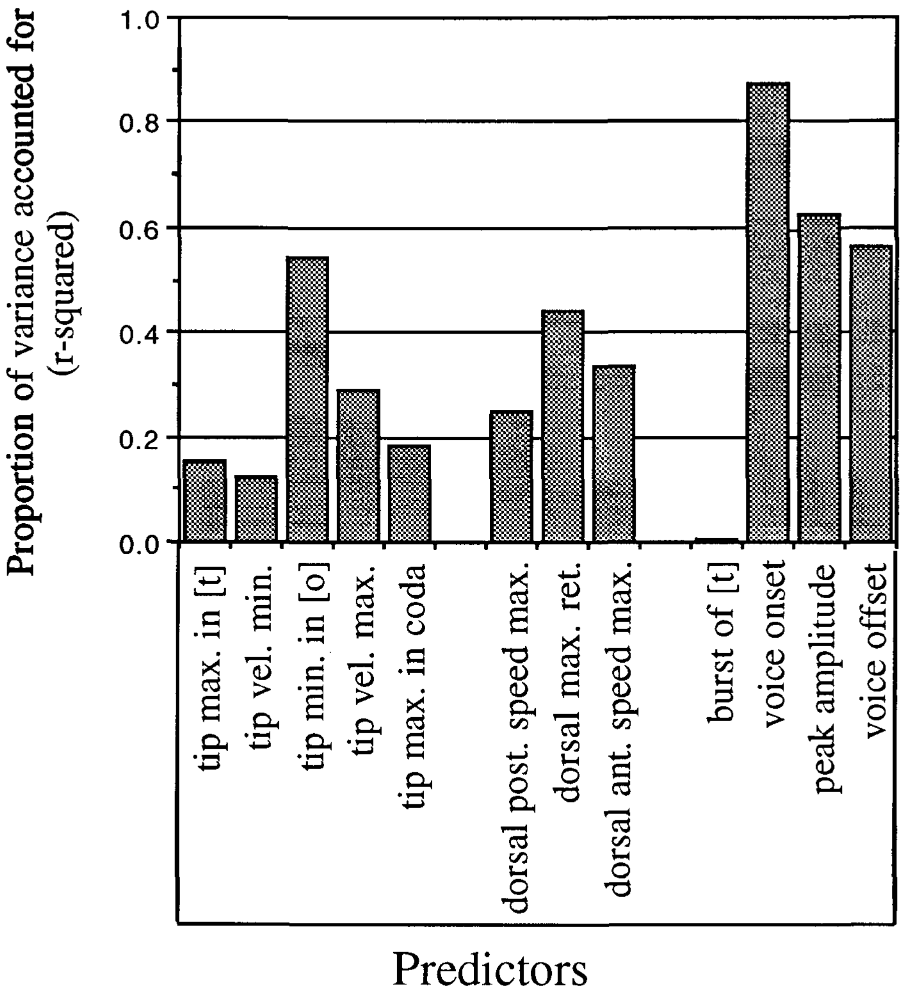


**Figure 3.** Pearson R-squared values for regressions between interevent anisochronies and average perceptual adjustments for Experiment 1.

The best predictor of p-center perception in Experiment 1 is the location of the onset of voicing, predicting almost 25% more of the variance in p-center perception than any other predictor. To test for other factors in the listener responses, they were subtracted from the anisochronies of the voicing onsets. These deviations from voicing onset isochrony were then subjected to a stepwise regression analysis against the duration of various parts of the stimuli, such as those reported in Table 2 of de Jong (1992). Each of the stimuli was segmented into 4 or 5 portions by acoustic, tongue tip, and dorsal events; and the duration of each portion was used to predict the residual variance of the responses from voice-onset to voice-onset isochrony. Three such analyses were performed: using acoustic events, using tip events, and using dorsal events.

With a cut-off of F > 4 in each of the analyses, none of the predictors were included in the model. The greatest partial correlation was between the acoustic duration of the codas and the residual (r-squared = 0.143). Thus, the results of this experiment were not complicated by additional factors. To further compare the adjustments in Experiment 1 to those of de Jong (1992), the perceptual adjustments were subtracted from dorsal velocity peak to dorsal velocity peak isochrony, and the stepwise regression analyses were rerun, using acoustic, jaw, and tip events to mark off the segment durations. Using a cut-off of F > 4, no predictors were included in the analyses involving tip and acoustic events. Two segments were included in the jaw analysis, from jaw minimum to maximum closing velocity (unique r-squared = 0.329), and from maximum closing velocity to highest position of jaw in the coda. Thus, perceptual adjustments in Experiment 1 are most simply modeled by voice onset isochrony.

## 3. Experiment 2.

One possible problem with Experiment 1 involves its use of coda variation to induce variations in p-center location. Earlier studies either had difficulties finding a consistent coda effect on p-center location, such as Fox and Lehiste (1987), or found a considerable amount of inter-listener variability in the coda p-center effect, as in Cooper, et al. (1988). This problem is aggravated by the small amount of temporal difference between the stimuli in Experiment 1. Two listeners produced no consistent adjustments for the coda-varied stimuli. Even for the remaining subjects, it may be the case that the coda p-center effect, which typically is quite small, was simply too small to be detected. To avoid these problems, Experiment 1 was repeated with stimuli taken from a different X-ray microbeam database which included syllables with various onset consonants. Onset variation on p-center location is much more consistently reported in the literature.

Stimuli for Experiment 2 differed from one another in the phonemic aspiration of the initial consonant. Comparing syllables with aspirated and unaspirated onsets is especially useful for comparing a vowel gesture model of p-center location with more acoustic accounts of p-center location. Aspirated and unaspirated onsets differ dramatically in terms of the distance of the voice onset from the release of the consonant. By contrast, the time course of lingual articulation is likely to be quite similar. This relationship is illustrated in Figure 4, which plots the acoustic tracing along with a jaw movement trace for two of the stimuli used in this experiment. As is typical of all of the aspiration minimal pairs, there is a large anisochrony in voice onset in the initial presentation where onsets occur isochronously, while the anisochrony of other articulatory events will be either small (such as the jaw maxima in the onset) or non-existent (such as the jaw minima in the present case).

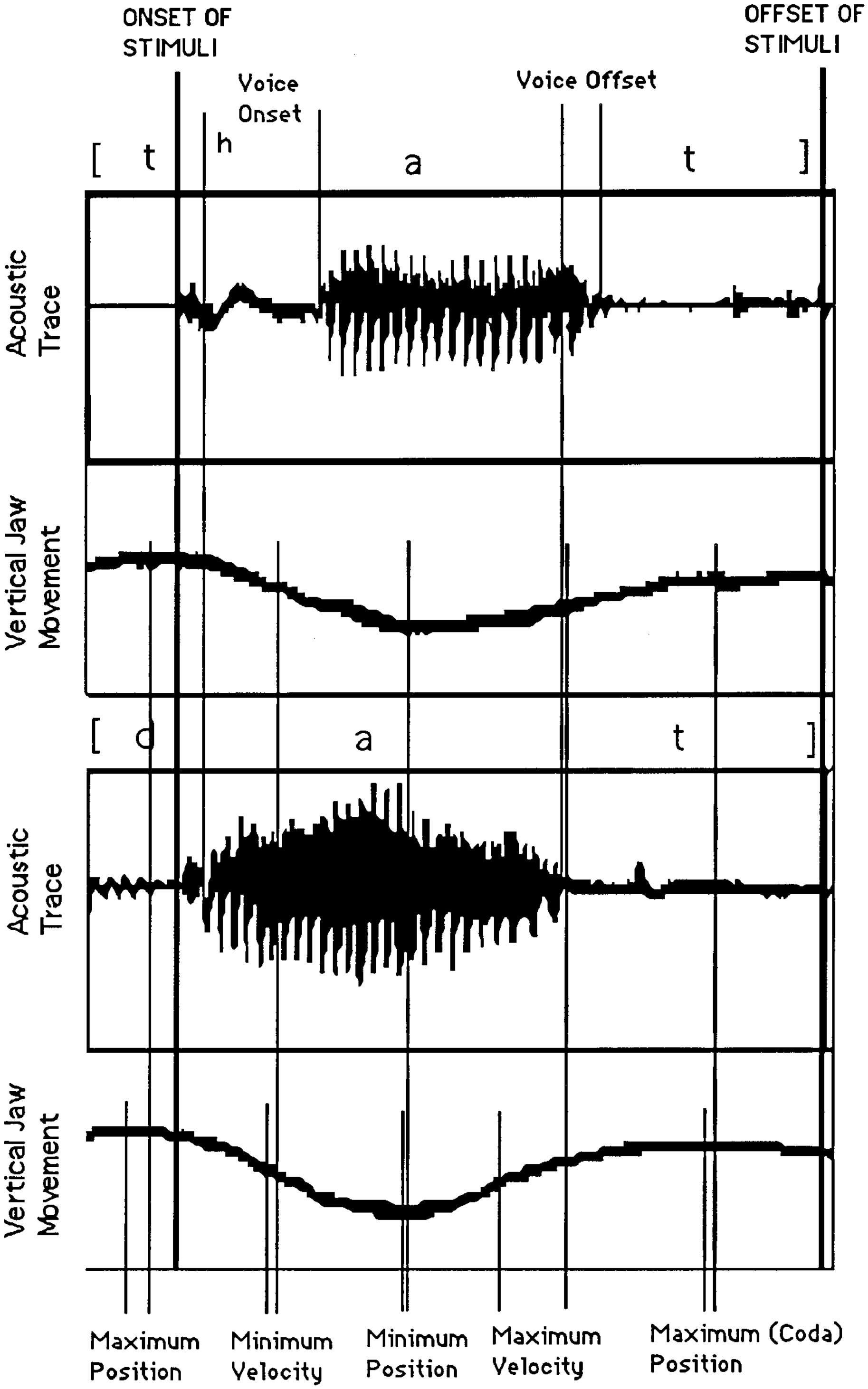


**Figure 4.** An aspirated and unaspirated pair of stimuli from Experiment 2 aligned at the onset of the stimuli file. Below each acoustic trace is a tracing of the vertical movement of the jaw during the stimulus. The anisochrony of voicing onset is indicated by the misalignment of the vertical marks in the upper and lower panels. By contrast, jaw minimum position for these two stimuli are almost isochronous in the initial presentation.

*Methods.*

The stimuli consisted of minimal pairs contrasting initial aspiration, extracted from a different microbeam database to that used in Experiment 1. This database consisted of CVC combinations uttered in the frame sentence, *"Say a _____ of a _____ again."* Nuclear (sentence) accent had been placed consistently on the CVC combination which occurred on the first blank. In order to avoid problems associated with streaming, the CVC combinations chosen were real lexical items; listeners reported no problems with streaming. In order to facilitate articulatory analyses, the stimuli had lingual consonants and low vowels. This restriction entailed that the tongue had to execute a clear movement into the vowel. Word pairs chosen were *gap/cap, gob/cob, gab/cab, dot/tot,* and *dab/tab.* The speaker spoke a St. Louis dialect of American English. For more detailed information concerning this database, see de Jong (1991b).

Stimuli were extracted from their original context, and presented to listeners using the same paradigm as in Experiment 1. Each word pair was presented in both orders. Thus, for each listener, there were five word pairs by two orders by five repetitions for a total of 50 trials. Listeners consisted of three of the subjects from de Jong (1992; S1, S3, and S4; here numbered S1, S6, and S7, respectively) and one more graduate student in linguistics (S8), who was naive to the purposes of the experiment.

Articulatory and acoustic measures were taken from the stimuli as in Experiment 1. The release of the initial stop, the onset of voicing, the offset of voicing, and two events in the amplitude envelope: peak amplitude, and a point at which the amplitude curve ceased increasing sharply at the beginning of the vowel were marked as potentially interesting acoustic events. Jaw movement events were included as an index of the opening of the vocal tract. Since not all of the initial consonants were coronals, the tip trajectory could not be relied upon for indexing the size of oral opening as in Experiment 1. Three events in the consonant articulator movement (where consonant articulator = tongue tip in [d] and [t], but consonant articulator = tongue mid-section in [g] and [k]) were also marked: maximum position for the initial consonant, minimum position in the vowel, and the speed maximum in between these two points. Three events were extracted from the movement of the tongue mid-section and also for the tongue dorsum for the articulation of the vowel: the point of maximum lowering during the vowel, and maximum speed points into and out of this position. As in Experiment 1, anisochronies in these events were correlated with the average responses for each listener.

*Results.*

To determine whether listeners performed different adjustments for different tokens, the listener responses were subjected to a three-way ANOVA with listener, word order (aspirated or unaspirated first), and word pair as factors. There was a strong main effect of word order ($F(1,160) = 700.90$, $p < 0.0001$) and an interaction between word order and listener ($F(3,160) = 27.22$, $p < 0.0001$). There was also a weaker, but still significant main effect of word pair ($F(4,160) = 3.95$, $p < 0.01$), as well as an interaction between word pair and word order ($F(4,160) = 4.82$, $p < 0.01$), and three-way interaction ($F(12,160) = 2.54$, $p < 0.01$). Other effects were insignificant. To insure that all of the listeners were giving different responses to the different tokens, each listener's data was subjected to a two-way ANOVA with word pair and word order as factors. Each listener had main effects of word order with $p < 0.0001$. In addition, S1 and S7 had significant interactions (at the 0.05 level) between word pair and word order. Thus, although there were some differences between subjects, they were consistent in making different adjustments for aspirated and unaspirated tokens. Means for each listener are given in Table 2.

Table 2. Experiment 2. Mean adjustments performed by each listener for each target word pair in each order.

**GAB / CAB**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -27.6 ( 4.6) | -14.4 ( 3.6) | -24.0 ( 7.5) | -16.0 ( 6.8) | -20.5 ( 7.8) |
| aspirated | 28.0 ( 7.9) | 15.6 ( 3.8) | 15.2 ( 4.8) | 7.6 (17.2) | 16.6 (11.8) |

**GOB / COB**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -19.2 ( 4.8) | -14.4 ( 2.2) | -22.8 ( 7.3) | -5.2 ( 9.2) | -17.9 (11.0) |
| aspirated | 25.2 ( 8.0) | 12.4 ( 4.6) | 8.8 ( 6.6) | 1.6 (12.4) | 12.0 (11.7) |

**GAP / CAP**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -14.0 ( 5.8) | -13.6 ( 9.1) | -8.4 ( 9.8) | -4.8 ( 6.6) | -10.2 ( 8.3) |
| aspirated | 19.6 ( 7.0) | 10.4 ( 9.7) | 10.0 (13.7) | 5.2 (12.9) | 11.3 (11.6) |

**DOT / TOT**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -17.6 ( 4.6) | -12.4 ( 5.9) | -32.0 ( 4.9) | -17.8 ( 5.0) | -20.7 ( 8.2) |
| aspirated | 14.8 ( 4.4) | 3.6 ( 4.8) | 15.6 (12.8) | 6.0 ( 7.2) | 10.0 ( 9.1) |

**DAB / TAB**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -20.0 ( 8.8) | -15.6 ( 4.3) | -18.0 ( 9.8) | -6.0 ( 7.3) | -14.9 ( 9.1) |
| aspirated | 19.6 ( 3.8) | 13.2 ( 1.8) | 22.4 ( 7.9) | 9.6 (12.5) | 16.2 ( 8.8) |

**AVERAGE**

| First | Listener | | | | |
|---|---|---|---|---|---|
| token | S1 | S6 | S7 | S8 | Avg. |
| unaspirated | -21.7 ( 8.1) | -14.9 ( 5.2) | -21.1 (10.8) | -9.8 ( 8.6) | -16.9 ( 9.6) |
| aspirated | 21.4 ( 7.6) | 11.0 ( 6.6) | 14.4 (10.2) | 6.0 (12.0) | 13.2 (10.8) |

(Figures in parentheses indicate standard deviations.)

Regressing average subject responses against the various articulatory and acoustic predictors yields the results illustrated in the bottom panel of Figure 5. As in Experiment 1, the best predictor of p-center location is the onset of voicing. Most articulatory predictors perform very poorly, especially those taken from the earlier half of the syllable. An exception is the maximum position for the jaw during the onset consonant. Curiously, the next best articulatory predictor of p-center location is the timing of the velocity peak of the tongue body moving *out of* the vowel in the second half of the syllable. Thus, even though the p-center adjustments were larger and more consistent in the present stimuli than for those in Experiment 1, the best predictor of their location is the onset of voicing for the vowel.

To assess the cross-listener consistency of this predictor, separate regressions were performed for each listener. The results are summarized in Table 3. The pattern of results is, on the whole, consistent across listeners. Acoustic predictors perform better than articulatory predictors. Subjects 6 and 7 gave responses best predicted by the location of the end of the amplitude rise for the vowel -- another acoustic predictor.

Table 3. Experiment 3: Pearson R-squared values for regressions between interevent anisochronies and perceptual adjustments.

| Events | Listeners | | | | |
|---|---|---|---|---|---|
| | S1 | S3 | S4 | S8 | Avg. |
| Jaw maximum in onset | 0.903 | 0.907 | 0.800 | 0.776 | 0.747 |
| Jaw minimum in vowel | 0.100 | 0.086 | 0.039 | 0.133 | 0.072 |
| Jaw maximum in coda | 0.035 | 0.013 | 0.015 | 0.075 | 0.025 |
| Jaw velocity minimum | 0.167 | 0.080 | 0.035 | 0.002 | 0.064 |
| Jaw velocity maximum | 0.065 | 0.052 | 0.063 | 0.215 | 0.070 |
| C-articulator maximum in onset | 0.175 | 0.153 | 0.040 | 0.060 | 0.094 |
| C-articulator minimum in vowel | 0.147 | 0.063 | 0.048 | 0.047 | 0.072 |
| C-articulator maximum in coda | 0.140 | 0.068 | 0.138 | 0.148 | 0.111 |
| dorsal extremum in vowel | 0.134 | 0.169 | 0.025 | 0.055 | 0.078 |
| dorsal speed into vowel | 0.193 | 0.131 | 0.016 | 0.002 | 0.068 |
| dorsal speed out of vowel | 0.234 | 0.302 | 0.394 | 0.574 | 0.303 |
| mid extremum in vowel | 0.552 | 0.402 | 0.257 | 0.182 | 0.330 |
| mid speed into vowel | 0.132 | 0.106 | 0.020 | 0.004 | 0.055 |
| mid speed out of vowel | 0.370 | 0.568 | 0.684 | 0.694 | 0.592 |
| burst of onset C | 0.146 | 0.221 | 0.346 | 0.239 | 0.205 |
| voice onset | 0.961 | 0.934 | 0.892 | 0.751 | 0.816 |
| end of sharp amplitude rise | 0.852 | 0.871 | 0.919 | 0.894 | 0.796 |
| peak amplitude | 0.829 | 0.461 | 0.348 | 0.311 | 0.384 |
| voice offset | 0.239 | 0.199 | 0.123 | 0.244 | 0.174 |

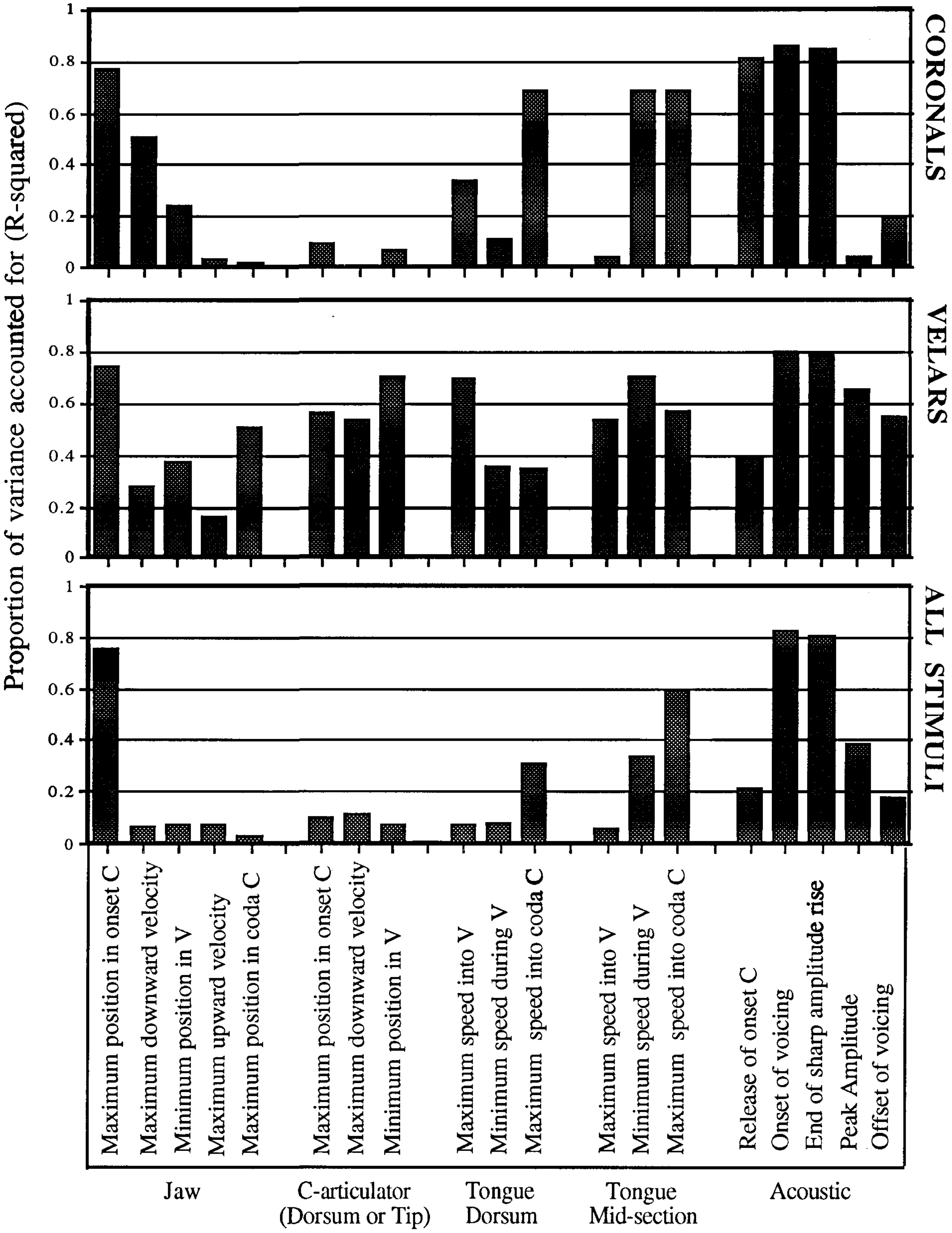


**Figure 5.** Pearson R-squared values for regressions between interevent anisochronies and average perceptual adjustments for each listener in Experiment 2.

Since the stimuli in the present experiment differ in the point of articulation of the initial consonant, it is possible that the measures of tongue movement performed so poorly because they are inconsistent in their relationship with the underlying gestures for the vowels. For example, the mid-section pellet movement will be more strongly affected by the initial consonant in the velar cases than in the coronal cases. Thus, the velocity maxima in the velar cases will likely occur later with respect to the initial consonant than in the coronal cases. To avoid the problem of such inconsistencies due to consonant place of articulation, the responses to velar initial stimuli were separated from the responses to coronal initial stimuli, and the regressions were performed again. The results of these separate analyses are illustrated in the top two panels of Figure 5. With the coronal cases eliminated, the tongue body movement predictors perform much better; however, they still do not perform as well as the voicing onset.

Considering these results in more detail, the difference in effectiveness of the voice onset predictor and the tongue body velocity predictor lies in the large difference in listener responses to aspirated and unaspirated stimuli. Figure 6 plots the average responses of each listener to each word pair against the response predicted by voicing onset isochrony and tongue mid-section velocity peak isochrony. The upper panel shows that the prediction by the voicing onset predictor, that listeners will adjust aspirated and unaspirated tokens in opposite directions, is borne out. Tokens lie in the upper right and lower left quadrants, not in the upper left and lower right quadrants. The tongue body predictor in the lower panel, however, does not predict this result. There is a general tendency for the velocity peaks to be later in the aspirated velars -- yielding a correlation with listener adjustments in the right direction for velars only. In one respect, however, the results illustrated here favor the velocity peak predictor. The slope of the regression line in the lower panel is very close to 1. Anisochronies in velocity peaks are compensated in a one-to-one fashion by listener adjustments. The slope of the regression function in the upper panel, by contrast, is far less than one -- indicating that listeners undershoot adjustment for voicing onset isochrony.

It is possible that a voice onset factor is combined with the effects of some other p-center factor. To test for multiple factors in p-center location, the average responses for each listener were subjected to a stepwise regression against all of the event predictors used in the simple regressions above. Using a criterion of $F > 4$, only voicing onset was included in the model. The next highest increment would occur with the inclusion of the end of the amplitude rise (r-squared = 0.065). Thus, there are no demonstrably independent additional factors to voicing onset.

## 4. Discussion and Conclusion.

To summarize, then, the present experiments have sought to reproduce earlier perceptual p-center studies using naturally varied acoustic stimuli with articulatory recordings. Results of de Jong (1992) were partially consistent with a vowel gesture account of p-centers in that the timing of tongue dorsum velocity peaks predicted perceptual adjustments better than acoustic markers did for one of the stimulus words. However, the same predictor in the present experiments does not perform as well. The other stimulus word in de Jong (1992) pointed toward a different articulatory predictor – the timing of maximum tongue-tip lowering. Experiment 2 points toward yet another predictor which could be associated with the vowel gesture, that of the maximum position of the jaw. This predictor performed reasonably well in de Jong (1992), though not as well as other articulatory predictors in the first half of the syllable, and not consistently across listeners.

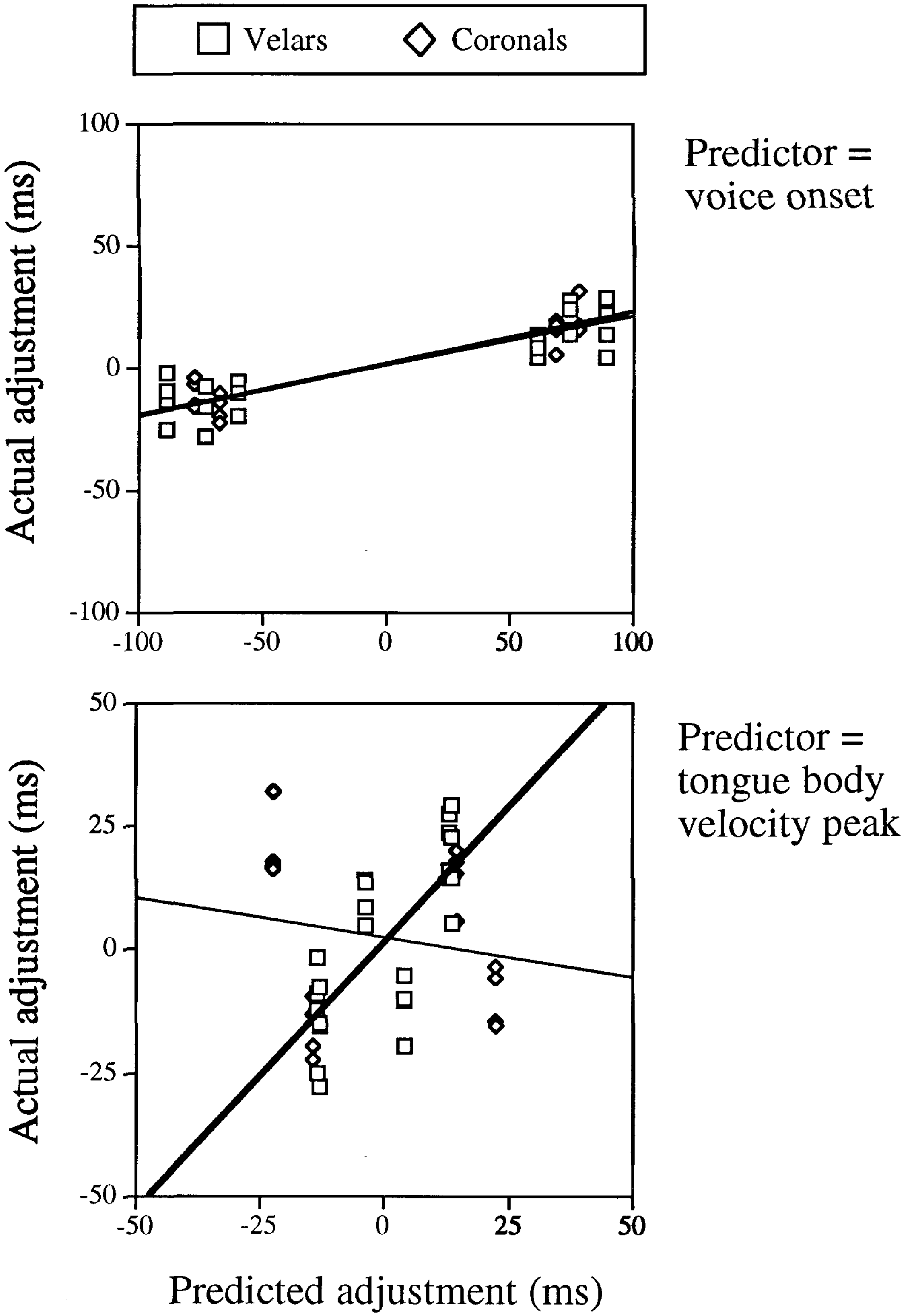


**Figure 6.** Average adjustment for each listener plotted against the adjustment predicted by voicing onset (upper panel) and tongue mid-section peak velocity going into the vowel (lower panel). Regressions lines for coronals are light; those for velars are heavy. Regression lines in the top panel lie atop one another.

Thus, there are reasonably good kinematic correlates of p-center location; however, they do not perform consistently across the experiments. The results of the present set of experiments do not support a model in which p-center location corresponds to a particular articulatory kinematic event.

Evaluating the present results with respect to a model of p-center location corresponding to the timing of an abstract, underlying gesture, such as those proposed in Articulatory Phonology (Browman and Goldstein, 1986; 1990) is more difficult. In this model, observed articulatory movement can be the result of the forces exerted by more than one active underlying gesture at any given time. Thus, the kinematic events must be considered as possible indexes of the underlying gestural events, not as parts of the gestures themselves. Kinematic events in tongue-body movement should clearly index an underlying vowel gesture to the extent that the tongue body movement is not being differentially affected by the presence of neighboring consonants. This indexing is likely to be most clear when there are successive demands on that articulator which are incommensurate. One example of this is tongue body advancement during the production of coronal consonant - back vowel sequences, such as in the stimuli used in de Jong (1992). The coronal consonant demands an advanced tongue body supporting the constriction of the vocal tract at the alveolar ridge, while the back vowel demands a retracted tongue body to make a velar or uvular constriction. Another example is that of the tongue mid-section in velar-low vowel sequences, as in some of the stimuli used in Experiment 2. Velar consonants demand a high position of the tongue body to make an occlusion at the velum, while low vowels demand an open oral cavity for the low vowels. In these two sets of examples, there are fairly large movements of the tongue pellets, and thus, the velocity peaks are readily apparent, and are reasonably interpreted as indexing a consistent location within the underlying vocalic gesture. To the extent that these two predictors, dorsal speed in a [to] sequence and mid-section speed in a velar-low vowel sequence have performed well, there is some support for a gestural account of p-center location.

Also promising for a gestural account is the performance of the dorsal velocity predictors in the later half of the syllable in Experiment 2. Such results suggest that there might be an articulatory relationship between events in the later half of a syllable and the timing of articulatory behavior in the first half of the syllable. Finding and describing such a relationship is necessary for any purely articulatory account to explain coda effects on p-center location.

However, one result of the present experiments poses a large problem for an articulatory analysis of p-centers. The most consistent predictor of p-center perception in the present experiments is an acoustic measure, the onset of voicing. For Experiment 1 (as well as that in de Jong, 1992), one could say that the acoustic predictor -- voicing onset -- is indirectly indexing vocalic gestures. Voicing onset indexes glottal gesture timing, which in turn may be synchronized with some aspect of the oral gestures for consonants and vowels. However, for Experiment 2, this approach does not work. The glottal gesture is phonemically specified to have grossly different timing with respect to the lingual gestures associated with the consonant and vowels. Experiment 2, then, suggests that p-center subjects are locating the timing of glottal gestures in particular. Although, one might argue that global tongue body gestures for vowels have a special function in the organization of speech events (see Fowler, 1983 for several arguments for this view), no such arguments are available for glottal gestures being central to the temporal organization of speech.

Note that the present results do not point unequivocally to an acoustic locus of p-center location, either. The extensive literature on p-center location has produced fairly complicated acoustic predictors of p-center location. The present experiments' results at first blush point to a fairly simple p-center location, voice onset. Thus, it would be tempting to blame the complicated results of earlier studies on the fact that their stimuli had been digitally edited. However, the present results are not that simple. For example, Figure 6 shows not just that voice onset anisochronies predict listener adjustments fairly well, but it also shows that listeners consistently undershoot the amount of adjustment needed for voicing onset isochrony. This is apparent in the slope of the regression line being far less than one. Thus, it is not the case that listeners are simply aligning voice onsets. This result is consistent with the results of earlier articulatory studies (especially Rapp, 1971, and Fowler and Tassinary, 1981). More direct acoustic modeling of the present results is necessary before one can draw firm conclusions about acoustic correlates of p-center location.

One general methodological conclusion can be drawn. If experimental studies are to effectively evaluate gestural accounts of p-center location, the stimuli used must be kept as natural as possible. Given the present state of our knowledge of the dynamics of natural speech production, it is difficult to assess which aspects of the speech signal might bear relevant information about rhythmically important aspects of speech production. Similarly, the nature of the psychological link between production and perception which plays a role in a gestural account of p-center adjustments needs further investigation. The streaming effects which occurred on occasion in the experiments reported here suggest that this psychological link may be quite fragile, and care must be exercised to keep it in tact during experimental sessions.

## Acknowledgments.

I would like to thank Keith Johnson for extensive aid in developing the experiment running software, as well as for extensive discussion of the research reported here. I acknowledge also the contributions of Peter Ladefoged, Doug Whalen, and two anonymous reviewers in helpful comments on earlier versions of this paper. Also thanks go to Victoria Anderson, Barbara Blankenship, Aaron Shryock, and Richard Wright for allowing themselves to be subjected to the P-center paradigm. Work supported by the NIDCD under grant number T32DC00029.

## References.

BROWMAN, C.P. and GOLDSTEIN, L. (1986). Towards an articulatory phonology. *Phonology Yearbook,* **3**, 219-252.

BROWMAN, C.P. and GOLDSTEIN, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299-320.

BREGMAN, A.S. (1990). *Auditory Scene Analysis: the Perceptual Organization of Sound.* Cambridge, Mass.: MIT Press.

COOPER, A.M., WHALEN, D.H. and FOWLER, C.A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, **16**, 231-241.

DE JONG, K.J. (1991a). The articulation of consonant-induced vowel duration changes in English. *Phonetica*, **48**, 1-17.

DE JONG, K.J. (1991b). *The Oral Articulation of English Stress Accent.* Unpublished Ph.D. dissertation, Ohio State University, Columbus, Oh.

DE JONG, K.J. (1992). Acoustic and articulatory correlates of P-center perception. *UCLA Working Papers in Phonetics*, **81**: 66 - 75.

FOWLER, C.A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, **102(3)**, 386-412.

FOWLER, C.A., and TASSINARY, L. (1981). Natural measurement criteria for speech: the anisochrony illusion. In Long, J., and Baddeley, A. (eds.), *Attention and Performance, IX.* Hillsdale, N.J.: Earlbaum.

FOX, R.A., and LEHISTE, I (1987). Effect of unstressed affixes on stress beat location in speech production and perception. *Perceptual and Motor Skills*, **65**, 35-44.

MARCUS, S.M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception and Psychophysics*, **30**, 247-256.

RAPP, K. (1971), A study of syllable-timing. *Speech Transmission Laboratory, Quarterly Progress and Status Report* (Stockholm), 1971(1), 14-19.

# Customized 3-D Electropalatography Display

Cheng Cheng Saw

## Introduction

This paper describes a method for displaying dynamic electropalatographic data in a three-dimensional display scaled to the shape and size of the individual's palate. Computer-based electropalatography (EPG) is a method by which palatal-lingual contact can be recorded and dynamically displayed on the computer screen. The Kay Elemetrics Palatometer is one such EPG system. Using a clear retainer-like pseudo-palate implanted with 96 electrodes, the Palatometer stores contact information in a binary file. The file is usually then opened and viewed in the Computer Speech Laboratory (CSL). As in conventional EPG systems, the data can only be viewed on a standardized 2-D display. Data from a subject with a huge and steeply sloped palate is shown on the same display as data from another subject with a narrow and flat one. Yet, an individually-scaled 3-D display is often necessary to attribute aspects of the data to individual anatomy. Figure 1 illustrates the contrast between the two kinds of displays. Thus the goal of the project described here is to display the time-varying contact information on a roughly customized 3-D image of the palate allowing views from all perspectives, one that can be presented in a movie format. Chui and Shadle (199x) proposed a very sophisticated approach to a similar task resulting in a very different display.

This project was performed on a Macintosh Quadra with a scanner, an IBM compatible PC with CSL and the Kay Elemetrics Palatometer, and an IBM RS/6000 workstation. The IBM PC was used to gather the contact information. The RS/6000 had the Data Explorer program to display and manipulate images. The Macintosh had QuickTime, MacroMind Director, Adobe Premiere and SoundEdit Pro for producing the final movie. An FTP program was necessary to move text and TIFF files across all the machines.

The main visualization program used, Data Explorer, is a product of IBM which runs under UNIX on various workstations. It is a specialized program to provide enhanced methods of viewing data in 3-D. It can either display an object given a set of positions with a list of connections, or display data given another set of positions with a specified data value. Using Data Explorer, the palate was constructed as a wire-frame object defined by a set of x-y-z values interconnected by triangles. Likewise, the electrode contact data appear as little spheres superimposed onto the wire-frame object. The combined image of a sample or single frame of contact is saved as a TIFF file; the entire sequence of TIFF files from an utterance is then transferred onto the Macintosh and processed. The next two sections describe how to compose the necessary text files before the object and the data can be viewed. An explanation of how to generate frames of images and how to put together a movie then follow. Data collected for a flaps and taps experiment were processed in this way. A discussion of the results of that study concludes the paper.

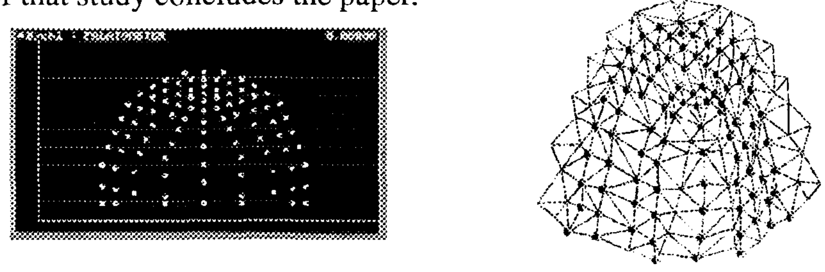


**Figure 1** Standardized 2-D vs customized 3-D palate display

## The Object: Defining the Positions and Connections of the Palate

The 3-D object was developed from the EPG pseudo-palate for speaker AK. It was defined by a set of x-y-z points specifying positions on the surface of the palate. The x axis was defined to be the width which runs horizontally from side to side, the y axis was the length which runs from the anterior to the posterior, and the z axis was the height which runs from the bottom of the teeth to the roof of the palate. First the positions of the 96 electrodes were collected. Then the positions of some additional points were filled in, roughly defining the general shape of the palate.

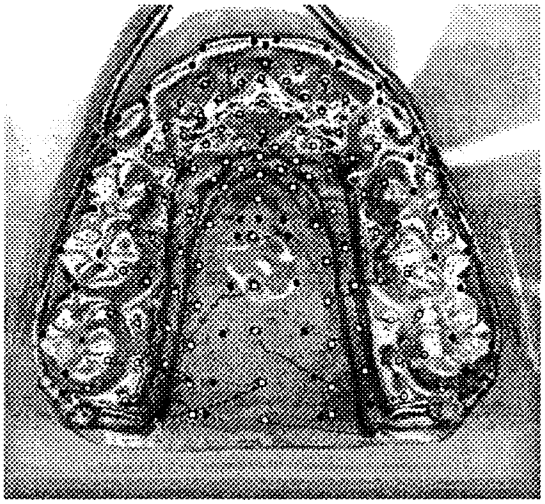


**Figure 2** Scanned image of palate with measured points for x and y values

Calculating the x and y values was quite straightforward. The palate was scanned in as an image scaled at 200%, as shown in Figure 2, to reduce the likelihood of error. The resulting PICT file was opened in SuperPaint. The SuperPaint rulers were shown and the unit of measurement set to millimeters. The horizontal ruler crossed the top of the page so the anterior edge was at 0 and the vertical ruler crossed the left end of the page so the furthest left tooth in view was 0. Little grey circles were placed at the location of the electrodes as they were measured.

Calculating z values was a bit trickier because it required measuring the height in sagittal planes. An impression was made from a cast of the speaker's palate. The impression material only reached up to the teeth to thinly cover their surface. It was then cut along the mid-sagitttal plane, placed along a straight line and traced. Figure 3 shows the trace that was scanned in at 200%. Vertical measurements were taken at y positions where the electrodes were located. The plane of the bottom

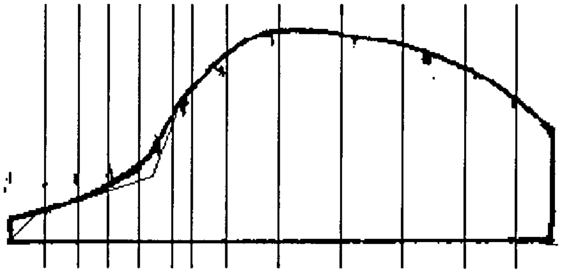


**Figure 3** Mid-sagittal plane for measuring z values

of the teeth has value 0 and the z values increase positively towards the roof of the mouth. The z values of the rest of the electrodes can be found similarly by making additional slices along the impression. However, in the present study they were actually measured by finding the electrode's relation to the z values of the mid-sagittal electrodes.

After the x-y-z values of the electrodes were measured, an additional 51 points were chosen, especially around the teeth and the roof of the mouth. These points were marked with black circles for reference as well. The 147 points measured, including the 96 electrode positions, are shown in Figure 2 as grey and black circles. The z measurements, as with the electrodes, are estimated from looking at the pseudo-palate and finding each point's relation to the z values of the mid-sagittal electrodes.

The final set of positions was compiled in the text file "ak.palpos", with a separate line for the x-y-z values for each of the 147 points measured on the palate. The first few lines are listed in Appendix A. All text files used here consist solely of characters and spaces, no tabs; additionally,

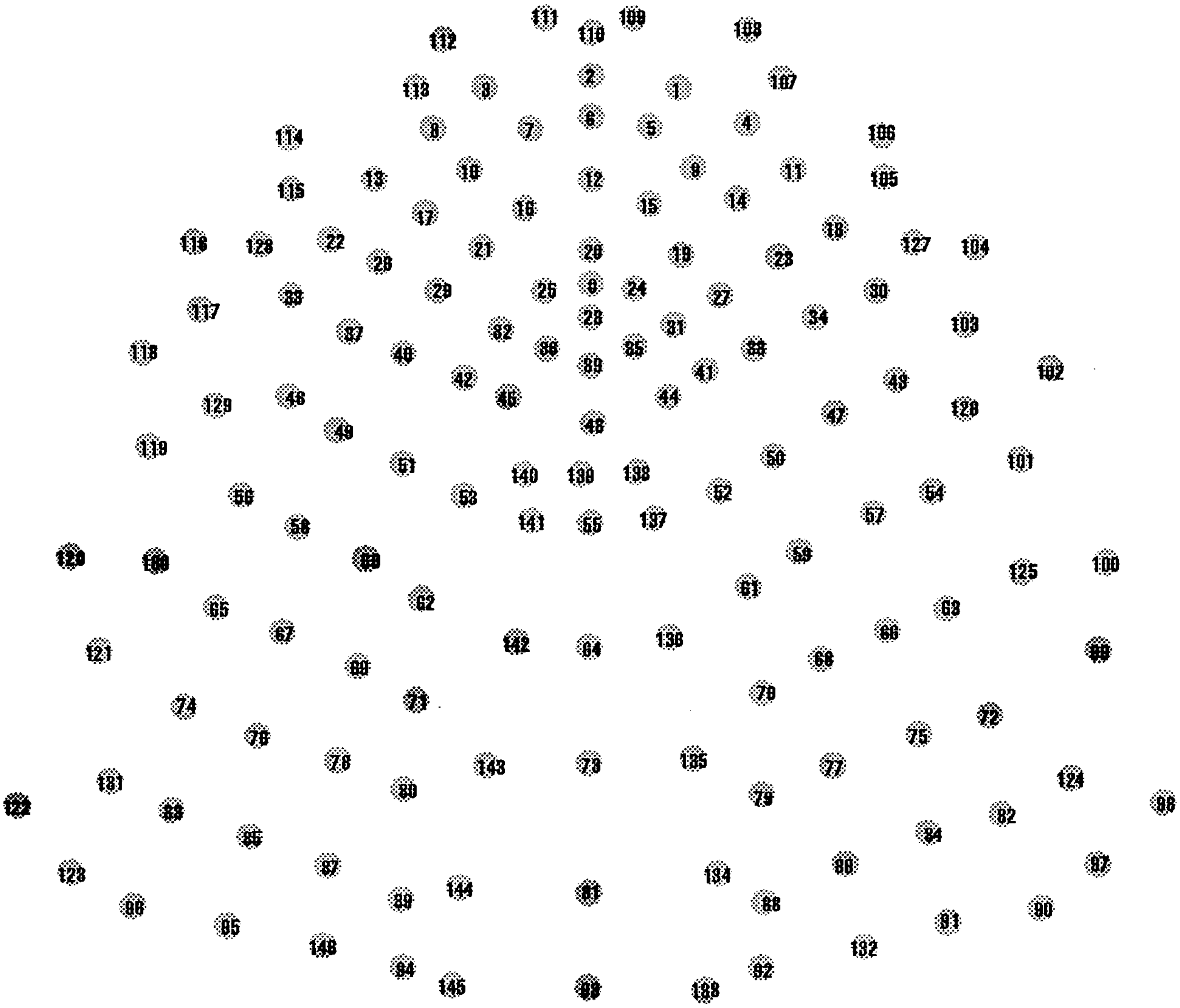


**Figure 4.** Numbering system for electrodes and additional points.
1 through 96 are electrodes; 0 and 97 through 146 are the other points.

there should be no other information such as headers or labels. However, label are provided in the Appendices for ease of reading. Just for ease of record keeping, the position of a non-electrode point was listed as point 0, so that electrodes 1 through 96 could be referred to as points 1 through 96, even though they are on lines 2 through 97. The 1 through 96 numbering of electrodes is provided by the Kay Palatometer. The complete numbering system used is shown in Figure 4.

Another text file "ak.palcon" was necessary to define how the points are connected to form faces on the palate. Since the points tend to be irregularly spaced, the faces were defined by triangles which are easier to specify compared to rectangles. On the scanned image where the circles were marked, a copy of those circles was printed. The dots were then connected with 254 triangles in a way that preserved the structure and depth quality of the object. The palate was split into smaller regions as shown in Figure 5. "ak.palcon" is the result with the three corners of each triangle on a separate line. Its first few lines are also listed in Appendix A. Each corner is listed by the number assigned in the positions file "ak.palpos". Now, there is enough information from the two text files specifying the positions and connections of the palate to display the palate as a wire-frame object in 3-D.

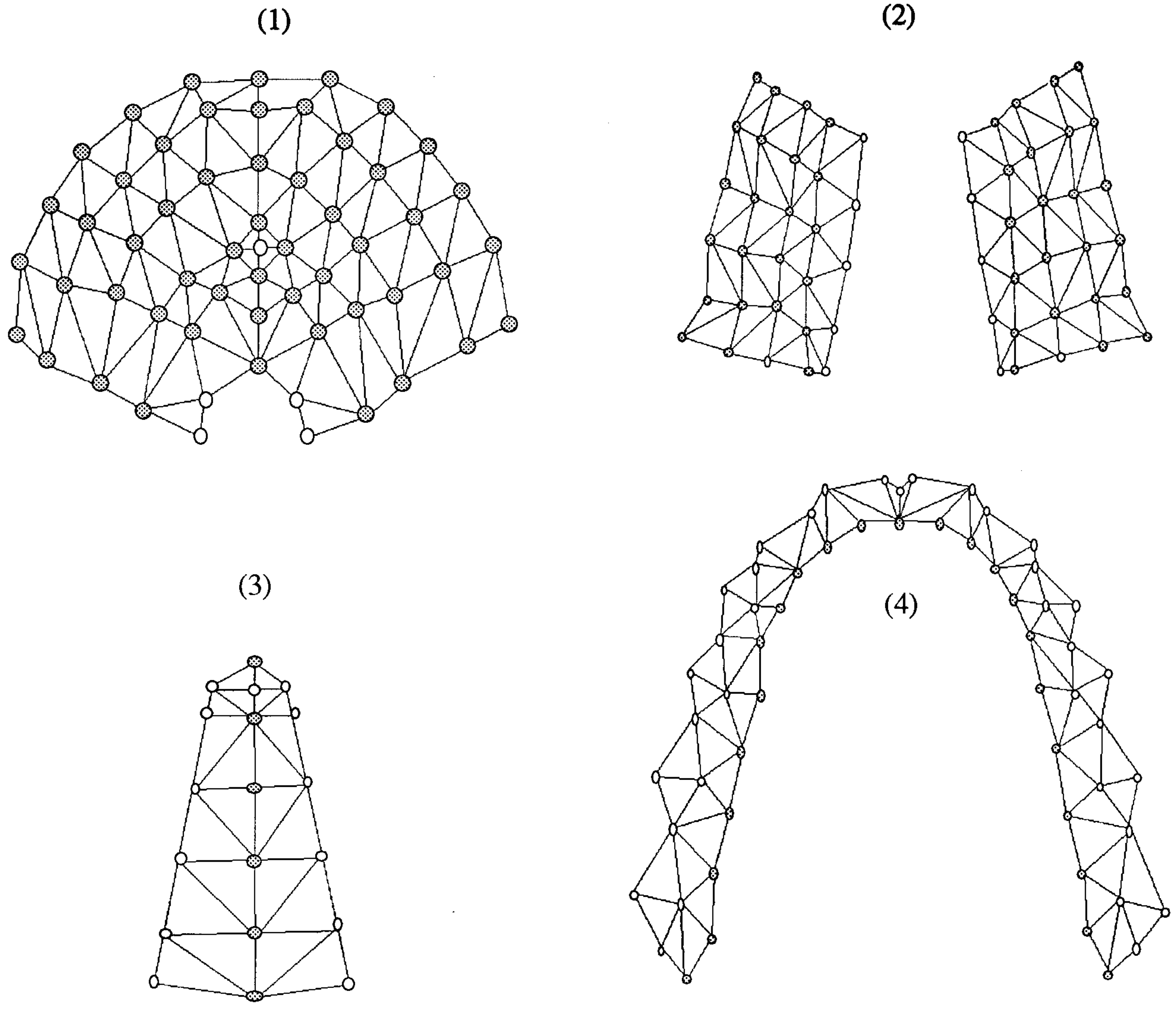


**Figure 5.** Triangulation of Alveolar Ridge (1), Sides (2), Roof (3), and Teeth (4).

## The Data: Defining the Positions and Data Values of the Electrodes

The palate by itself does not offer any information regarding the location of the electrodes at which palatal-lingual contact data is collected. To display electrodes as little spheres on the wire-frame palate, a set of positions and data values for each point is required. For the positions, a copy of the positions of points 1 through 96 from the palate's position file, corresponding to the electrode positions, was used. The data values at these positions for a given utterance were interpreted from CSL files with the .NSP extension. At each time point, a data value of 1 indicates that there was contact made at the given electrode whereas a value of 0 indicates no contact. A C program, ELECDAT.C read the binary file, for example "A1FLAP.NSP", searched for the beginning of the contact information, and output a text file, for example "A1FLAP.BIN", containing sets of 0s and 1s. A listing of this program is given as Appendix B. There is one set of 96 values for each 10 msec sample in the utterance. "A1FLAP.BIN" is an example data file illustrating a flap in "party". The first 5 samples of the text file are listed in Figure 6. The whole utterance has 240 samples (2.4 msec) and the entire output runs ten pages of 0s and 1s.

The 96 data values in each sample represent the state of the 96 electrodes on the pseudo-palate. Yet, the order of 0s and 1s (called the data channels) does not match the order of electrode numbers. For example, the data value for electrode 1 could be the 56th channel, in other words the 56th data value in the set of 0s and 1s, and the data value for electrode 2 could be the 85th channel. The possibilities are seemingly random and vary from speaker to speaker because the pseudo-palates were individually hand-fabricated. Therefore, the electrode channel mapping must be found in the speaker's user file provided by Kay Elemetrics, which for this speaker is "AK.USR". The data positions text file "ak.elecpos" was then sorted by channel mapping instead of electrode numberings. The first few lines of this file is also listed in Appendix A. With the text files containing data positions sorted by channel numbers and the large data list of 0s and 1s, there is enough information to display contact information of a given utterance as gathered from an electropalatography system.

```
1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0
1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1

1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0
1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1

1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0
1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1

1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0
1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1

1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 0 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0
1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1
```

**Figure 6.** First 5 samples of A1FLAP.BIN.

### Assembling the .dx Files

Having the four text files created as mentioned earlier (the positions and triangle connections of the palate and the positions sorted by channels and data value of the electrodes), the stage is set to use Data Explorer to display images. Data Explorer imports objects in a specific format described in files with the extention .dx. At least two .dx files are required as shown in Table 1 below. The first one describes the palate and will be named "ak.pal.dx". In that file, "palate_pos" is an array variable of 147 elements. The elements are the positions of the 147 points consisting of 3 items each, the x-y-z values from the text file "ak.palpos". Similarly, "palate_con" is another array of 254 elements, defining each of the 254 triangles from the text file with the connections called "ak.palcon". These two arrays are put together in a field variable called "palate".

**Table 1**. Data Explorer input files.

```
#Data Explorer file: ak.pal.dx
#
object palate_pos array type float
  rank 1 shape 3 items 147 data ak.palpos, 0
object palate_con array type int
  rank 1 shape 3 items 254 data ak.palcon, 0
attribute element type string triangles
attribute ref string positions
object palate field
  component positions palate_pos
  component connections palate_con
end
```

```
#Data Explorer file: ak.elec.dx
#
object electrodes_pos array type float
  rank 1 shape 3 items 96 data ak.elecpos, 0
object electrodes_dat0 array type float
  rank 0 items 96 data A1FLAP.BIN, 0
object electrodes_dat1 array type float
  rank 0 items 96 data A1FLAP.BIN, 196
object electrodes_dat2 array type float
  rank 0 items 96 data A1FLAP.BIN, 393
object electrodes0 field
  component positions electrodes_pos
  component data electrodes_dat0
object electrodes1 field
  component positions electrodes_pos
  component data electrodes_dat1
object electrodes2 field
  component positions electrodes_pos
  component data electrodes_dat2
end
```

The second .dx file, named "ak.elec.dx", is much longer. It requires that each frame of data values be imported into an array. Then, a separate field must also be created for each frame of data values. The array variable "electrodes_pos", like "palate_pos", has 96 elements specifying the positions of the 96 electrodes from "ak.elecpos" (the electrode positions file sorted by channels). Next, "electrodes_dat*n*" where the integer suffix *n* is the frame number, stores the 0 or 1 data values for each channel for the given frame. Remember, "A1FLAP.BIN" is the output file from ELECDAT.EXE with "A1FLAP.NSP" as the input file. The integer following the file name is the starting byte number where the frame begins. A frame is one 10 msec sample representing the state of the 96 electrodes. Here, frame 0 starts at byte 0 in the text file, frame 1 at byte 196 then each successive frame after that starts at 197 byte increments. Only three frames are defined above for illustration but in reality, "ak.elec.dx" has 240 frames, one for each sample. Since typing each line is a tediously repetitive process, the C program DX.C generates a .dx file given the file name with electrode positions, the file name with the electrodes' data values, each offset, and total number of frames. A listing of this program is provided in Appendix C.

Certain scripts and macros given in the next sections expect variables named "palate" and "electrodes" with associated positions, connections, or data values. These will be the actual objects imported and displayed in Data Explorer.

### Creating the Images

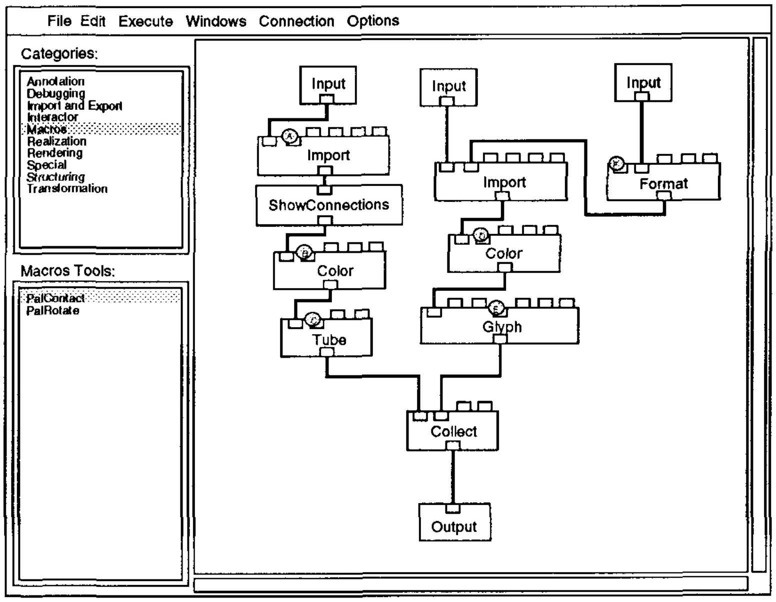
The Data Explorer runs on scripts with the extention .net which are created in its graphical interface as this is easier to use than explicitly writing code. Various block units called modules are placed on the large blank canvas. The Data Explorer automatically converts the arrangement of modules into a .net script, representing actual code. A brief overlook is provided here specifically for the task of viewing time-varying palatal-lingual contact information. Some of the task performed, and the Data Explorer modules that are used to perform them, are listed below in Table 2. The categories are in bold followed by a list of modules they contain. For example, the Import module imports data or objects, Export module exports data or objects, and WriteImage module saves a generated image. They all fall under the category "Import and Export".

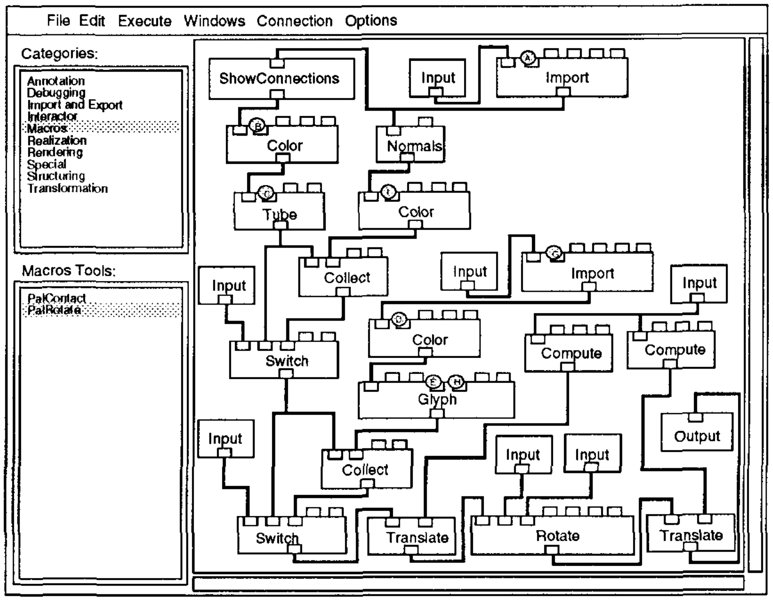
**Table 2.** Data Explorer modules.

**Annotation:** Format, Glyph, Tube
**Debugging:** Print
**Import and Export:** Import, Export, WriteImage
**Realization:** ShowConnections
**Rendering:** AmbientLight, AutoCamera, Camera, Display, FaceNormals, Image, Light, Normal, Render, Rotate, Translate
**Special:** Sequencer, Colormap
**Structuring:** Collect
**Transformation:** Color, Compute

Modules are little blackboxes that perform a set of instructions given some inputs and that usually produce an output, much like a function or procedure in a programming language. Depending on the individual module, there will be several tabs on the top side representing inputs and at most one tab on the bottom representing the output. If an input tab is folded down, it means that the user has specified a value; otherwise, a default value will be assumed.

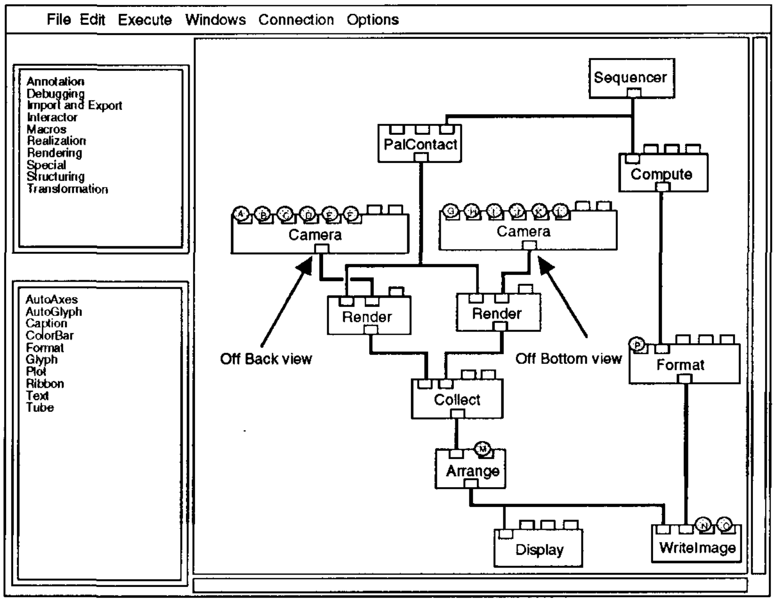
Since importing the palate and a frame of electrode contact will be performed repeatedly, a macro titled PalContact in the category "Macros" shown, in Figure 7 was written to execute the task. A macro is basically a user-defined module. There are two branches above the Collect modules. The left branch (Import through Tube) imports the palate and constructs the triangle connections with a specified color and tube-size. The right branch (Format through Glyph) imports and constructs the electrodes with a specified color as glyphs which are little tiny spheres. When data is displayed, the spheres appear only for those electrodes which have a contact data value of 1. Input and Output modules are used in place of some parameters to allow more general, user-specified usage. PalContact takes 3 inputs: .dx file name where the palate information is stored, .dx file name where the electrodes information is stored, and the frame number *n* to import. Remember that in ak.elec.dx, the variable "electrodes*n*" has a positions component and also a data value component with contact information for a given frame *n*. The variables "palate" and "electrodes*n*" are expected to exist within

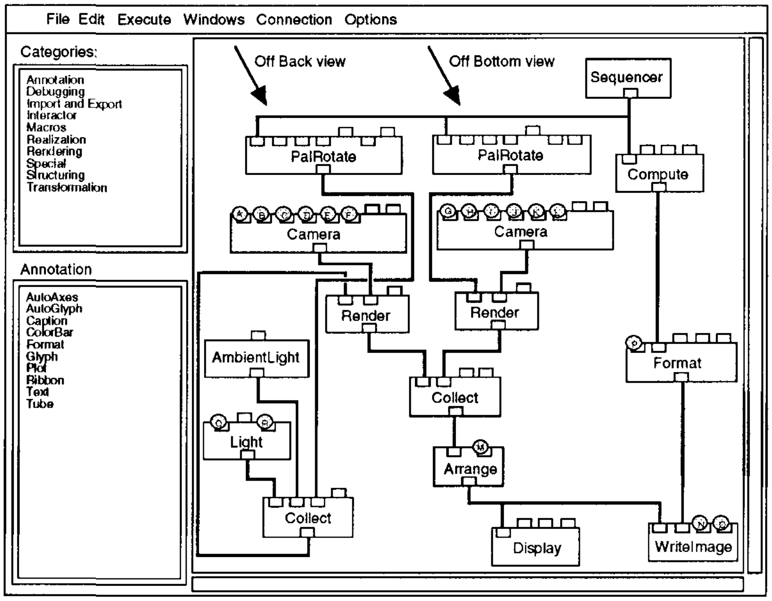




| | |
|---|---|
| Ⓐ "palate" | the variable "palate" is expected in the palate's .dx file |
| Ⓑ "red" | the tubes of the palate are red |
| Ⓒ 0.75 | the width of the tubes are size 0.75 |
| Ⓓ "white" | the electrodes are white |
| Ⓔ 2.25 | the spheres representing the electrodes are size 2.25 |
| Ⓕ "electrodes%d" | a variable "electrodes*n*" is expected for each frame *n* of contact in the electrodes' .dx file |
| Ⓖ "electrodes0" | a variable "electrodes0" is expected in the electrodes' .dx file as dummy variable |
| Ⓗ 1.0 | ratio of on and off electrodes is 1.0 so all electrodes are shown |
| Ⓘ "red" | the covered surface of the palate is red |

**Figure 7.** The networks for macros PalContact (top) and PalRotate (bottom) with descriptions.





Ⓐ [60 50 21]
Ⓑ [135 126 -257]
Ⓒ 130.0000
Ⓓ 320
Ⓔ 1.0
Ⓕ [0 -1 0]

Ⓖ [60 90 21]
Ⓗ [135 -190 96]
Ⓘ 130.0000
Ⓙ 320
Ⓚ 1.0
Ⓛ [0 1 1]

Ⓜ 2
Ⓝ local
Ⓞ tiff
Ⓟ *filename*
Ⓠ [5 -10 5]
Ⓡ 1

**Figure 8.** The networks "utterance.net" (top) and "rotation.net" (bottom) with assigned values.

those files. The module's output is a collective object containing the electrode contact at frame *n* superimposed onto the palate.

Another macro, the PalRotate module, performs a rotation of the palate. This is also shown in Figure 7. Rotation is accomplished fairly easily with a Rotate module sandwiched between two Translate modules, resulting in an object which can be rendered. The first Translate module is necessary to align the axes so they cross at the center of the object. This way, the object will be rotating around an internal axis and not flying through space around an external axis. The Rotation module performs the actual calculation to rotate, given the axis and degree of rotation. The final Translate module shifts the object back to its original position.

PalRotate also gives an option to cover the surface of the wire-frame and to show the electrodes. The Switch module takes a set of objects and selects one to pass through. The macro implements Normals to cover the surface with the default light pointing in the position y direction. When rotating, the electrodes are either all present or all absent. In summary, PalRotate takes 7 inputs: the degree of rotation, the .dx file name with palate defined, the .dx file name with the electrodes defined, whether the surface should be covered, whether the electrodes should be shown, the axis to be rotated about, and the coordinates of the center. The Sequencer, a module that generates integers sequentially, is usually connected to specify the first input. Running it from 0 to 360 at 10 degree increments provides smooth movement without taking too much time. A more detailed documentation of both PalContact and PalRotate is in Appendices D and E.

Sample networks illustrate how to use new modules from the macros. "Utterance.net" displays a sequence of frames in an utterance, and "rotation.net" rotates the palate; both are shown in Figure 8. The scripts may be easily incorporated to generate other frames simply by specifying the file names. The output is a sequence of TIFF files showing the palate in two views. The palate is shown from the bottom off to the right, and also from the anterior. After all the frames had been created they were imported onto the Macintosh.

### Making the Movie

The movie uses a split screen with the two views generated by "utterance.net" and "rotation.net". It begins with a few rotations to orient the viewers. The first rotation is just the covered surface; this is most helpful for the view from the bottom which can easily create an illusion of how the palate is placed and turned. In the next rotation, all 96 electrodes appear in their proper positions on the wire-frame. Then the tongue contact pattern of a given utterance is presented four times at 3 different speeds: in real time, one-eighth of real time, one-half of real time, and real time again.

Clips were initially created in the QuickTime MovieConverter. The number of frames per second depends on the clip. Rotation at 10 degree increments ran at 5 frames per second; frames started getting lost at any higher speeds. Utterance displays have little change so higher speeds were fine. 100 frames per second is real time and seems incredibly fast. 50 frames per second which is one-half real time, and 12.5 frames per second which is one-eight real time, were appropriate speeds for contrast and careful study. In all cases, a control panel is available to step through frame-by-frame.

After this initial clip was made, MacroMind Director was used to place a caption at the bottom and to add audio to match the articulation, creating a more complete clip. The audio samples were prepared using SoundEdit Pro. The "Tempo" option was used to slow down the speech. The same

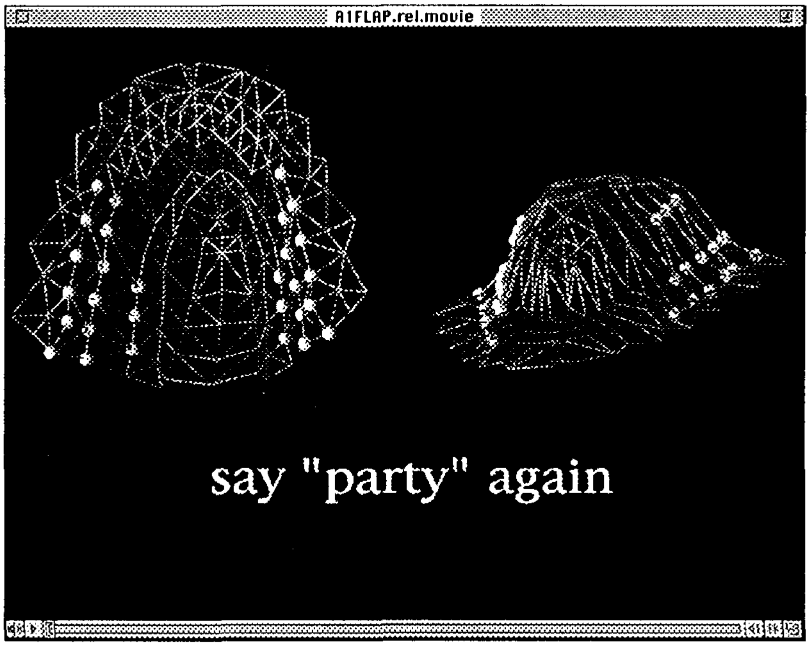


**Figure 9.** First frame of A1FLAP.rel.clip.

steps were repeated for a total of 10 tokens in the flaps and taps experiment. To put the whole movie together the resulting clips were imported into Adobe Premiere and placed into a new sequence. A sample of the setup is shown in Figure 9. The complete movie was then be recorded onto a VCR.

## Study of Flaps and Taps

As dynamic articulations, flaps and taps are ideal for viewing on a electropalatography display as opposed to a static palatogram. In the 1993 edition of A Course in Phonetics, Ladefoged explains, "A tap or a flap is caused by a single contraction of muscles so that one articulator is thrown against another. It is often just a very rapid articulation of a stop." He then further distinguishes between taps and flaps.

> In a tap [ɾ], the tip of the tongue simply moves up to contact the roof of the mouth in the dental or alveolar region, and then moves back to the floor of the mouth along the same path. In a flap [ɽ], the tip of the tongue is first curled up and back in a retroflex gesture, and then strikes the roof of the mouth in the post-alveolar region as it returns to its position behind the lower front teeth. The distinction between taps and flaps is thus to some extent bound up with what might be called a distinction in place of articulation. Flaps are retroflex articulations. (168)

What we called a tap [ɾ] is sometimes labeled "alveolar flap" and a flap [ɽ] is sometimes labeled "retroflex flap". The hypothesis of this study is that flaps and taps both occur as post-stress allophones of /t/ in American English; if there is an adjacent rhotic the articulation will be a flap, but otherwise a tap. We will address the following questions:

(1) Do both flaps and taps occur in American English?
(2) How are flaps and taps different from stops?
(3) Are there any differences between oral taps and nasal taps?

Data were collected consisting of a total of ten sentences repeated five times each by one speaker. The sentences consisted of the carrier phrase "Say ___ again" and contained a word with an anticipated flap, oral tap, nasal tap, oral alveolar stop [t], or prenasalized alveolar stop [t]. All target consonants of interest were followed by [i]. Each set of consonants was also tested after both low and non-low vowels. This enabled us to identify and rule out aspects that were effects of the adjacent vowel rather than characteristics of the articulation itself. The list of target words recorded is given in Table 3.

**Table 3.** Target words recorded in the experiment.

| TARGET WORDS | TARGET CONSONANTS |
|---|---|
| Bertie | flap after [ɚ] |
| petty | oral tap after [ɛ] |
| plenty | nasal tap after [ɛ] |
| repetition | oral [t] after [i] |
| intonation | prenasalized [t] after [ɨ] |
| | |
| party | flap after [ɑɹ] |
| potty | oral tap after [ɑ] |
| Pontiac | nasal tap after [ɑ] |
| autistic | oral [t] after [ɑ] |
| auntie | prenasalized [t] after [æ] |

Initially, data was compiled by recording the total frames of central contact in each consonant. Central contact was defined as contact in any electrodes in the first five rows of the three middle columns. Notice that this definition did not guarantee complete closure. For each frame, the frontmost and backmost electrode row numbers of central contact were also recorded. Each frame is a 10 msec sample and Row 1 starts at the teeth. Then to help answer the questions above, the degree of movement, the place of articulation, the duration, and the width of constriction were also calculated. Thus the following observations and conclusions were made:

(1) Both flaps and taps occur in this speaker's American English, differing in both movement and place of articulation.
(2) Stops had longer duration and generally a wider constriction than both flaps and taps.
(3) Oral and nasal taps differed slightly in duration, width of constriction, and perhaps direction of release.

In answering question (1), whether or not flaps and taps both occur was tested based on Ladefoged's definition. First of all, flaps and taps should differ in place of articulation; flaps contact the post-alveolar region while taps contact the alveolar ridge. Place of articulation is a function of electrode row numbers measured at the steady-state or at the middle of each articulation. If more than one row was contacted, the mean of the row numbers was used. To maintain a meaningful relation between the row numbers and place of articulation, the fourth electrode on the middle column was considered as Row 3 rather than Row 4. To comprehend what the row numbers imply, take advantage of the 3-D display and notice that Row 3 is at the edge of the alveolar ridge. Observe that the fourth

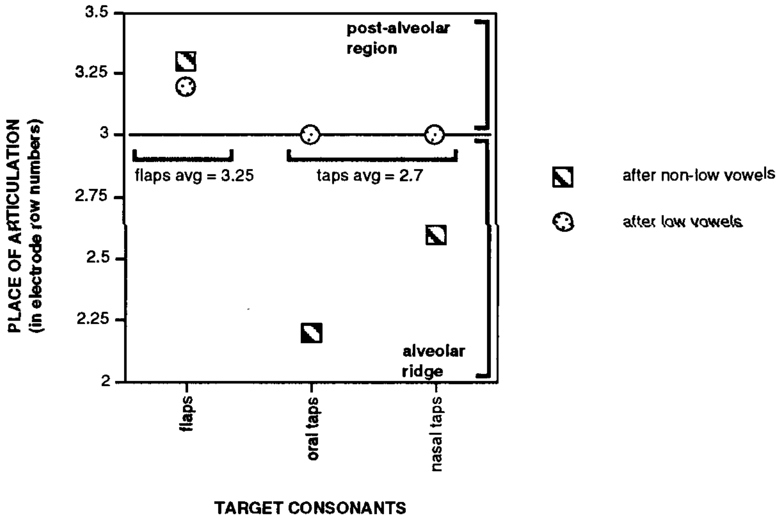


**Figure 10.** Contrasting place of articulation between flaps and taps. The line drawn at Row 3 divides the alveolar and post-alveolar regions.

electrode on the middle column mentioned earlier is also still on the edge of the alveolar ridge while the other electrodes on Row 4 are clearly higher up on the post-alveolar region. Thus, anything with values slightly greater than Row 3 was considered to be on the post-alveolar region and Row 3 or less to be on the alveolar ridge. Figure 10 illustrates the average electrode row number at the middle of each articulation for the various flaps and taps broken down by vowel enironment and nasality. Flaps generally had values greater than Row 3, indicating that they were on the post-alveolar region. In fact, flaps reached Row 4 in 80% of the tokens and in 0% of all other target consonants. As predicted, taps were always at Row 3 or more forward, meaning that they were on the alveolar ridge. Speaker AK has an unusually steep palate so the distance from Row 3 to Row 4 (about 10 mm) is a considerable larger than from say Row 1 to Row 2 (about 2 mm).

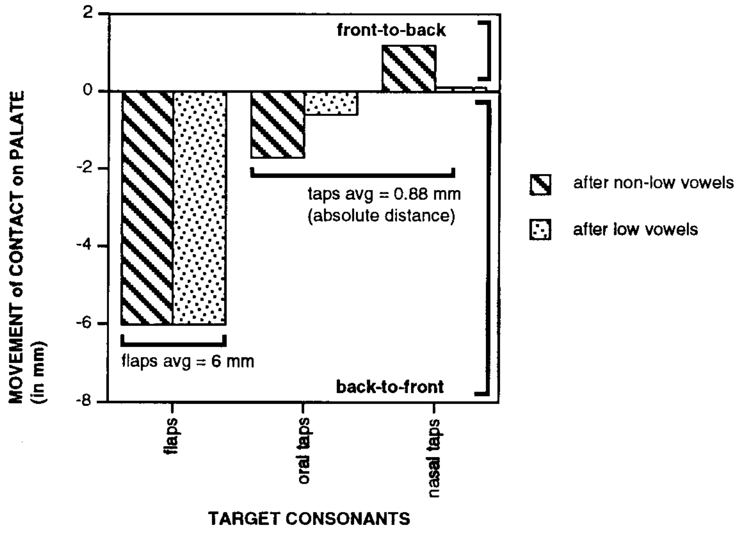


**Figure 11.** Contrasting distance and direction of movement between flaps and taps.

Next, a tap should have very little movement of the contact on the palate because it is an up-and-down motion, whereas flaps would have a back-to-front movement of the contact on the palate. Degree of movement was measured as a change in row numbers contacted. It is calculated by first finding the difference in row numbers between the initial contact and final release. This absolute row difference is then converted roughly into distance in terms of mm. For example, the distance from Row 3 to Row 4 is about 10 mm , Row 2 to Row 3 is 3 mm, and Row 1 to Row 2 is 2 mm. These may not necessarily be the actual distance covered by the tongue but they give accurate weights. Figure 11 shows the average movement of flaps and taps. Positive values are front-to-back movements while negative values are back-to-front. As a whole, taps did not indicate a clear direction for their movement. Flaps generally moved from Row 4 to Row 3 with an overall average of 6 mm. They began on the post-alveolar region, migrated forward, and seemed to "drop off" at the end of the alveolar ridge in Row 3. On the other hand, if taps showed any movement at all, it was between Row 1, Row 2, or Row 3 with an average less than 1 mm. The speaker has a retroflex [ɹ], that is, the tongue tip is curled back for the /r/. It would be interesting to see whether speakers with bunched [ɹ] would have a flap here.

Addressing question (2) to compare stops with flaps and taps, we measured the duration and the width of the constriction. Figure 12 shows the duration of contact for each articulation. The duration is the number of frames of central contact, each lasting 10 msec. The oral taps had the shortest duration, followed by flaps and nasal taps, then [t]. As the definition states, flaps and taps are "a very rapid articulation of a stop"; in these cases, they are often only half the duration of a stop [t] counterpart. Width of constriction is another characteristic distinguishing stops from flaps and taps. This was calculated as the average width throughout the entire constriction in terms of mm. The conversion from row numbers to distance in mm was the same as the one used to determine movement. Stops tend to have thicker or wider constrictions typically covering 2 or 3 rows of electrodes while flaps and taps tend to have thinner or narrower constrictions which appeared as only one row of electrodes. Figure 13 illustrates these calculations. On the average, stops had wider constrictions. However, flaps were also fairly wide due to the large area that must be covered when

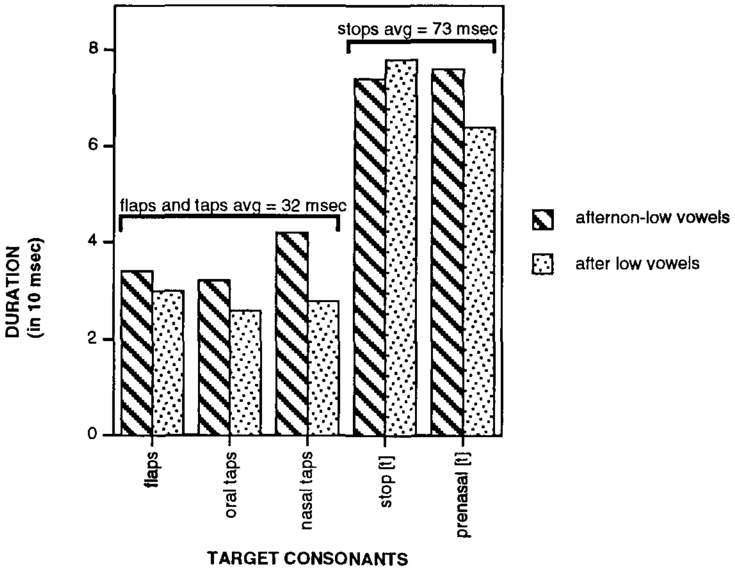


**Figure 12.** Comparing duration of stops with duration of flaps and taps.

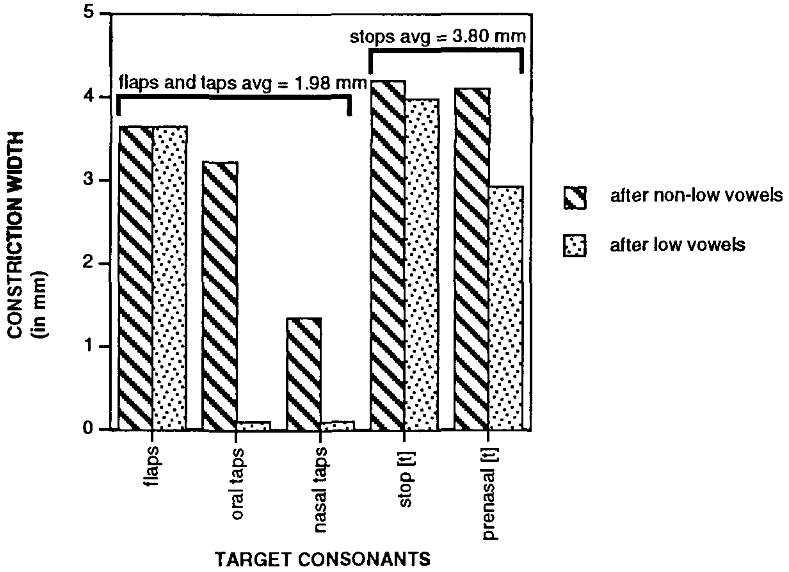


**Figure 13.** Comparing constriction width of stops with constriction width of flaps and taps.

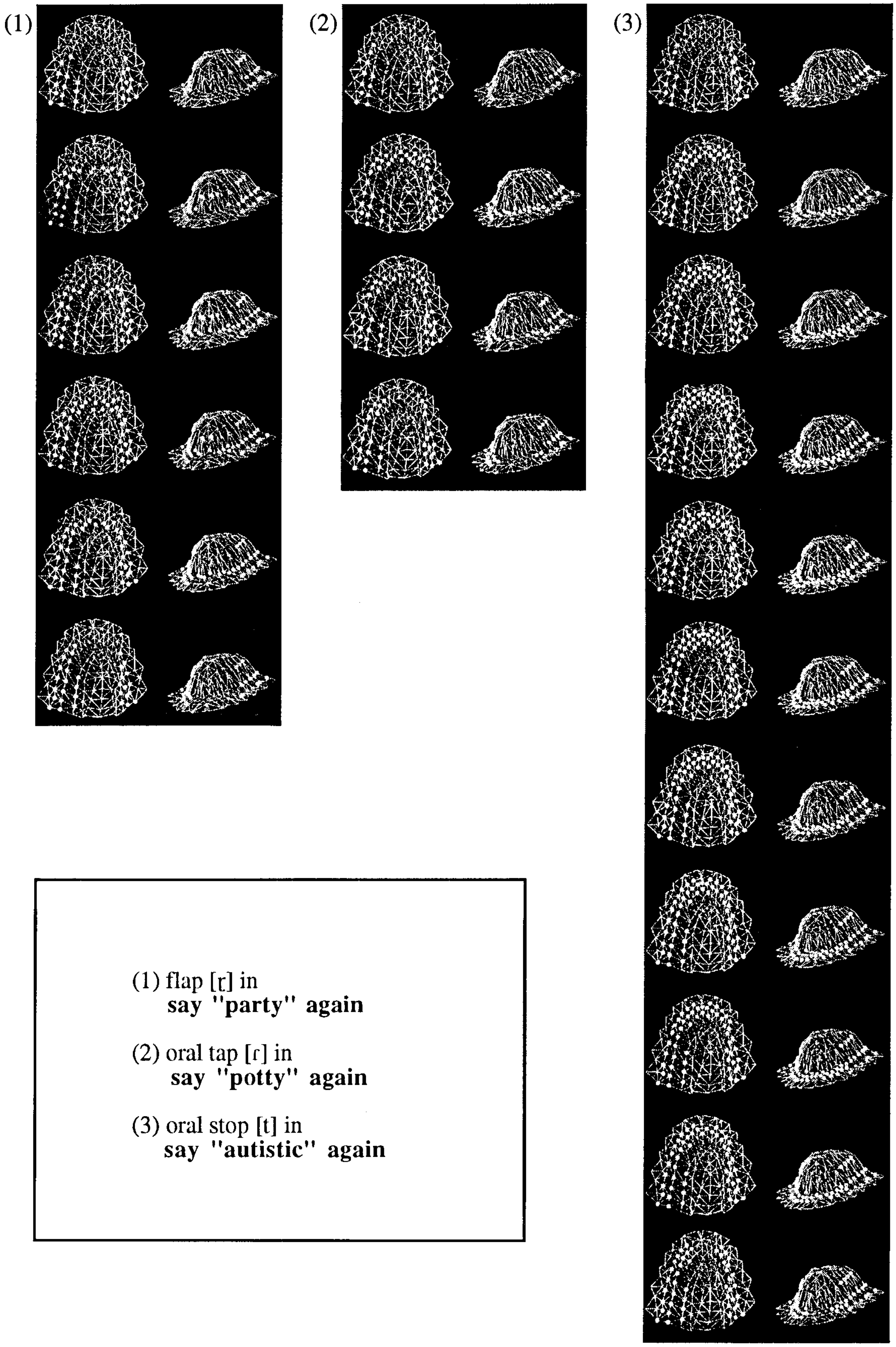
moving from Row 4 to Row 3. But this wide constriction usually only lasted about one or two frames. In any case, they were still not a wide as the widest stops although a little wider than a prenasalized [t] after a low vowel.

In response to question (3), oral and nasal taps are fairly similar in terms of place of articulation and amount of movement but have slight differences in other aspects. Table 4 compares the place of articulation, movement, duration, and constriction width. Taps, classified as an alveolar articulation, clearly remained on the alveolar ridge at all times although they were often very near the edge at Row 3. As shown in Figure 11, neither oral nor nasal taps show much movement, especially after low vowels. But after non-low vowels nasal taps initially contact Row 2 and Row 3 then release backward to Row 3. In oral taps, the base is at Row 3 and the tongue tip rolls forward widening the constriction as far as Row 2 or Row 1. Besides the opposite directions of release, oral taps are slightly shorter in duration and have somewhat wider constrictions than nasal taps. These differences remain to be explained.

Figure 14 illustrates the constriction of a flap, a tap, and a stop [t]. The duration and width of constriction differences distinguishing stops from flaps and taps are quite clear. But it is also a good indication of differences between the place of articulation and amount of movement in flaps and taps.

**Table 4.** Differences between oral taps and nasal taps.

| | ORAL TAPS | NASAL TAPS | CONCLUSION |
|---|---|---|---|
| place of articulation | 2.6 | 2.8 | very close |
| movement of contact on palate | -0.85 mm | 0.6 mm | very close absolute distance but in opposite directions |
| duration | 29 msec | 35 msec | oral taps slightly shorter |
| constriction width | 23 mm | 12 mm | oral taps slightly wider |



(1) flap [ɽ] in
**say "party" again**

(2) oral tap [ɾ] in
**say "potty" again**

(3) oral stop [t] in
**say "autistic" again**

**Figure 14.** Illustration of a flap, an oral tap, and an oral stop [t].

## Conclusion

A systematic method for collecting and converting electropalatography data for 3-D has been presented. The contact of the tongue with the palate, registered by the electrodes, is superimposed on a simplified shape of the palate. Such a display is especially enlightening compared to the commonly used 2-D display for visualizing speaker variations and tongue movement. The process involves manually collecting x-y-z values defining the points on the palate and translating the EPG binary file. The palate may be rotated to various views and the contacted electrodes may be displayed, providing a rough sense of what is truly happening inside the oral cavity.

Flaps and taps before [i] processed by this method clearly displayed certain features described in textbooks, differing in both movement and place of articulation. Both flaps and taps were shorter in duration than stops which generally had a wider constriction than both flaps and taps. Oral and nasal taps differed slightly in duration, width of constriction, and perhaps direction of release.

## Summary of Procedures

| | | |
|---|---|---|
| 1 | Collect x-y-z positions of palate | ak.palpos |
| 2 | Triangulate with connections | ak.palcon |
| 3 | Sort electrodes x-y-z positions by channels | ak.elecpos |
| 4 | Convert .NSP file to text by running DUMPBIN.EXE | |
| | A1FLAP.BIN, A1OTAP.BIN, A1NTAP.BIN, A1OTEE.BIN, A1NTEE.BIN | |
| | E1FLAP.BIN, E1OTAP.BIN, E1NTAP.BIN, E1OTEE.BIN, E1NTEE.BIN | |
| 5 | Create .dx files to define palate | ak.pal.dx |
| 6 | Create .dx files to define electrodes by running DX.EXE | ak.elec.dx |
| 7 | Create and load macros PalContact.net & PalRotate.net | |
| 8 | Create network and generate frames to view utterances | utterance.net |
| | ak.A1FLAP, ak.A1OTAP, ak.A1NTAP, ak.A1OTEE, ak.A1NTEE | |
| | ak.E1FLAP, ak.E1OTAP, ak.E1NTAP, ak.E1OTEE, ak.E1NTEE | |
| 9 | Create network and generate frames to view rotations. | rotation.net |
| | covered surface. | ak.rotation21 |
| | wire-frame. | ak.rotation12 |
| 10 | Make clips of utterances at 12.5 frames per second, one-eighth real time. | |
| | Make clips of utterances at 50 frames per second, one-half real time. | |
| | Make clips of utterances at 100 frames per second, real time . | |
| | Make clips of rotation. | |
| 11 | Add caption and audio to each clip. | |
| 12 | Place sequences in order and record onto VCR. | |

## Aknowledgement

This work was supported by an REU supplement to NSF grant #DBS 9213604 to Patricia Keating. Special thanks to Patricia Keating for coming up with the research idea, obtaining the funding, and reading over all the drafts; without her, this project would not have been possible. Credit is also due to Dani Byrd for her help with using the palatometer and providing the original test data, Joan Slottow from the OAC Visualization Laboratory for familiarizing me with the visualization equipment, and Abigail Kaun for her involvement as a research subject. I am grateful to Ian Maddieson for his insightful comments and members of the Phonetics Laboratory for their input.

## References

Chui, W.S.C., and Shadle, C.H. (1992) Use of palate shape data in an enhanced electropalatography system. *Proceedings of the Institute of Acoustics, Speech and Hearing Conference*, Windermere, November 19-22, 1992.

Ladefoged, P. (1993) *A Course in Phonetics, 3rd Edition*, Harcourt Brace Jovanovich, New York.

## Appendix A

"ak.palpos" defining the positions of the palate (page 15 of 147)

| point number | x value | y value | z value |
|---|---|---|---|
| 0 | 60 | 30 | 19 |
| 1 | 69 | 9 | 6 |
| 2 | 60 | 8 | 6 |
| 3 | 51 | 9 | 6 |
| 4 | 76 | 13 | 6 |
| 5 | 65 | 13 | 8 |
| 6 | 60 | 12 | 8 |
| 7 | 55 | 13 | 8 |
| 8 | 44 | 13 | 6 |
| 9 | 71 | 17 | 8 |
| 10 | 49 | 17 | 8 |
| 11 | 82 | 18 | 6 |

"ak.palcon" defining the triangle connections of the palate (page 15 of 254)

| triangle number | 1st point | 2nd point | 3rd point |
|---|---|---|---|
| 1 | 1 | 2 | 5 |
| 2 | 2 | 3 | 7 |
| 3 | 5 | 12 | 15 |
| 4 | 7 | 12 | 16 |
| 5 | 12 | 15 | 20 |
| 6 | 12 | 16 | 20 |
| 7 | 2 | 5 | 6 |
| 8 | 2 | 6 | 7 |
| 9 | 5 | 6 | 12 |
| 10 | 6 | 7 | 12 |
| 11 | 5 | 9 | 15 |
| 12 | 7 | 10 | 16 |
| 13 | 1 | 5 | 9 |
| 14 | 3 | 7 | 10 |
| 15 | 1 | 4 | 9 |

"ak.elecpos" defining the positions of the electrodes sorted by channel (15 lines of 96)

| *channel* | point number | *x value* | *y value* | *z value* |
|---|---|---|---|---|
| 1 | 56 | 25 | 51 | 2 |
| 2 | 85 | 24 | 86 | 9 |
| 3 | 19 | 70 | 26 | 9 |
| 4 | 93 | 60 | 102 | 30 |
| 5 | 62 | 43 | 60 | 32 |
| 6 | 35 | 65 | 36 | 29 |
| 7 | 23 | 81 | 26 | 9 |
| 8 | 11 | 82 | 18 | 6 |
| 9 | 52 | 74 | 51 | 32 |
| 10 | 16 | 54 | 22 | 9 |
| 11 | 81 | 60 | 92 | 36 |
| 12 | 54 | 95 | 51 | 2 |
| 13 | 22 | 34 | 24 | 8 |
| 14 | 82 | 101 | 85 | 1 |
| 15 | 76 | 26 | 76 | 9 |

## **Appendix B** ELECDAT.C

```
/*****************************************************************************
 *
 *   ELECDAT.C
 *
 *   Takes a *.NSP file and prints the inary Palatometer data
 *   in a *.BIN text file
 *
 *   by:      Cheng Cheng Saw 4/93 revised 8/93
 *
 *   for:     UCLA Phonetics Laboratory
 *            Department of Linguistics
 *
 *****************************************************************************/

#include <conio.h>
#include <stdio.h>
#include <string.h>
#include <stdlib.h>

#define TRUE                    1
#define FALSE                   0
#define UP_TO_AUDIO_DATA        112
#define UP_TO_PAL_DATA          520
#define MAX_ELECTRODES          96
#define BYTES_PER_SAMPLE        12      /* 12 bytes per sample = 96 bits */

void main (argc, argv)
    int     argc;
    char    *argv[];
{
    FILE    *ip, *op;                   /* infile and outfile pointers */
    char    *ptr;                       /* temporary string pointer */
    char    string[5],infile[20],outfile[20];
    long    audio_size,samples;
    int     X,Y,i,j,k;

    puts ("\nELECDAT -- prints the electrode contact data.\n");

/* taking file input parameters */
    if (argc != 2) {
        printf ("Enter prefix of the file (.NSP extension presumed): ");
        scanf  ("%s",infile);
    }
    else
        strcpy (infile,argv[1]);

/* opening *.NSP input file */
    strcpy(outfile,infile);
    strcat(infile,".NSP");
    printf("The input file is %s\n",infile);
    if ( (ip = fopen(infile, "rb")) == NULL) {
        printf ("Can't open %s\n",infile);
        exit(1);
    }

/* opening *.BIN output file */
    strcat(outfile,".BIN");
```

```
    printf("The output file is %s\n",outfile);
    if ( (op = fopen(outfile, "w")) == NULL) {
        printf ("Can't open %s\n",outfile);
        exit(1);
    }

/**********
 * The following criteria must be met in order for
 * the file to be considered a valid palatometer file:
 *    Contains audio data length in bytes        112-115
 *    Contains string "PALA" in bytes 116 + data length - same + 4 OR
 *    Contains string "PREF" in same bytes as above
 **********/

    (void) fseek (ip, UP_TO_AUDIO_DATA, SEEK_SET);
    if (fread (&audio_size, sizeof (long), 1, ip) != 1) {
        printf ("Error reading audio size from %s\n", infile);
    }
/**********
 * Here we look for the string "PALA".  If we find it, the
 * pal data starts next.  If the string is "PREF", the pal
 * data starts 520 plus audio data size bytes into the file.
 **********/
    (void) fseek (ip, audio_size + UP_TO_AUDIO_DATA + 4, SEEK_SET);
    ptr = fgets (string, 5, ip);
    if (strcmp (ptr, "PALA")) {
        if (strcmp (ptr, "PREF")) {
            printf
            ("%s is not a palatomer file.  Missing PALA/PREF header\n",
            infile);
            exit(1);
        }
        (void) fseek (ip, audio_size + UP_TO_PAL_DATA, SEEK_SET);
    }

/* getting the number of samples and prints for future use */
    fread (&samples, sizeof (long), 1, ip);
    samples /= BYTES_PER_SAMPLE;
    printf("There are %lu SAMPLES.\n",samples);

/* ip now points at beginning of palatometer data */
    printf("Printing the electrode contact data ...\n");
    for (i=1;i<=samples;i++) {
        for (j=1;j<=6;j++) {
            fread(&X, sizeof (int),1,ip);
            for (k=0;k<=15;k++) {
                Y=X;
                fprintf(op,"%d ",(Y>>k)&1);
            }
            if ((j%2)==0) fprintf(op,"\n");
        }
        fprintf(op,"\n");
    }

    puts( DONE. );
    fcloseall();
}
```

# Appendix C DX.C

```
/*****************************************************************************
 *
 *  DX.C
 *
 *  This is a general program to create a Data Explorer
 *  format file given the electrode positions file, the
 *  electrode contact data file, the total number of frames,
 *  and the output file name.
 *
 *  by:     Cheng Cheng Saw 8/93
 *
 *  for:    UCLA Phonetics Laboratory
 *          Department of Linguistics
 *
 *****************************************************************************/

#include <stdio.h>
#include <string.h>
#include <stdlib.h>
/* the offset between each frame in the electrodes data file is 197 */
#define FRAME_OFFSET 197

void main(argc, argv)
    int     argc;
    char    *argv[];
{
    FILE    *fp;
    unsigned long offset,frame;
    char    elecpos[20],elecdat[20],filename[20],string[5],quote[2];
    int     max_frame;

    puts("\nDX.C -- creates a Data Explorer format file for electrodes\n");

/* taking file input parameters */
    if (argc != 4) {
        printf ("Enter file name with the electrode positions: ");
        scanf  ("%s",elecpos);

        printf ("Enter file name with the electrode data: ");
        scanf  ("%s",elecdat);

        printf ("Enter the total number of frames in %s: ",elecdat);
        scanf  ("%d",&max_frame);
    }
    else
    {
        strcpy(elecpos,argv[1]);
        strcpy(elecdat,argv[2]);
        sscanf(argv[3],"%d",&max_frame);
    }

    printf ("Enter file name where the output should be stored: ");
    scanf  ("%s",filename);
    printf("Creating Data Explorer file %s ...\n",filename);
    if ( (fp = fopen(filename, "w")) == NULL) {
        printf ("Can't open %s\n",filename);
        exit(1);
```

```
    }

/* fixing problems with misplaced quotation marks */
    sprintf(quote,"%c",'"');

/* printing file header */
    fprintf(fp,"# Data Explorer file: %s\n#\n",filename);

/* creating electrode positions object "electrodes_pos" */
    fprintf(fp,"object %selectrodes_pos%s array type float\n",quote,quote);
    fprintf(fp,"    rank 1 shape 3 items 96 data %s,0\n\n",elecpos);

/* creating electrode data object "electrodes_datn" for each frame n */
    for (frame=0;frame<max_frame;frame++)
    {
        sprintf(string,"%d",frame);
        strcat(string,quote);
        fprintf(fp,"object %selectrodes_dat%s array type float\n",quote,string);

        if (!frame)
            offset = 0;
        else
            offset = FRAME_OFFSET*frame-1;

        fprintf(fp,"    rank 0 items 96 data %s,%lu\n",elecdat,offset);
    }
/* carriage return before starting next object */
    fprintf(fp,"\n");

/* creating electrode object "electrodesn" for each frame n */
    for (frame=0;frame<max_frame;frame++)
    {
        sprintf(string,"%d",frame);
        strcat(string,quote);
        fprintf(fp,"object %selectrodes%s field\n",quote,string);
        fprintf(fp,"    component %spositions%s %selectrodes_pos%s\n",
            quote,quote,quote,quote);
        fprintf(fp,"    component %sdata%s %selectrodes_dat%s\n",
            quote,quote,quote,string);
    }
    fprintf(fp,"end\n");

    puts("DONE.");
     fcloseall();
}
```

**Name** PalContact - displays palate contact information

**Syntax** output = PalContact(palate,electrodes,frame)

**Inputs**

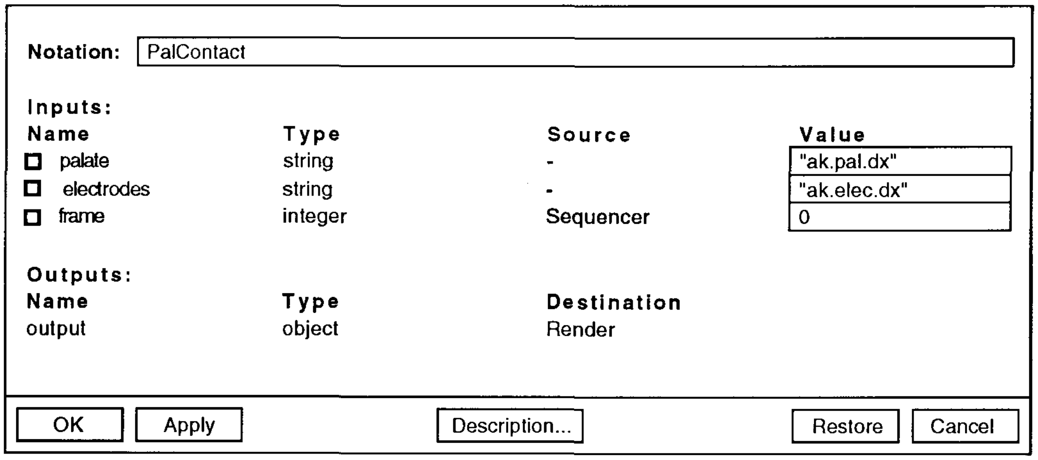
| NAME | TYPE | DEFAULT | DESCRIPTION |
|---|---|---|---|
| palate | string | no default | name of file with palate information |
| electrodes | string | no default | name of file with electrode information |
| frame | integer | 0 | frame number to import |

**Outputs**

| NAME | TYPE | DESCRIPTION |
|---|---|---|
| output | object | palate with electrode contact superimposed |

**Description** The input “palate” takes a string file name which has the definitions for the palate; thus, a variable called “palate” which has a positions and connections component is expected to exist. Likewise, the input “electrodes” takes a string file name which has the definitions for the electrodes. This file contains one set of positions for the 96 electrodes. Then for each 10 millisecond sample during the utterance, called a frame, a separate variable must store the data information listing whether each electrode is on or off. Additionally for each frame, a variable must have the electrode positions component as well as the data value component for that frame. The input “frame” specifies the frame number of electrode contact to be imported. Hence, the actual variable imported is “electrodes*frame*”. The result, “output”, is the electrodes which had contact super imposed on the palate. It is an object which can be rendered and displayed.

**Example** These are the values used in utterance.net:



**Notation:** PalContact

**Inputs:**

| Name | Type | Source | Value |
|---|---|---|---|
| ☐ palate | string | - | "ak.pal.dx" |
| ☐ electrodes | string | - | "ak.elec.dx" |
| ☐ frame | integer | Sequencer | 0 |

**Outputs:**

| Name | Type | Destination |
|---|---|---|
| output | object | Render |

**See Also** PalRotate

**Name** PalRotate - displays a rotated palate

**Syntax** output = PalRotate(degree,palate,electrodes,surface,show,axis,center)

**Inputs**

| NAME | TYPE | DEFAULT | DESCRIPTION |
|---|---|---|---|
| degree | integer | 0 | degree of rotation |
| palate | string | no default | name of file with palate information |
| electrodes | string | no default | name of file with electrode information |
| surface | integer, flag | 1 | 1=no covered surface<br>2=covered surface |
| show | integer, flag | 1 | 1=no electrodes on<br>2=all electrodes on |
| axis | string, integer | y | axis to rotate around |
| center | vector | [0, 0, 0] | center of object |

**Outputs**

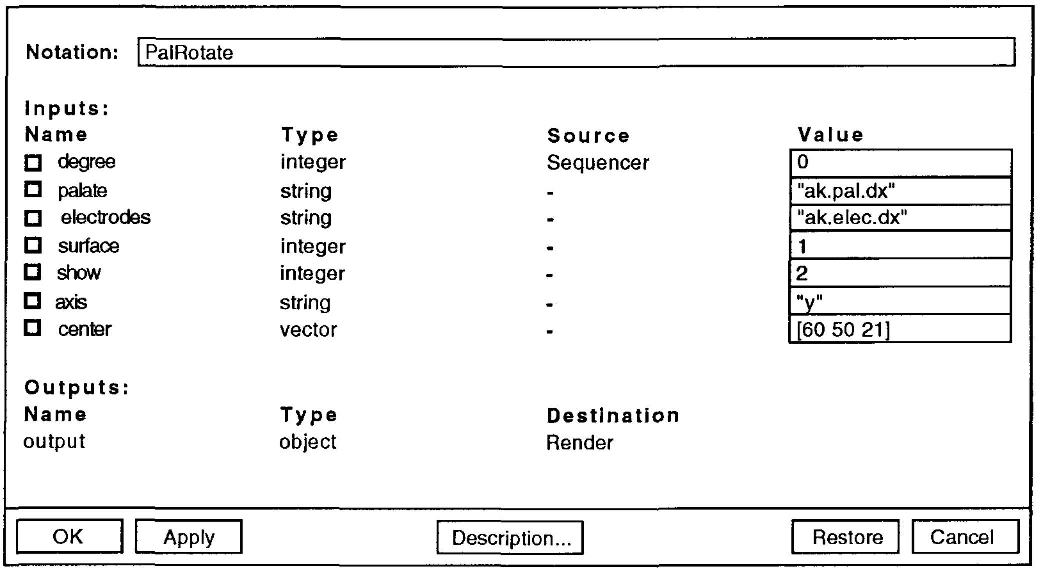
| NAME | TYPE | DESCRIPTION |
|---|---|---|
| output | object | palate with at given rotation |

**Description**

The Sequencer is usually connected to the "degree" parameter since it specifies the degree of rotation, usually between 0 and 360, from how the object is initially defined. The input "palate" takes a string file name which has the definitions for the palate; thus, a variable called "palate" which has a positions and connections component is expected to exist. Likewise, the input "electrodes" takes a string file name which has the definitions for the electrodes. This file contains one set of positions for the 96 electrodes. Then for each 10 millisecond sample during the utterance, called a frame, a separate variable must store the data information listing whether each electrode is on or off. The variable "electrode0" with the positions component and a component having the data values of electrode contact at frame 0 is imported. But only the positions component is used since all the electrodes, whether contacted or not, are displayed with the same size.

Both "surface" and "show" are switches. Surface determines whether the wire-frame palate will have a covered surface which is often helpful to distinguish how the object is being viewed. Lighting effects are also more apparent when there is a surface. Show, however, determines whether or not to show the electrodes; they are either all on or all off. The "axis" is used to specify which axis the object is rotated around. The choices are x, y, and z or 1, 2, and 3. "center" specifies where the center of the palate is. It is used to shift the palate so its [0 0 0] axes are internally at the center. This ensures that the palate will actually be rotating and not flying through space. The result, "output", is an object, shifted to the given rotation, which can be rendered and displayed.

**Example** These are the values used in rotation.net for the Off Back view on the left. It was a wire-frame with all the electrodes on rotated around the y axis. The Off Bottom view is usually rotated around the z axis.



Notation: PalRotate

Inputs:

| Name | Type | Source | Value |
|---|---|---|---|
| ☐ degree | integer | Sequencer | 0 |
| ☐ palate | string | - | "ak.pal.dx" |
| ☐ electrodes | string | - | "ak.elec.dx" |
| ☐ surface | integer | - | 1 |
| ☐ show | integer | - | 2 |
| ☐ axis | string | - | "y" |
| ☐ center | vector | - | [60 50 21] |

Outputs:

| Name | Type | Destination |
|---|---|---|
| output | object | Render |

**See Also** PalContact

# Aerodynamic Evidence for Articulatory Overlap in Korean

Daniel Silverman
Jongho Jun

**Abstract.** Aerodynamic evidence indicates the existence of overlapped labial//velar sequences in Korean. Oral pressure readings for [ipku] show a brief rarefaction in oral pressure during the consonantal sequence, indicating that tongue retraction during a front-back vowel sequence occurs simultaneously with full closures at the labial and velar places of articulation. This confluence of phenomena results in the observed pressure rarefaction due to expansion of the sealed oral cavity. Similarly, pressure readings for [upki] show a brief, marked increase during the consonantal sequence, indicating that tongue advancement during a back-front vowel sequence temporally overlaps with full closures at both the labial and velar places of articulation. This results in a pressure increase due to contraction of the sealed oral cavity. To our knowledge, the present study is the first to demonstrate consonant co-production in terms of oral pressure, and to report on coarticulatory effects involving four sequenced segments.

## 1. Introduction

In this study we present evidence for the existence of overlapped labial//velar sequences in Korean. A native speaker of the Seoul dialect (one of the authors) was recorded in the UCLA Phonetics Laboratory uttering [VCCV] sequences involving both labial and velar consonants, and front and back vowels, in a variety of combinations ([ipki, upku, ipku, upki, ikpu, ukpi, ikpi, ukpu]). Both nonsense words (Experiment One) and real words (Experiment Two) were employed. Oral airflow, as well as pharyngeal and suprapharyngeal pressure, were recorded. It was found that back-front vowel combinations, in conjunction with intervocalic -pk- sequences ([upki]), produced a marked increase in suprapharyngeal (henceforth oral) pressure during the course of the consonantal sequence -- a far greater increase than was found when the same consonantal sequence was flanked by vowels of identical phonemic quality ([ipki, upku]). We claim that these oral pressure readings are the result of tongue advancement during a back-front vowel sequence, temporally overlapping with full closures at both the labial and velar places of articulation. The result is a brief increase in oral pressure due to oral cavity contraction. Relatedly, front-back vowel combinations in the same consonantal environment ([ipku]) produced a marked rarefaction in oral pressure: tongue retraction during a front-back vowel sequence, occurring simultaneously with full closures at the labial and velar places of articulation, results in a decrease in oral pressure due to cavity expansion.

Previous studies investigating gestural timing relationships in terms of air pressure include Ladefoged (1962) and Demolin (1992), which report on phonological labio-velar stops in West and Central African languages, respectively. Ohala (1981) investigates epenthetic stop production in English through similar techniques. Kozhevnikov and Chistovich (1965), and Maddieson (1990), investigate medial consonant clusters using nasal and oral airflow measurements respectively. Marchal (1987) investigates consonant co-production employing electropalatography. Relatedly, Nolan (1992), and Zsiga (1993) employ electropalatography to investigate coarticulation. Spectrographic analyses are employed in Öhman (1966), and Zsiga (1992). Finally, Bird (1992) presents a study of consonant co-production involving synthesized speech production. To our knowledge, the present study is the first to demonstrate consonant co-production in terms of oral pressure.

Moreover, earlier studies have reported either (i) trans-consonantal vowel-to-vowel effects, to the exclusion of intervening consonant effects (Öhman 1966, Fowler 1983 *inter alia*), or (ii) inter-consonantal overlap, to the exclusion of flanking vowel effects (Browman and Goldstein 1986,1990, Zsiga 1992,1993, Bird 1992, Nolan 1992). The present study investigates the co-production of adjacent consonants in conjunction with their flanking vowels. To our knowledge, no previous study has reported on this confluence of phenomena.

## 2. Experiment One

The sole subject was an adult male speaker of the Seoul dialect (one of the authors). The subject has no history of pathological speech, and considers his Korean untainted by foreign languages. The subject was fitted with a mouth mask connected to pressure/flow transducers. One pressure tube was inserted behind the lips, thus recording oral pressure. A second tube was inserted nasally into the pharyngeal cavity, thus recording pharyngeal pressure. Finally, oral airflow was recorded. Eight combinations of VCCV sequences were employed, involving -pk- and -kp-clusters flanked by i-u, u-i, i-i, and u-u. The wordlist, read three times by the subject, is shown in (1).

(1) upki, upku, ipki, ipku, ukpi, ukpu, ikpi, ikpu

As can be seen, these nonsense forms consist of an onsetless syllable followed by a codaless one. Only two vowels, [i,u], were used, to control for features other than frontness and backness. Due to a regular process of post-obstruent fortition in Korean (See Kim-Renaud 1986 for more details), [pk] and [kp] will surface [pk'] and [kp'], respectively ([k'] and [p'] represent glottalized obstruents.) We assume that this fortition process has no effect on the coarticulatory phenomena under investigation, and therefore have not included this detail in our transcriptions. The list was read in the order shown, without a carrier phrase. An additional list of actual Korean words was recorded as well. This list is presented and discussed in Section 2.

*Results*

Figure (1) shows sample flow and pressure data for [upku] and [ipki]. These sequences consist of labial and velar voiceless stops flanked by vowels of identical phonemic quality. Comparable results were obtained in the other two trials.

The pressure records show that pharyngeal and oral pressure increase at exactly the same point in time (Point A). This indicates, as expected, that labial closure either temporally precedes or is simultaneous with dorsal closure, for an increase in oral pressure entails an increase in pharyngeal pressure as well. Were the dorsal closure to precede the labial one, there should be an increase in pharyngeal pressure *before* an increase in oral pressure. This is confirmed in Figure (2), which shows sample flow/pressure data for [ukpu, ikpi]. In Figure (2), increases in pharyngeal pressure temporally precede oral pressure increases (Point A precedes Point B), indicating that the dorsal closure temporally precedes the labial closure, as expected.

Returning now to Figure (1), observe that pharyngeal pressure is sustained as oral pressure returns to normal (Point B). That is, the increase in pharyngeal pressure is sustained for a longer period than the increase in oral pressure. This indicates both that dorsal closure precedes labial release, i.e., that the two closures temporally overlap for a period, (resulting in the phonetic equivalent of a labio-velar stop), and that the release of the dorsal closure occurs *after* the release of the labial closure.

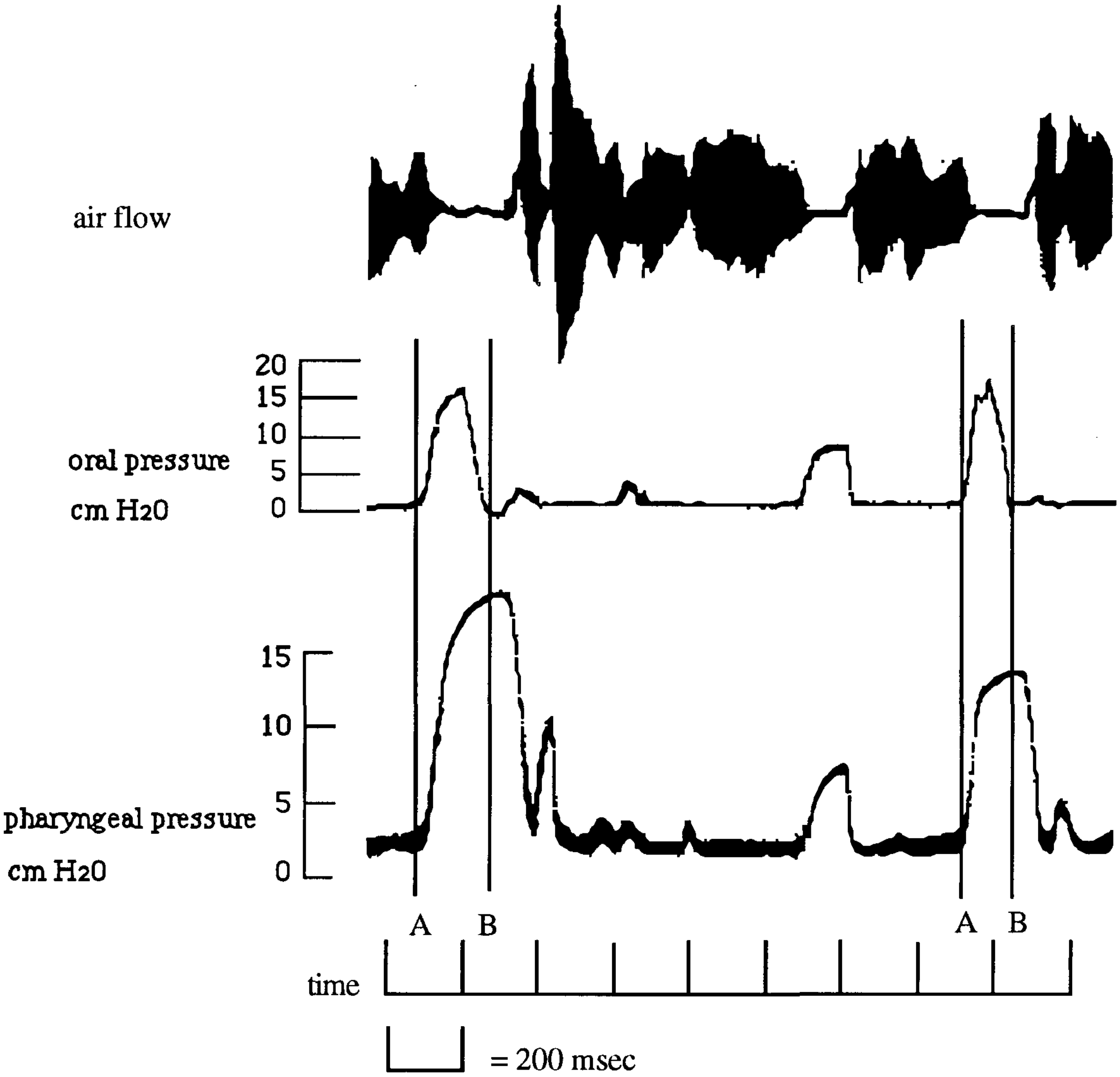


Figure 1. Sample airflow, oral pressure, and pharyngeal pressure records for [upku] and [ipki].

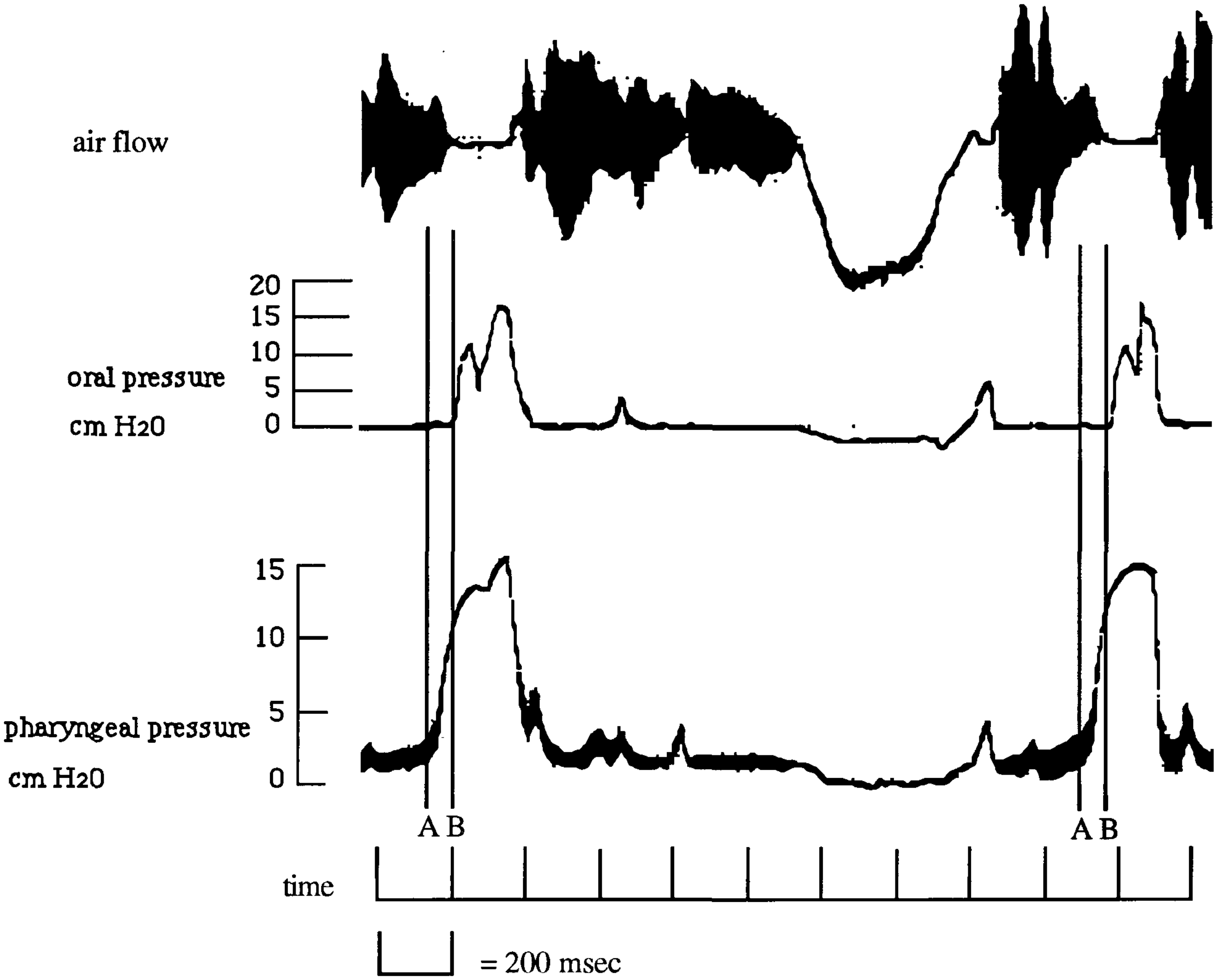


Figure 2. Sample airflow, oral pressure, and pharyngeal pressure records for [ukpu] and [ikpi].

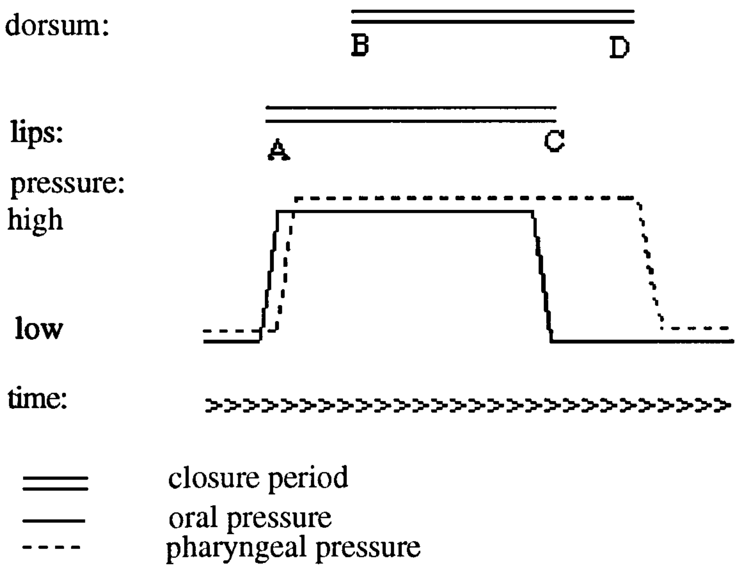


Figure 3. Schematic representation of dorsal and labial activity, as well as oral and pharyngeal pressure effects for [upku] and [ipki].

Figure (3) summarizes in schematic form the hypothesized articulatory configurations which produce the attested pressure data in Figure (1). Below the schematized closure record is the pressure data , also in schematized form.

At Point A, labial closure results in an increase in both oral and pharyngeal pressure. At Point B, dorsal closure occurs. (Point B, it should be noted, can only be estimated, as the pressure record for these data provides insufficient information to determine its exact point in time.) At Point C, the labial closure is released, resulting in oral pressure returning to normal, while pharyngeal pressure is maintained, due to the continuing dorsal closure. Finally, at Point D, the dorsal closure is released and pharyngeal pressure returns to normal.

If the schematic representation in Figure (3) truly reflects the actual state of articulatory affairs, it is predicted that manipulating the volume of the oral cavity during the point of closure overlap should result in pronounced pressure effects. That is, expanding the volume of the sealed cavity should result in a marked *decrease* in oral pressure. Conversely, contracting the sealed cavity should result in a marked *increase* in oral pressure.

This hypothesis is confirmed upon observing the pressure record for the same labial-velar consonantal sequence in a modified vocalic environment consisting of front-back vowels (i-u), and back-front vowels (u-i). Pressure/flow data for two tokens of [ipku] are presented in Figure (4).

Note the pronounced oral pressure rarefaction just after the labial closure (Point A). This, we claim, is due to the tongue retracting from a front position (for [i]) to a back position (for [u]) as the dorsal closure is implemented. Oral pressure rarefaction begins soon after the labial closure, indicating that the dorsal closure follows the labial closure almost immediately. The volume of the sealed oral cavity which results from overlapping labial and dorsal closures increases upon tongue retraction. This increase in volume is due, we claim, to the tongue body's sliding back across the soft palate. This articulatory configuration is schematized in Figure (5).

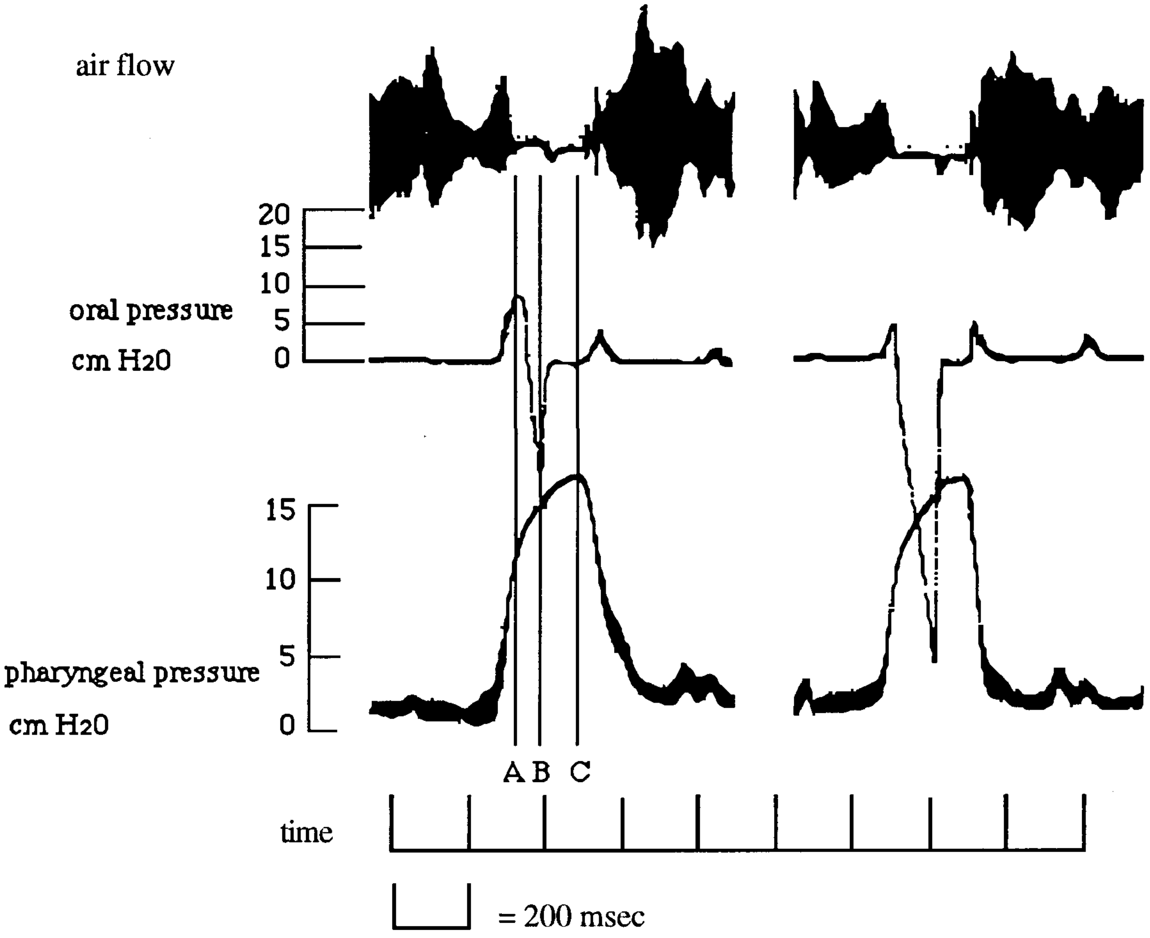


Figure 4. Airflow, oral pressure, and pharyngeal pressure records for two tokens

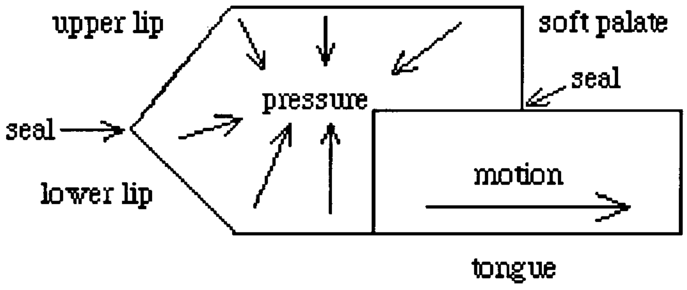


Figure 5. Schematic representation of articulator dynamics and resultant pressure effects during consonantal overlap in [ipku].

As the oral cavity is sealed at both ends, and as the labial closure is by and large fixed, cavity expansion due to tongue retraction across the soft palate is the only reasonable explanation of the pressure facts.

Returning now to Figure (4), as the labial closure is released, oral pressure increases back to normal (Point B). However, pharyngeal pressure remains high until dorsal release (Point C).

Results consistent with our hypothesis were found for the sequence [upki]. Figure (6) indicates a very pronounced increase in oral pressure upon labial closure for one token. Comparable results were obtained in the other trials.

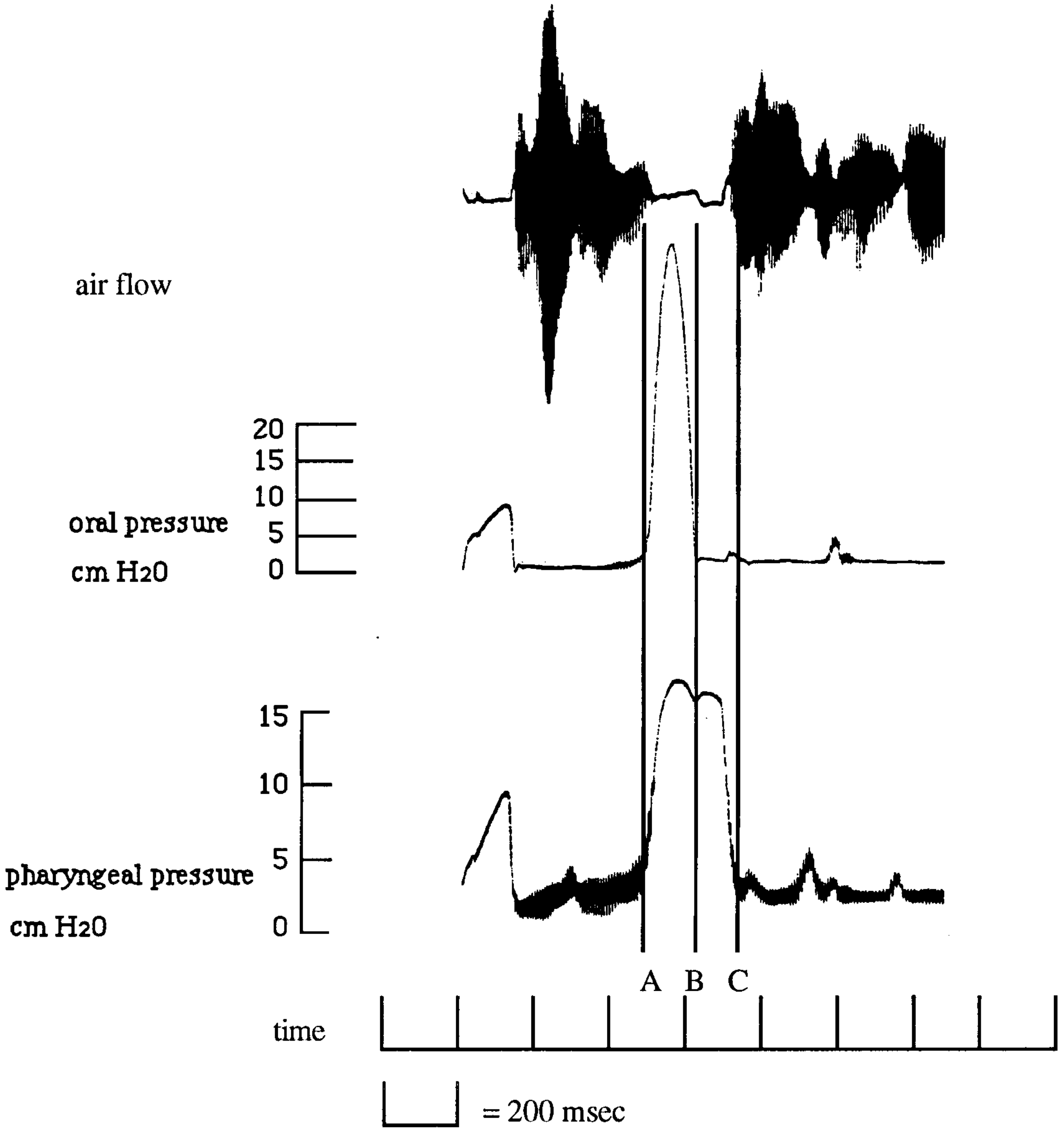


Figure 6. Airflow, oral pressure, and pharyngeal pressure records for [upki].

During the multiple closure, the oral pressure in Figure (6) is far greater than that observed when this same consonantal sequence is flanked by vowels of identical quality (cf. Figure (1)). This result is consistent with our hypothesis that the two closures overlap in time, and in addition, temporally co-occur with vocalic tongue movement: during the course of the overlapped labial and dorsal closures, the tongue body slides forward across the palate in its movement from back to front. The resulting contraction of the sealed oral cavity results in a marked increase in oral cavity pressure (Point A), which persists until the labial seal is broken (Point B). Pharyngeal pressure is maintained until the dorsal closure is released (Point C). This articulatory configuration is schematized in Figure (7).

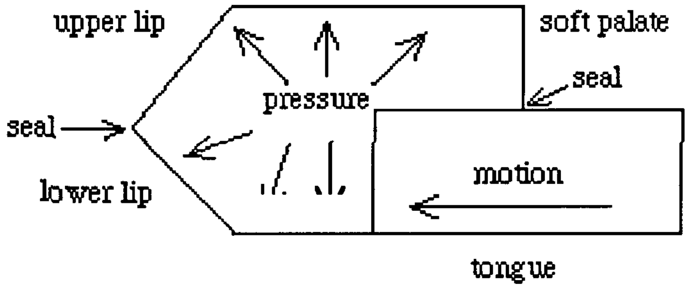


Figure 7. Schematic representation of articulator dynamics and resultant pressure effects during consonantal overlap in [upki].

Figures (8) and (9) provide a schematized representation of the data in Figures (4) and (6), with tongue body movement superimposed on the labial and dorsal records.

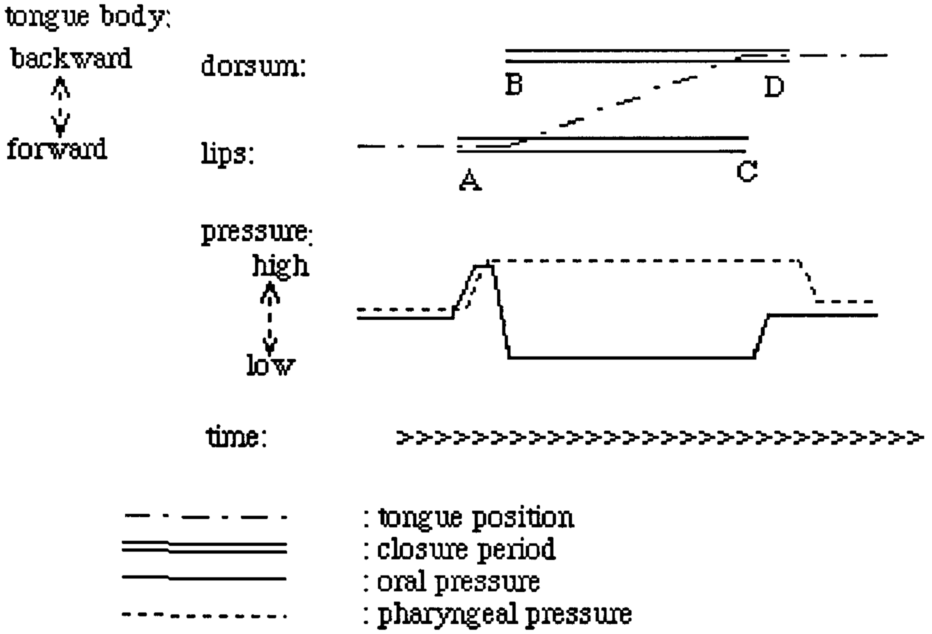


Figure 8. Schematic representation of dorsal and labial activity, tongue body activity, and pharyngeal pressure effects for [ipku].

In Figure (8), at Point A, a labial closure is formed, followed almost immediately by a dorsal closure (Point B). Simultaneously, the tongue body retracts, sliding back along the soft palate. Oral pressure rarefies due to oral cavity expansion. At Point C, the labial closure is released, and oral pressure returns to normal. Finally, at Point D, the dorsal closure is released, and pharyngeal pressure returns to normal. A similar scenario obtains for Figure (9), the only difference being that tongue advancement during the multiple closure results in greatly increased oral pressure. Note finally sample pressure records for the sequences [ukpi, ikpu] (Figure 10).

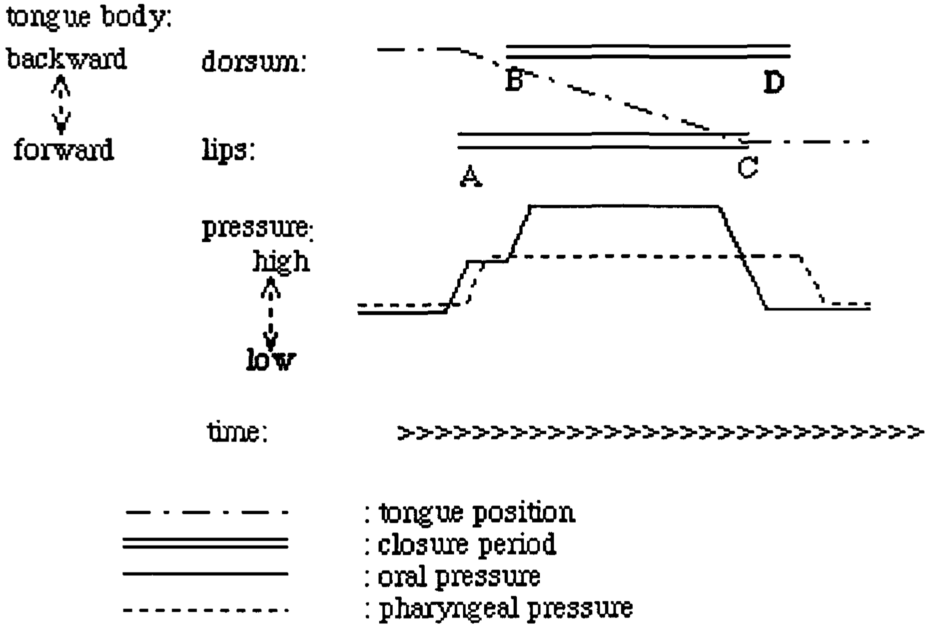


Figure 9. Schematic representation of dorsal and labial activity, tongue body activity, as well as oral and pharyngeal pressure effects for [upki].

For [ukpi], after an increase in pharyngeal pressure (Point A), a marked increase in oral pressure is observed (Point B). In contrast, for [ikpu] an oral pressure rarefaction is observed (Point B). Here again, these pressure effects are a result of the confluence of consonantal and vocalic articulations.

For [ukpi], at Point A, dorsal closure for [k] results in an increase in pharyngeal pressure. Immediately following, at point B, labial closure proceeds, with concomitant tongue advancement. The now sealed oral cavity experiences a marked pressure increase, just as observed for [upki]. At Point C, the dorsal closure is released, and oral pressure is reduced to equivalency with pharyngeal pressure. Finally, at Point D, the labial seal is broken, and both pressure records return to normal.

For [ikpu], comparable results were obtained. At Point A, dorsal closure results in heightened pharyngeal pressure. At Point B, labial closure with concomitant tongue retraction for the i-u vowel sequence results in a brief period of oral pressure rarefaction. At Point C, dorsal release results in an oral pressure increase to match pharyngeal pressure. Finally, at Point D, the labial closure is released, and both pressure records return to normal.

These results are schematized in Figures (11) and (12).

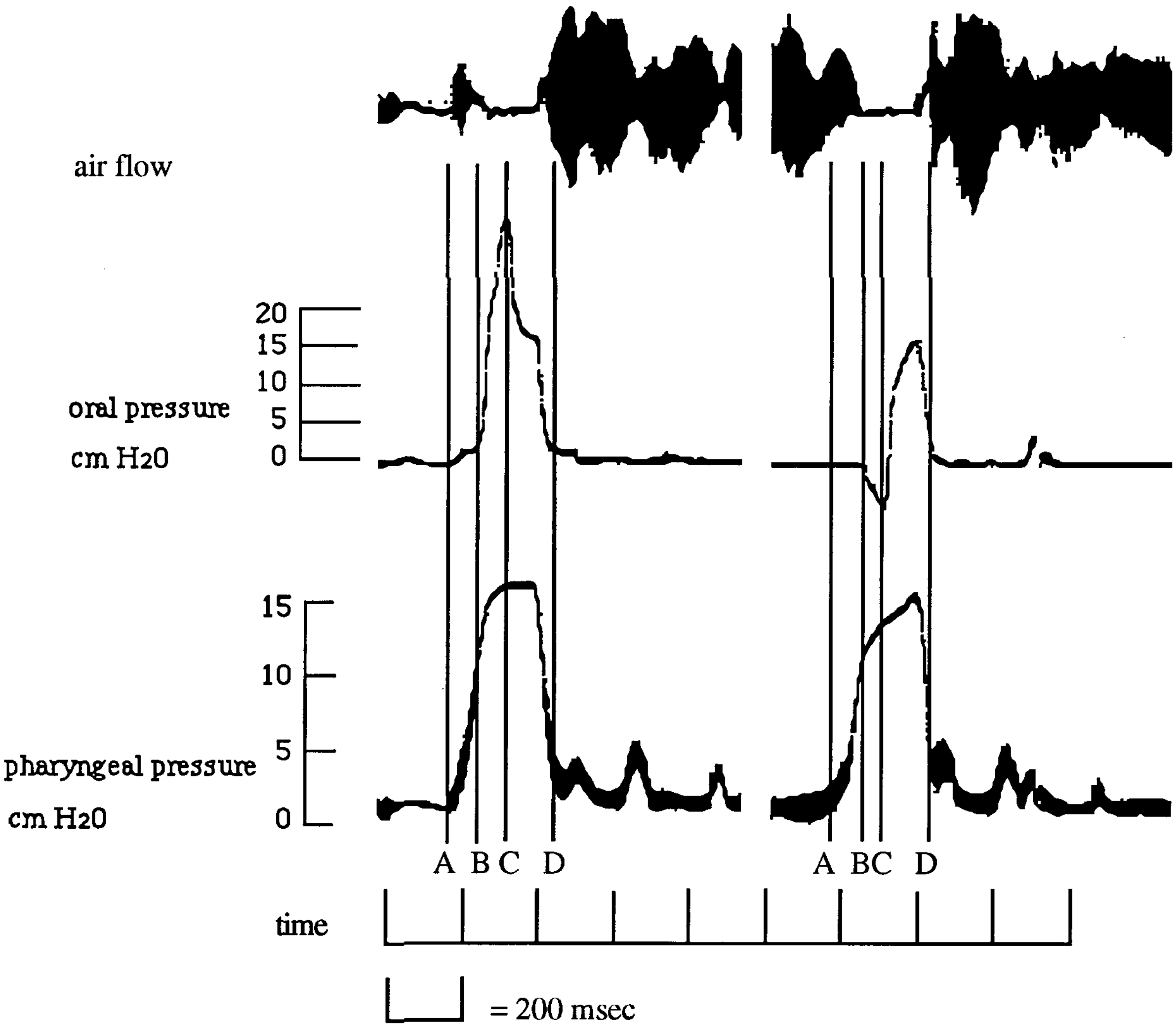


Figure 10. Sample airflow, oral pressure, and pharyngeal pressure records for [ukpi] and [ikpu].

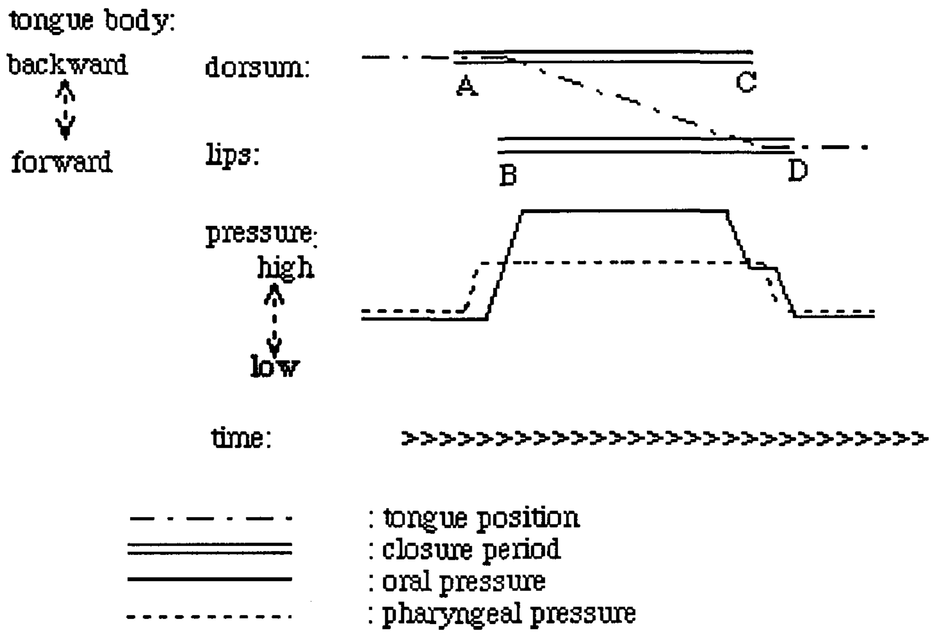


Figure 11. Schematic representation of dorsal and labial activity, tongue body activity, and oral and pharyngeal pressure effects for [ukpi].

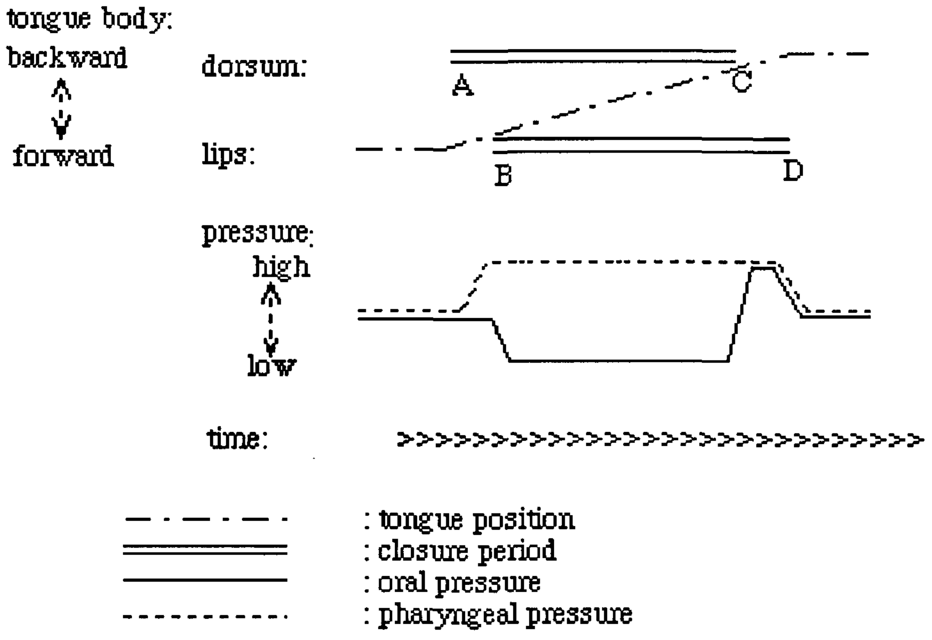


Figure 12. Schematic representation of dorsal and labial activity, tongue body activity, and oral and pharyngeal pressure effects for [ikpu].

*Summary*

Back-front vowel combinations, in conjunction with intervocalic -pk- or -kp-sequences produced a far greater increase in oral pressure during the course of the consonantal sequence than was found when the same consonantal sequence was flanked by vowels of identical phonemic quality. Front-back vowel combinations in conjunction with intervocalic -pk- or -kp- sequences produced a marked rarefaction in oral pressure.

Note finally the telling asymmetry in oral pressure between [ipku] (Figure 4) and [ikpu] (Figure 10). In [ipku], the pressure drop occurs after the pressure increase, whereas in [ikpu] the pressure drop occurs before the pressure increase. This shows that phonological ordering persists into the phonetics, despite the lengthy period of temporal overlap. Thus for [ipku], the labial closure precedes the dorsal closure, resulting in an initial oral pressure increase, followed by a rarefaction upon dorsal closure and concomitant tongue retraction. For [ikpu] however, oral pressure initially rarefies. The labial closure occurs only after dorsal closure. This second closure, with concomitant tongue retraction, results in an initial oral pressure rarefaction.

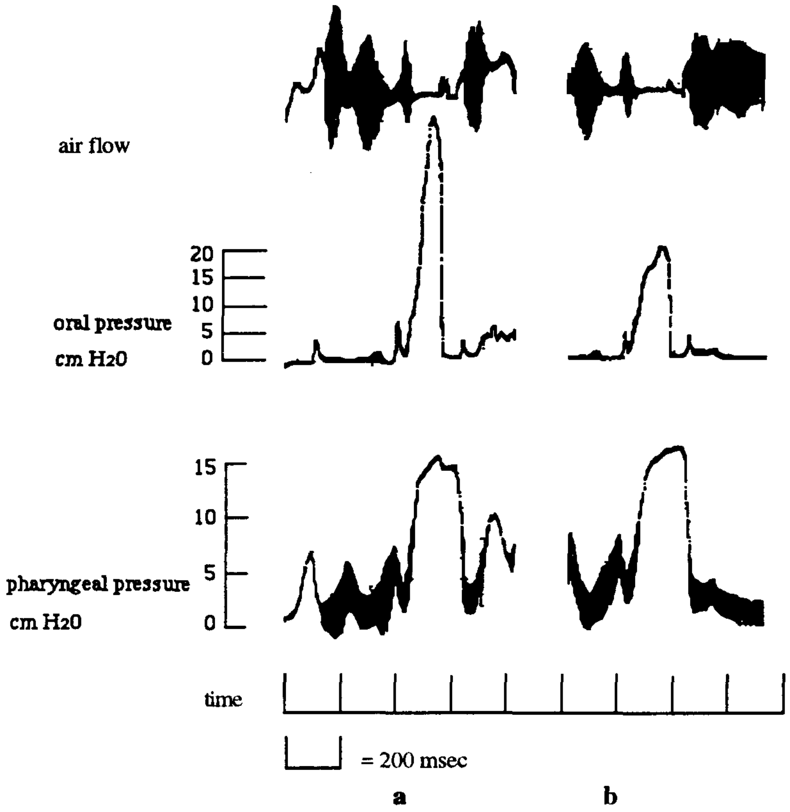
## 3. Experiment Two

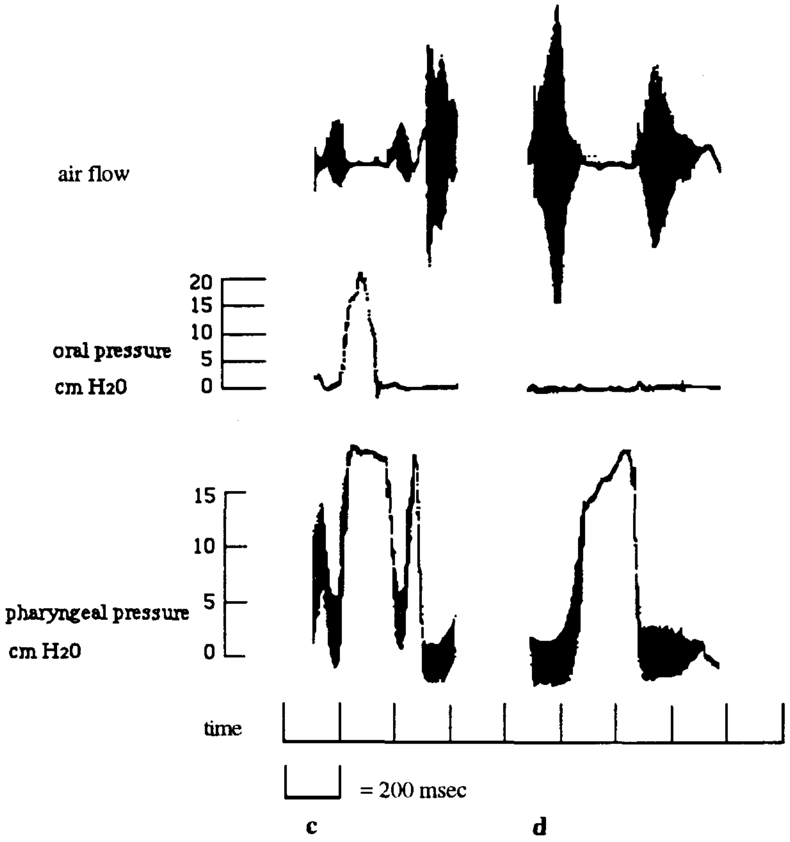
The same subject and methods were employed in Experiment Two. Also, the same VCCV sequences were employed as in Experiment One. However, this time real Korean words were used in phrasal/sentential contexts. In (2) is a complete list of the employed phrases/sentences. The relevant segmental sequences are underlined. The abbreviations used in (2) indicate the following: Con = Verb connective; SE = Sentence Ender; Acc = Accusative case marker; Nom = Nominative case marker

(2) a. /cəki nupki siləjo/
there lying hate
I hate lying over there
b. /cəki nup-ku sip-ne/ ([ku] is a free variant of the standard pronunciation, [ko].)
there lie-Con want-SE
I want to lie over there
c. /os-l ipki-ka himtɨl-ta/
clothes-Acc wearing-Nom hard-SE
It is hard to wear the clothes
d. /hakkyo ipku/
school entrance
school entrance
e. /kukpi juhak/
awarded by the nation studying abroad
overseas study fellowship
f. /pukpu cipaŋ/
north province
north area
g. /sikpi putam/
food expense burden
burden of food expenses
h. /sikp$^{h}$um-pu/
food section
food section

### *Results*

The results of Experiment Two are shown in Figures 13a-h.





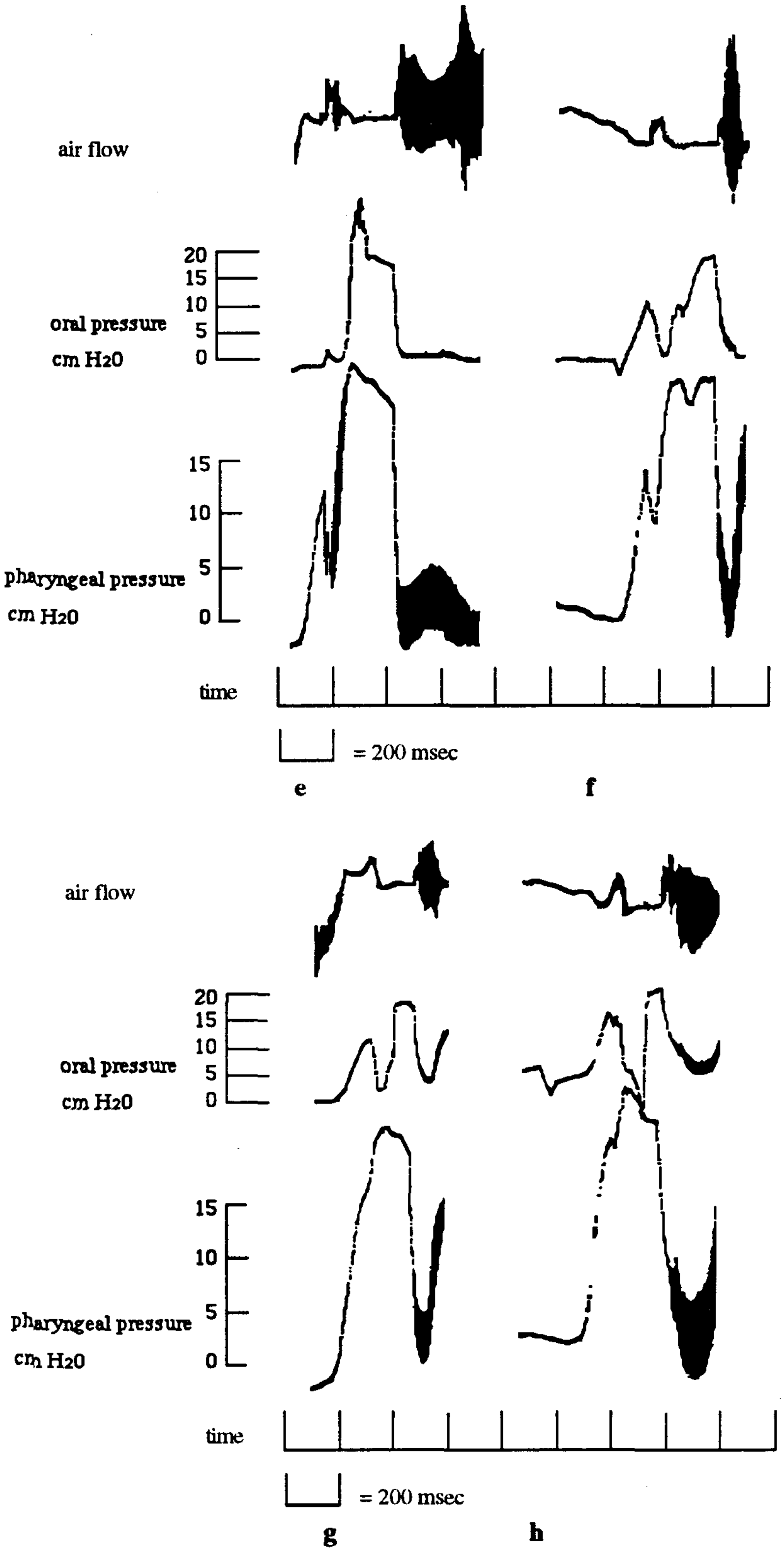


Figure 13. Sample airflow, oral pressure, and pharyngeal pressure records for [upki, upku] (a,b), [ipki, ipku] (c,d), [ukpi, ukpu] (e,f), and [ikpi, ikpu] (g,h), extracted from real words.

Results are identical, except for the pressure readings for [ipku]. Figure 13c indicates no change in oral pressure for this form. Identical results were obtained in another trial, for this form only. This, we think, is due to a process of Korean place assimilation whereby labial consonants assimilate in place to a following velar consonant. If /ipku/ is realized [ikku], no change in oral pressure is expected, as there is no labial closure. It should be noted that this is an optional process, thus accounting for the mixed results obtained. In the following section, this result is considered in more detail. We conclude that the results of Experiment Two are consistent with those of Experiment One.

## 4. Discussion

In Experiment Two, no change in oral pressure for /ipku/ was observed. This, we suspect, is due to an optional process of place assimilation. Based on previous work in Korean phonology (Kim-Renaud 1974; Cho 1990 among others), we assume that this process is optional, and dependent on the style and the rate of speech. In casual speech, coronal obstruents assimilate in place to a following consonant (3a, b); labials assimilate only to a following velar (3c, d).

(3) Korean Place Assimilation

| | | | | | | |
|---|---|---|---|---|---|---|
| a. | /tat | + | ko/ | --> | [takk'o] | 'close + and' |
| b. | /nac | + | pam/ | --> | [napp'am] | 'daytime (and) night' |
| c. | /ip | + | ko/ | --> | [ikk'o] | 'wear + and' |
| but | | | | | | |
| d. | /ip | + | ta/ | --> | [ipt'a] | 'wear + Sentence ender' |

We briefly discuss the implications of these results for the theory of Articulatory Phonology (Browman and Goldstein 1986, 1990, 1992). Within this theory, casual speech alternations such as those in (4) (and presumably (3)) are seen as the result of gestural overlap and/or gestural reduction. Gestural overlap involves the obscuring of one gesture by another temporally co-occurring gesture. Gestural reduction involves the reduction in magnitude of a gesture.

(4) Browman and Goldstein (1990:359)

| | | | |
|---|---|---|---|
| (a) /'mʌst bi/ | --> | ['mʌsbi] | ("must be") |
| (b) /hʌndrəd 'paʊndz/ | --> | [hʌndrəb 'paʊndz] | ("hundred pounds") |
| (c) /'graʊnd 'prɛʃɚ/ | --> | ['graʊm 'prɛʃɚ] | ("ground pressure") |

Thus, what a phonologist might model as the regressive place spreading of /p/ with place delinking of /d/ in (4b), emerges from the gestural model as /p/ superimposed on /d/, with the /d/ gesture maintained, though possibly in reduced form.

The present data are in accordance with this approach to casual speech alternations. Experiment One shows that nonsense forms with labial//velar sequences involve a high degree of gestural overlap. Experiment Two, in which real words are employed, also shows this high degree of gestural overlap. In addition, however, those -pk- sequences in which no oral pressure change is observed indicate that gestural reduction is playing a role here as well: when no oral pressure change in -pk- sequences is observed, we may conclude that labial closure does not take place. Instead, this labial gesture is reduced, perhaps to zero.

Jun (in prep.) investigates in greater detail the distinct roles of both gestural overlap and gestural reduction in the production and perception of casual and formal speech involving /pk/

sequences in Korean. Thus far fourteen native Korean speakers have been tested. Preliminary results suggest that gestural reduction plays a decisive role in the perception of /pk/ sequences: /pk/ sequences displaying pressure changes, and those not displaying pressure changes, are seemingly perceived as unassimilated and assimilated respectively. That is, /pk/ with no labial closure is perceived as [kk], while /pk/ with labial closure is perceived as [pk].

## 5. Conclusion

The significance of the present study rests in the methodology employed, in the results obtained, and, potentially, in the theoretical conclusions that may be drawn. To our knowledge, no previous study demonstrates consonant co-production in terms of oral air pressure, and further, no previous study reports on coarticulatory effects involving four ordered segments. The theoretical implications of the obtained results may prove useful to both phoneticians and phonologists in their investigations of phonetic coarticulation and phonological spreading processes.

## References

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

Browman, C.P. and L. Goldstein (1990) Tiers in Articulatory Phonology, in John Kingston and Mary E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge University Press: 341-376.

Browman, C.P. and L. Goldstein (1992) Articulatory Phonology: an overview. *Phonetica* 49:155-180.

Cho, Y.Y. (1990) *Parameters of Consonantal Assimilation*, Ph.D. dissertation. Stanford University.

Demolin, D. (1992) *Le Mangbetu. Etude Phonetique et Phonologique.* Doctoral Dissertation, Universite Libre de Bruxelles.

Fowler, C.A. (1983) Converging sources of evidence on spoken and perceived rhythms of speech: cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112.3:386-412.

Jun, J. (in prep.) On gestural reduction and gestural overlap in Korean and English /pk/ clusters. UCLA.

Kim-Renaud, Y.K. (1986) *Studies in Korean Linguistics*, Seoul: Hanshin.

Kozhevnikov, V.A. and L.A. Chistovich (1965) *Speech: Articulation and Perception..* Joint Publication Research Service (U.S. Dept. of Commerce) No.30,543.

Ladefoged, P. (1962) *A Phonetic Study of West African Languages.* Cambridge University Press.

Maddieson, I. (1990) Shona velarization: complex consonants or complex onsets? *UCLA Working Papers in Phonetics* 74: 16-34.

Marchal, A. (1987) Des Clics en Francais? *Phonetica* 44.30-37.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation, in G.J. Docherty and D.R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge University Press. 261-280.

Ohala, J.J. (1981) Speech timing as tool in phonology. *Phonetica* 38:204-212.

Öhman, S. E. G. (1966) Coarticulation in VCV utterances: spectrographic measurements. *JASA* 39:151-168.

Zsiga, E. (1992) Acoustic evidence for gestural overlap in consonant sequences. *Haskins Laboratories Report on Speech Research 1992,* SR-111/112:1-20.

Zsiga, E. (1993) *Features, Gestures, and the Temporal Aspects of Phonological Organization.* Ph.D. dissertation, Yale University.