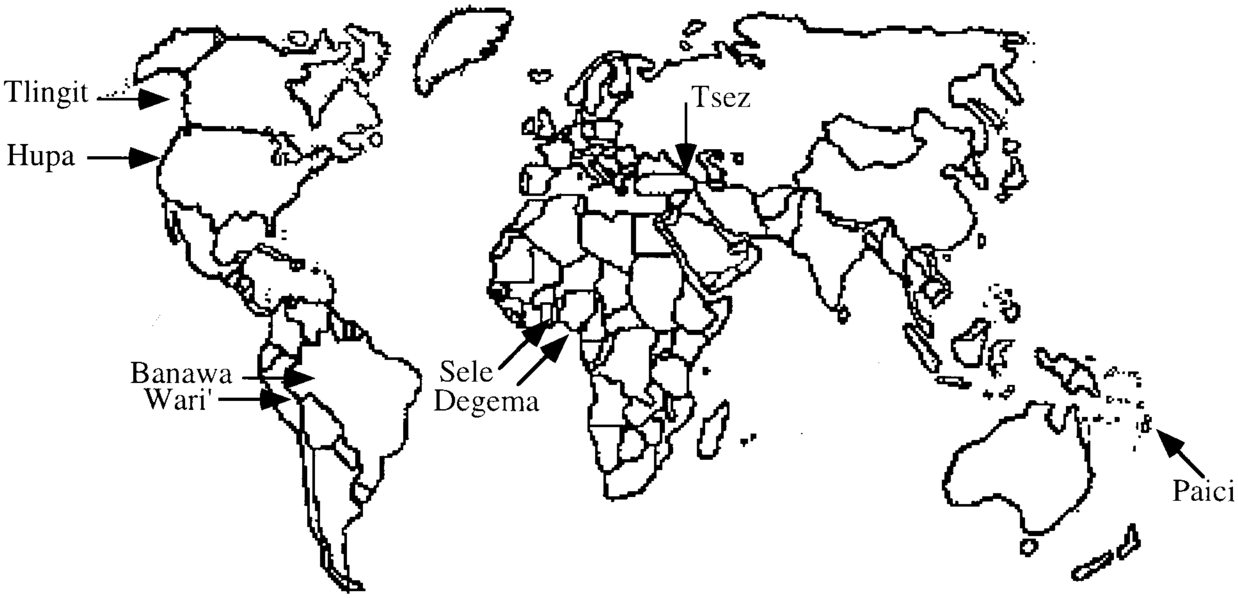
Title

WPP, No. 93: Fieldwork Studies of Targeted Languages IV

Publication Date

1996-12-01

Fieldwork Studies of Targeted Languages IV



UCLA WPP Volume 93

December 1996

The UCLA Phonetics Laboratory Group

Victoria Anderson
Marco Baroni
Barbara Blankenship
Taehong Cho
Sandy Disner
Christina Foreman
Susan Meyers Fosnot
Cécile Fougeron
Sean Fulop
Elissa Fultz
Matt Gordon
Bruce Hayes
Chai-Shune Hsu
Sun-Ah Jun
Pat Keating
Haejin Kim
Keeho Kim
Hyun-Gi Kim
Sergei Knyazev
Jenny Ladefoged
Peter Ladefoged
Ian Maddieson
Peggy MacEachern
Elizabeth Menard
Kie Ross
Aaron Shryock
Dan Silverman
Caroline L. Smith
Donca Steriade
Henry Teheranizadeh
Kimberly Thomas
Moto Ueyama
Jie Zhang

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:
USPHS grant 5 T32 DC00029
NSF grant SBR 9319705
NSF grant SBR 9511118
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 90095-1543

The General Editor of UCLA Working Papers in Phonetics is Ian Maddieson.

UCLA Working Papers in Phonetics

Number 93

December 1996

Fieldwork Studies of Targeted Languages IV

Table of Contents

Recording the phonetic structures of endangered languages

Peter Ladefoged and Ian Maddieson

Making a record of the phonetic structures of an endangered language involves finding suitable items that illustrate all the phonological contrasts, and finding a group of fluent speakers. The material is recorded on DAT recorders and then analyzed using standard acoustic analysis techniques. Aerodynamic, palatographic and glottographic data is also recorded when appropriate.

INTRODUCTION

There are about 7,000 languages spoken in the world today; but soon, probably by our great grandchildren's time, only a few hundred may be left. Every language that dies represents a significant loss to human culture, often causing great suffering among those affected. In addition, for those of us who are concerned with the sounds of languages, the disappearance of a language is a loss of a resource for the scientific study of human speech communication. Our notions of what constitutes a possible human language depend on studies of actual human languages. Many linguistic concepts are shaped by ideas of what is normal or unmarked in a language. This paper will describe a project that aims to provide descriptions of the salient phonetic structures of some endangered languages before it is too late.

Making a record of the phonetic structures of an endangered language is, in principle, no different from making a record of the phonetic structures of any language. From a linguistic point of view, we want to record the facts about each language so that we can compare them with other languages. This implies that we have some universal phonetic framework that will accommodate the description of the sounds of any language. Less obviously, it also implies that we are working within some phonological framework that can describe all the differences in sound systems that we are likely to observe. We cannot describe the sounds of a *language*, without first considering the patterns of sounds, the phonological contrasts, that are necessary for it to be a language. As phoneticians we are interested in the physical manifestations of these patterns of sounds.

MATERIAL TO BE RECORDED

All languages have consonants and vowels, and we can start describing the salient phonetic structures of a language by describing these entities. Consonants and vowels occur only as parts of words, the smallest components of a language that can occur on their own. Generally the first task in describing the phonetic structures of a language is to draw up a list of words exemplifying the minimal contrasts. Ideally we want to record the language as it occurs in normal, everyday, spoken utterances. But this notion conflicts with the desire to describe all the contrasting sounds in a manner such that they can be compared with those of other languages. If we want to describe, for example, the vowels of English, it is impractical to just record a lot of free speech and hope that it will contain utterances in which somebody says each vowel between the same consonants, and with the same degree of stress and the same intonation. We cannot wait around in hopes that our speakers will all say 'bead', 'bid', 'bade', 'bed', 'bad', 'bod', 'bud', 'bird', 'booed,' etc. We must ask them to say the words that we want to record. It may not be the most natural form of speech, but a word list provides the only way in which one can record the complete set of sounds of a language as produced by a number of speakers.

Our normal procedure is to try to make a word list illustrating each vowel in at least two different stressed syllables, one beginning with a coronal stop, and one with some other consonant. If it is appropriate, we also record examples of unstressed vowels. Thus our word list for Banawa, an Arawakan language with four vowels, spoken in the Amazonian rain forest, included vowels in stressed penultimate syllables such as **bita** 'mosquito', **befa** 'other', **bata** 'pick', **bufa** 'put on water', and in the unstressed final syllables such as **ibi** 'each other', **ibe** 'a strip', **iba** 'put, place, **ibufa** 'dump into water'. The structure of the language often does not permit the compilation of such simple sets of words. Languages with complex morphology, a large number of consonants and vowels that contrast in tone, length and nasalization, may not have many minimal pairs. As a result, it may not be easy to compile a word list that includes all the contrasts in comparable forms. In Apache, an Athabaskan language with only three vowels, the morphological complexities and the elaborate consonantal system forced us to use a word list that was not as elegant as that for Banawa. Each of the three vowels can contrast in nasalization, length and tone as well as quality. In order to observe the interactions of all these features, nearly all of which can co-occur, we needed to record 48 contrasting items.

The form of the word list required for illustrating consonants depends on the structure of the language. In a language with only CV or CVV syllables and no coda consonants, we try to record each consonant preceding at least two different vowels in a stressed syllable, or in syllables with matching tones. In Pirahã, a Mura language of this type spoken in the Amazonian rain forest, we were able to extend the word list so that we recorded each of the eight consonants before each of the three vowels. But in Toda, a Dravidian language spoken in the Nilgiri Hills in India, it was more important to record the consonants in final position, as this language contrasts three places of articulation word initially, but six places in final position.

Devising an appropriate set of material to record is the key element in investigating the phonetic structures of any language. Our normal procedure before going into the field is to examine the phonology of the language as thoroughly as possible, considering dictionaries, grammars, and whatever is available, and often consulting extensively with other linguists who have been working on the language. Then, with a good knowledge of the probable phonetic structures and a preliminary word list in hand, we go into the field and spend a considerable amount of time checking this word list with one or two language consultants, often revising it extensively as we hear the language being spoken and are able to take into account the comments of native speakers on our own attempts to produce the sounds. We find that we cannot over-prepare before going into the field, and we cannot over check a list of words in the field before making a recording.

SPEAKERS

When we are satisfied that we have an appropriate set of material we record several speakers saying these words. A language is the property of a group of speakers, and it cannot be validly illustrated without recording several individuals. Any one speaker may have some individual speaking habits that are not typical of the language as a whole. Successful comparisons of one language (or one dialect) with another require measurements of the phonetic characteristics of a group of people, so that the mean and standard deviation of various properties can be determined. We like to record at least six women and six men, although we often have to settle for less. It is important to record both sexes as, quite apart from the physiological differences in the vocal tracts, the men may have a different dialect from the women. In Pirahã, the language previously mentioned in which there are only eight consonants, the women have a palatal fricative **ç** for one of those consonants, where the men have **h**.

Speakers need to be carefully chosen from several points of view. Firstly good dialect control is necessary. It is usually possible to ask speakers whether they regard themselves as typical speakers of the local dialect, and also to ask them whether they can hear any different local characteristics in the speech of the other individuals being recorded. If possible, we try to avoid speakers that are hard of hearing. Profoundly deaf speakers almost always have different speech characteristics, and even the slightly deaf may have been affected. Shy and easily embarrassed speakers are also difficult to record. Middle aged or younger speakers with strong voices and all their wits and teeth make the best language consultants for phonetic purposes, but they are often not available in endangered languages. In some languages only older speakers are still sufficiently fluent to be able to provide the required data. In such cases we make use of the best speakers available.

AUDIO RECORDING

The principal data in any study of the phonetic structures of a language are the audio recordings. At the moment the best way of making such recordings is on a DAT (Digital Audio Tape) machine. A short while ago analog tape or cassette recorders were considered most suitable; and in a year or two recordings will probably be made directly onto a computer and then transferred onto a more permanent storage medium. DAT and computer recordings have the unfortunate property that they can become completely unusable because of a single small fault in the recording medium that would cause only a local noise in an analog magnetic recording. It is therefore vital to make more than one backup copy of any DAT or computer recording. Fortunately copies of digital recordings can be made without any loss of quality. It is probably advisable to make analog copies of digital recordings as well. The field is changing so rapidly that it is difficult to reproduce digital records made more than a dozen years ago. But magnetic recordings made in 1959 were still usable for producing spectrograms in 1995 (Ladefoged & Maddieson 1996); and the old wax or tin cylinders made at the turn of the century may be noisy but they are still a usable source of material for a contemporary dictionary (Bergsland 1994) or text collection (Dauenhauer & Dauenhauer 1990).

The type of microphone used for making the recording is just as important as the type of recording machine. Many recordings of endangered languages have to be made in noisy surroundings, in villages in which there are engines running, children crying, cocks crowing, and all the sounds of community life. The best way of dealing with this situation is to use a close-talking noise-canceling microphone, held by a head mount a few centimeters to the side of a speaker's lips.

Audio recordings are the prime source of data for the detailed description of the phonetic structures of every language. Of course much information is gained from direct listening and observation of speakers. But far more is learned from recordings that can be played over and over again, in small sections, and, if appropriate, at a slow rate without loss of quality. We make a great deal of use of computer systems that permit audio recordings to be examined in detail while still in the field as well as on returning home, where acoustic analyses of these recordings may be carried out.

Some acoustic analyses of audio recordings can be conducted in the field; examination of tonal contrasts is often helpful, especially for investigators not accustomed to analyzing pitch phenomena. But most acoustic analyses are better conducted in a laboratory where more extensive facilities are available. Typical phonetic properties that need to be described in acoustic terms are the mean formant frequencies of the vowels, the voice onset time of stop consonants, and the durations of vowels and consonants in specified contexts. As such measures are well known, they will not be further exemplified here.

PHYSIOLOGICAL DATA RECORDING

Audio recordings are far from the only type of data that has to be collected for studying the phonetic structures of endangered languages. The acoustic analysis of audio data will provide much of the data needed for a description of the phonetic structures, but direct recordings of physiological data are also required. It is possible to gain useful information on the airstream mechanism using aerodynamic data recording techniques; the place and manner of articulation can be studied by means of palatography; video recording is useful for observations of labial gestures; and electroglottography provides information on the state of the glottis. In addition some kinds of perceptual experiments can be conducted in the field. The remainder of this paper will exemplify data from endangered languages using each of these techniques. More detailed accounts of their use are given in Ladefoged (1997).

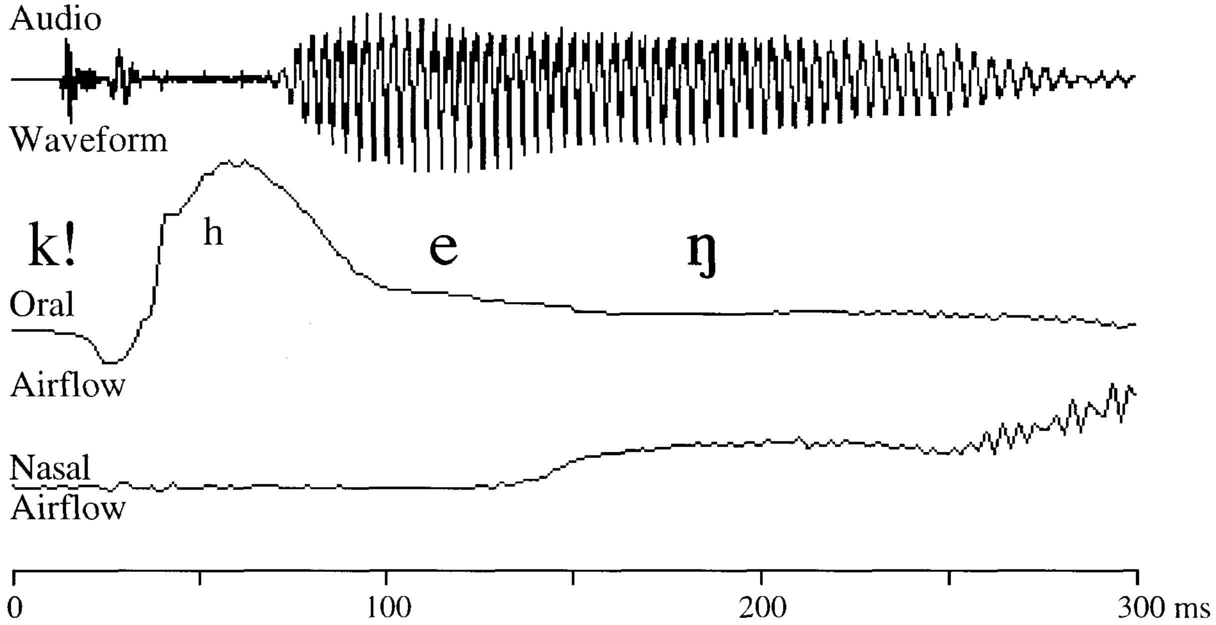


Figure 1. Aerodynamic record of the voiceless aspirated post-alveolar click in **k!ʰeŋ** ‘tongue’ in Sandawe.

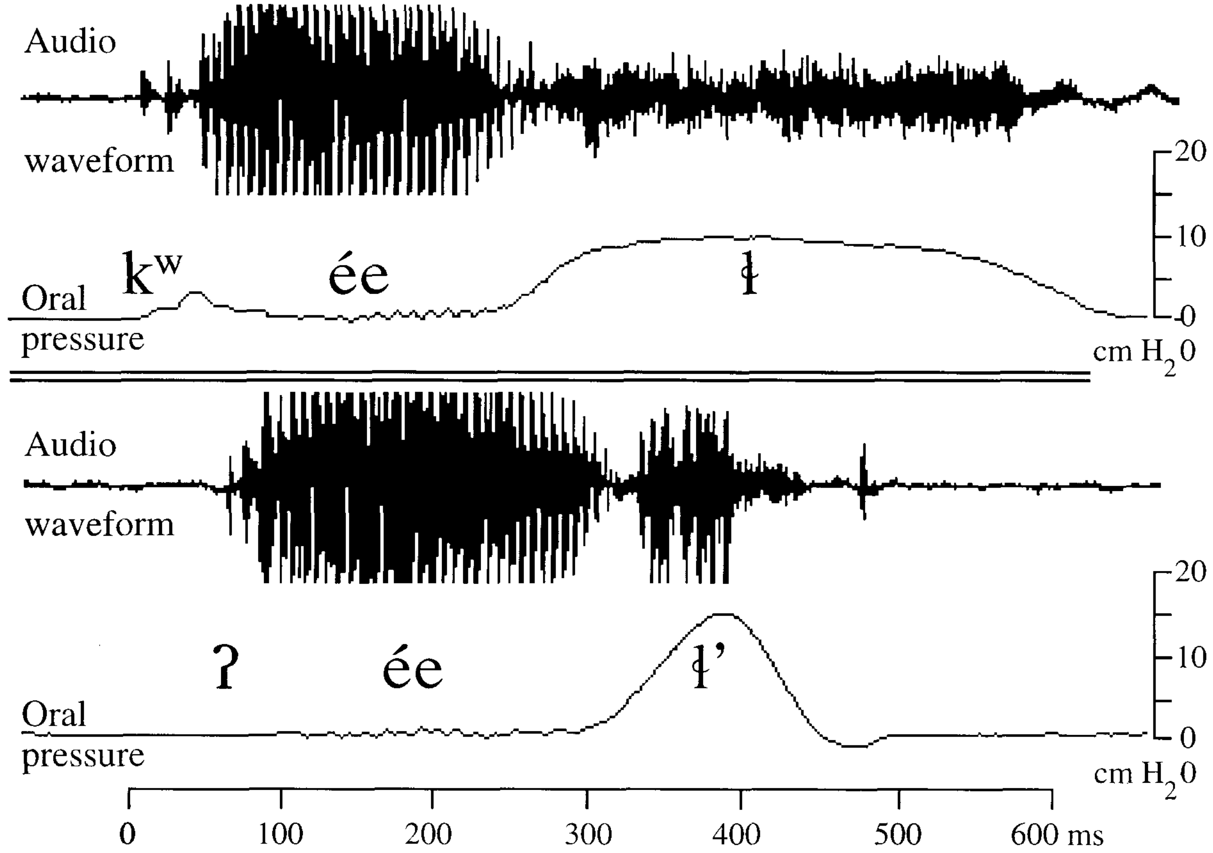


Figure 2. Aerodynamic records showing pulmonic and ejective lateral fricatives in Tlingit.

Aerodynamic data on a voiceless aspirated post-alveolar click in Sandawe (possibly Khoisan, spoken in Tanzania) are shown in Figure 1 (from Wright, Maddieson, Ladefoged and Sands, 1995). Following the inward airflow due to the click release, the aspiration is marked by the high-volume outward oral airflow, and a considerable delay before vocal fold vibration begins for the vowel. Nasal airflow is apparent for the final consonant, trailing off into a breathy voiced nasal at the end of the word.

A second example of aerodynamic data is shown in Figure 2, which illustrates the contrast in Tlingit between a pulmonic lateral fricative (in the upper part of the figure) and an ejective lateral fricative (in the lower part). The ejective has a much higher oral pressure, and a shorter duration.

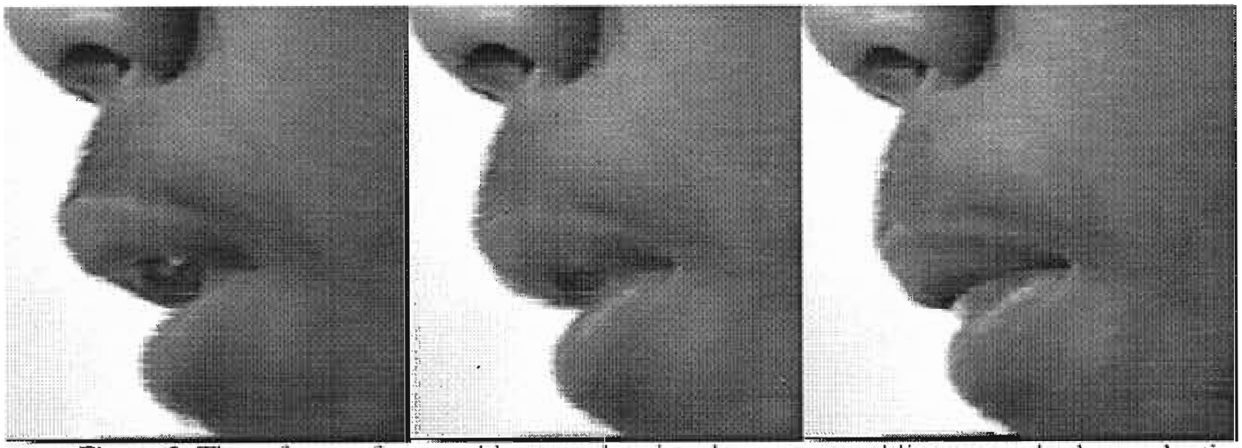


Figure 3. Three frames from a videotape showing the tongue and lip contact in the production of the word **nan̼d̼ak** ‘bow’ by a speaker of Vao.

Video recording is especially useful for obtaining data on lip positions. It is often important for making accurate descriptions of the degree of rounding in vowels. Some languages have unusual labial consonants that can be well documented by video recording. For example, Vao, an Austronesian language spoken in Vanuatu, has linguo-labial plosives, nasals and fricatives. Figure 3 (from Ladefoged and Maddieson 1996) shows three frames from a videotape of the consonant in the middle of the word **nan̼d̼ak** ‘bow’. This record makes it clear that not only does the tongue come out to meet the upper lip, but that in addition the upper lip is drawn inward to meet the tongue.

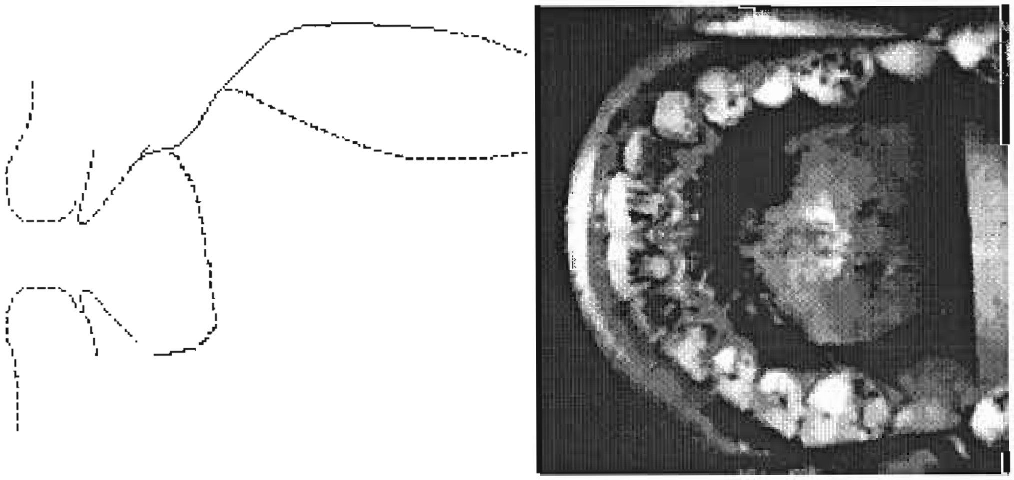


Figure 4: Palatogram and inferred sagittal view of the click in the Sandawe word **k!amba** ‘spleen’.

Changes in the articulatory contacts during a word are best recorded by electro-palatography, but this technique can rarely be used in studies of endangered languages. Butcher (forthcoming) was able to make false palates with embedded electrodes for recording tongue contacts for a number of speakers of Australian languages, but it is often not possible to do this in fieldwork situations. However, a great deal of information can be obtained from static palatography, which records all the tongue contacts that occurred in an utterance. If the shape of the palate is known from a dental impression, and the part of the tongue that made the contact is known from a separate photograph, it is possible to infer possible positions of the vocal organs. Figure 4 presents palatographic data (from Wright et al. 1995) for the unaspirated counterpart of the Sandawe click in Figure 1.

Laryngeal activity can be investigated concurrently with aerodynamic data, as illustrated in Figure 5 (from Flemming, Ladefoged and Thomason,1994), which shows the pronunciation of the Montana Salish word **tʃ'tʃen'** 'Where to'. The top line is an expanded section of the part of the second line between the arrows. These two lines of EGG data reflect the laryngeal movements. The unexpanded record shows the extensive laryngeal activity for the ejective, some of which occurs as the glottal closure is formed, with the greater part occuring during the release of the ejective and the accompanying intrusive laryngealized vowel. The expanded section of the EGG record (the top line) shows the shape of the glottal pulses during the glottalized nasal at the end of the word. The aerodynamic records in the remaining lines of data show the intrusive nasal vowel, and the varying pressures and airflow associated with the ejective affricate (which, somewhat surprisingly, has a lower oral pressure) and the pulmonic affricate.

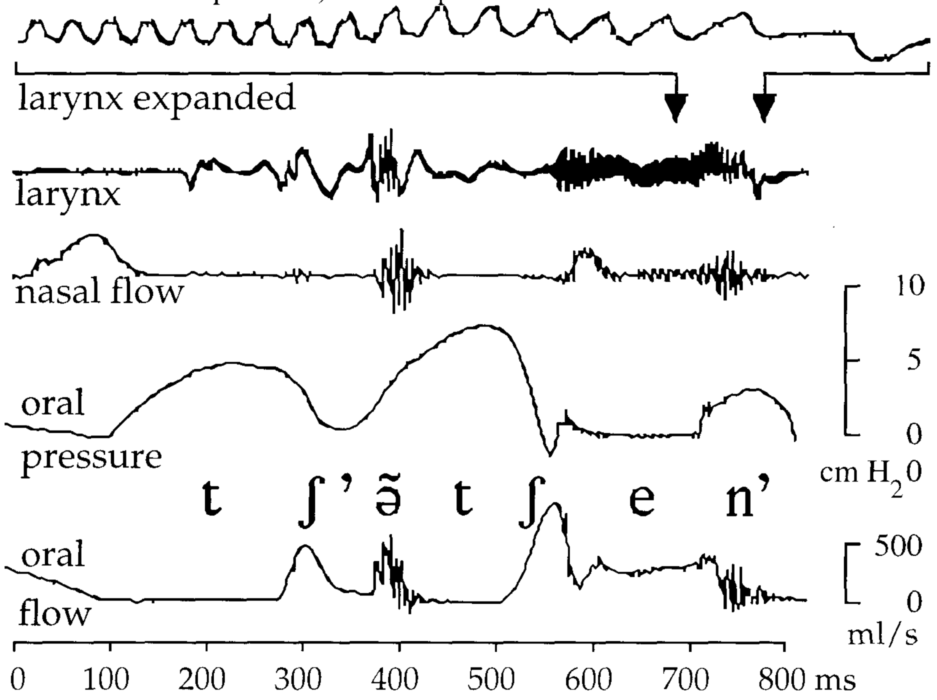


Figure 5. Aerodynamic records of a plosive and an ejective in the Montana Salish word **tʃ'tʃen'** 'Where to'.

PERCEPTUAL EXPERIMENTS

Simple perceptual experiments involving computer manipulations of recorded utterances can be easily performed in the field. We have investigated the pereception of voice onset time (VOT) by adding and deleting 10 ms sections from the waveforms of Mazatec utterances. Another investigation involved using a program that allows us to make independent variations of the pitch and length of different syllables, enabling us to study the perception of stress in Pirahã. One of the difficulties of this kind of research is that many speakers of endangered languages are not test sophisticated. Some of our Mazatec speakers became too bored or too frustrated to be able to complete the task. The Pirahã found it difficult to decide which of two words was being played without also deciding who the speaker was; if they could not do both these things they were apt to call the response simply 'crooked speech'. But valuable data is sometimes forthcoming. Wright (1996) manipulated complex consonant onsets in Tsou, and was able to show the importance of the stop release. Kari and Ladefoged (forthcoming) obtained interesting results from a perceptual experiment involving on-line synthesis in a study of vowel harmony in Degema. It is always useful to be able to supplement analyses of acoustic and physiological data by reference to data on what listeners find important.

CONCLUDING COMMENTS

Original data on endangered languages must be made as widely available as possible. This should include the text, glosses and phonetic transcription of audio reordings. Phonetic archives that have recordings with no accompanying written texts are seldom of any use. Instrumental records should also be carefuly preserved, along with interpretive comments. Linguists can manage with written records of dead languages for many aspects of their work. But in phonology and phonetics we must build our theories on explicit accounts of what people do, and what they sound like. These theories often refer to notions of universality, so we must be able to provide material for as many languages as possible. If this is not done soon, the number of languages on which our ideas can be based will be severely limited.

REFERENCES

Bergsland, K. (1994) *Aleut Dictionary*. Fairbanks: Alaska Native Language Center, University of Alaska.

Butcher, A. (forthcoming) *The Phonetics of Australian languages.*

Dauenhauer, N.M. and Dauenhauer, R. (1990) *Haa Tuwunáagu Yís, For Healing Our Spirit: Tlingit Oratory*. Seattle: University of Washington Press.

Flemming, E., Ladefoged, P. & Thomason, S. (1994) "Phonetic structures of Montana Salish." *UCLA Working Papers in Phonetics* 87: 1-33.

Kari, E. & Ladefoged, P. (forthcoming) "A perceptual investigation of vowel harmony in Degema."

Ladefoged, P. (1997) "Instrumental techniques for linguistic phonetic fieldwork." In *A Handbook of the Phonetic Sciences*, ed by W. Hardcastle and J. Laver. Oxford: Blackwells.

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages.* Blackwells: Oxford.

Wright, R. (1996) *Consonant clusters and cue preservation in Tsou*. Ph.D. thesis. UCLA.

Wright, R., Maddieson, I., Ladefoged, P. and Sands, B. (1995) "A phonetic study of Sandawe clicks." *UCLA Working Papers in Phonetics* 91: 1-24.

The status of phonetic rarities

[To be published in *Language*, 72.4. December 1996]

Peter Ladefoged and Daniel Everett*

Most sounds can be described in terms of a standard set of phonological features, or in terms of values of well known phonetic parameters. But some languages, particularly the smaller endangered languages of the world, also contain many unusual sounds that test our traditional descriptive theories. An example is the dental plosive followed by a bilabial trill, found in the Chapakuran languages. We suggest that there is a set (with fuzzy boundaries) of more common sounds that participate in a wide range of general linguistic processes and another set of rarer sounds that have been observed in only one or two languages.

1. Investigations of little described languages often produce data that help us test our linguistic theories. A good example is the unusual sound in the Chapakuran languages Wari' and Oro Win described by Everett and Kern (1996) as a "voiceless apico-dental plosive [followed by a] voiceless labio-labial trill ...". This sound, which we will generally transcribe [t͡ʙ̥], is rare in two ways. Firstly it has been observed in only a relatively small number of words in the two languages Wari' and Oro Win. Secondly it is like no other sound in the world's languages. Both these facts give rise to questions of theoretical interest. How frequent does a sound have to be before it can be considered as part of the phonological system of a language? Is the occurrence of an unusual sound in a single infrequent word enough, or must it occur in several words, at least some of them being fairly common? How widespread in the world's languages does a sound have to be before we take note of it when devising a universal feature set? This paper will first document the facts as far as the Wari' and Oro Win sounds are concerned, and will then consider these questions.

Wari' is spoken by about 2,000 speakers on the banks of the Pacaas Novos river in Rondonia, Brazil. Oro Win was spoken in an adjacent area, but there are now only five living speakers in Brazil and perhaps a few more in Bolivia (Jean-Pierre Angenot, p.c.). The only other living Chapakuran language is Moré, which is spoken by a single elderly speaker. A count of the occurrence of the sound [t͡ʙ̥] in a Wari' dictionary (Kern, unpublished) shows that it occurs in about two dozen words. In these words it contrasts with other more frequently found sounds, as is exemplified by the words in Table 1. There has been little extensive investigation of Oro Win, but [t͡ʙ̥] probably occurs in a similar number of words, some of which are illustrated in Table 2. This sound has not been recorded in Moré.

Table 1. Words contrasting [t͡ʙ̥] and [t] in Wari'.

t͡ʙ̥ot͡ʙ̥o	'to be pleasant'	**toto**	'to paint'
t͡ʙ̥ot͡ʙ̥oweˀ	'chicken'	**towe**	'to be fat'
t͡ʙ̥owem t͡ʙ̥owem	'dragonfly'		
t͡ʙ̥um	'to be green'	**tom**	'to burn'

Table 2. Words containing [t͡ʙ] in Oro Win.

t͡ʙot͡ʙok inan	'I walk on logs'
kot͡ʙinan	'I start a motor'
kat͡ʙʉ na	'it is an owl'
t͡ʙunt͡ʙʉ na	'it is a helicopter'
t͡ʙum	'small boy'

---

* Daniel Everett is at the Linguistics Department, University of Pittsburgh.

2. The sound [t͡ʙ̥] was first described by Royal Taylor, a missionary linguist working on Wari'. It was mentioned by Key (1975), who had heard Taylor's description, and first described as contrastive by Everett, who had observed it in his own fieldwork on Wari'. The present account is based on data recorded in Guajara Mirim, Brazil, in July 1995. Palatograms and linguagrams were made of two speakers of Wari' saying [t͡ʙ̥um] 'to be green'. Full face and side face videos were made of the same two speakers of Wari' saying all the words in Table 1, and of three of the five speakers of Oro Win saying all the words in Table 2. (We have not heard the two other Oro Win speakers living in Brazil. Angenot (p.c.) reports that there are few more speakers living in a nearby area in Bolivia.) In addition to the video recordings, we also recorded the same words from the three Oro Win speakers and a further 10 Wari' speakers on a DAT recorder, using a close-talking noise-canceling microphone so as to ensure a high signal/noise ratio. The 12 Wari' speakers were recorded as part of a larger project describing the sounds of this language.

Everett and Kern (1996) noted that [t͡ʙ̥] is produced by forming a dental stop, and then allowing the blast of air that occurs when this is released to set the lips in vibration. The palatogram of Wari' [t͡ʙ̥um] 'to be green' reproduced on the left in Figure 1 shows that there was full lingual contact on the upper teeth and the front part of the alveolar ridge. The linguagram on the right shows the part of the blade of the tongue that makes this contact. There is only light contact in the midline near the tip of the tongue, which can be seen more clearly on the original photograph in the area between the dashed white lines. The part of the tongue in the midline further back on the blade did not contact the roof of the mouth. The shaping of the tongue blade and tip forms a narrow jet of air when the closure is released.

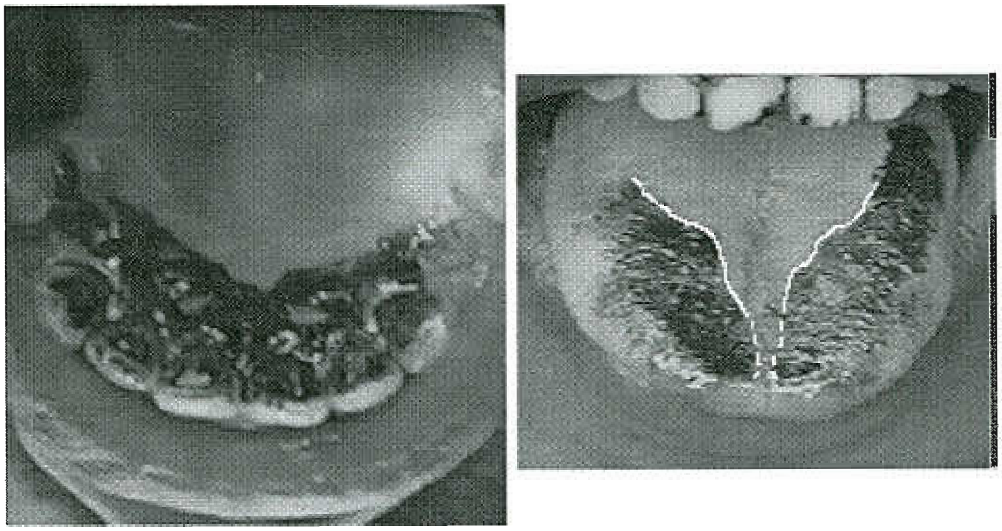
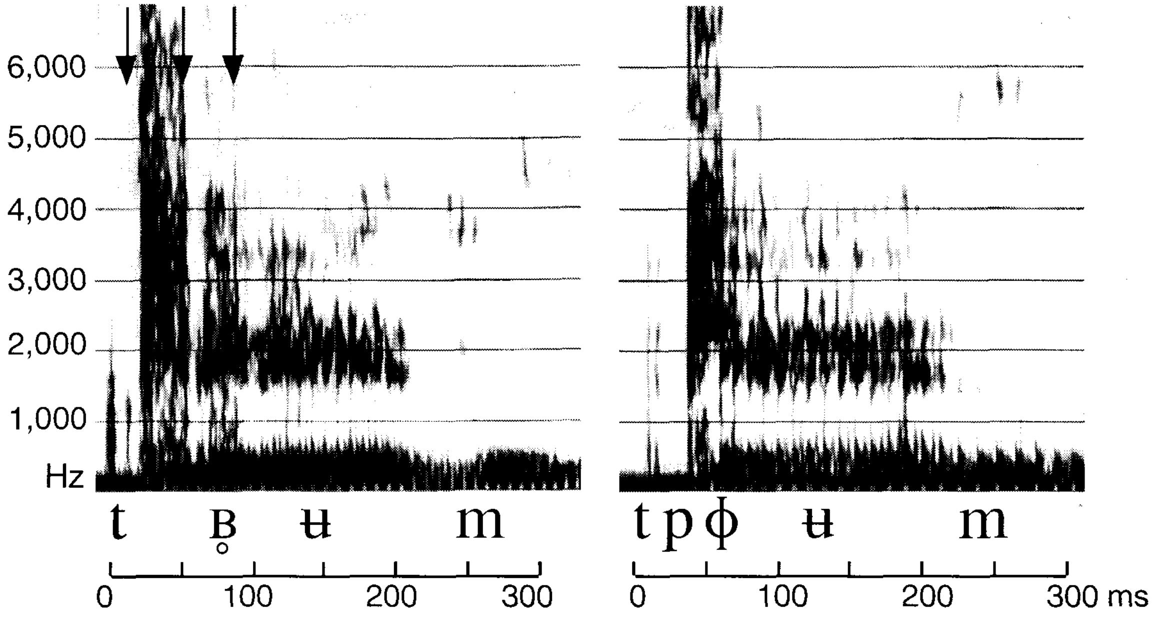


Figure 1. Palatogram and linguagram of Wari' [t͡ʙ̥um] 'to be green'.

Spectrograms were made of the recordings of the five words in Table 1 as pronounced by all 12 Wari' speakers. Two of these words contain two examples of [t͡ʙ̥] , making a total of seven examples within the table. Each speaker said each word twice, but due to odd individual behavior and faulty recording procedures only 156 of the potential 168 occurrences of [t͡ʙ̥] were analyzable. The acoustic consequences of vibrations of the lips were directly observable on only 5 occasions, four of those being in the word [t͡ʙ̥um], the only word containing the high back rounded vowel. The position of the lips required for this vowel is more likely to induce full trills in which the lips form one or more complete closures. An example is given in Figure 2(a). In this token the lips were slightly apart when the [t] closure was released at what has been marked as time 0. They were sucked together at the time marked by the arrow at the top of the picture, and then blown apart, forming a bilabial fricative. By the time of the second arrow at the top of the picture they had been drawn together again so that a complete closure was formed. Almost instantly they were blown apart and sucked together again at the time of the third arrow, and released a little later into a very fricative vowel. There were thus three taps in this trill.



(a) (b)

Figure 2. Spectrograms of the Wari' word [t͡ʙ̥um], 'to be green' (a) pronounced with an observable trill; (b) the more usual form in which the tremors of the lips form a bilabial fricative rather than a trill.

The most common variant of [t͡ʙ] among our 12 speakers was a laminal dental stop followed by what appears on spectrograms as a voiceless bilabial affricate with no vibrations observable in the acoustic record, as illustrated in Figure 2(b). In the video recordings of the two Wari' speakers who were photographed (one of whom was the speaker for both tokens in Figure 2), tremors of the lips are clearly visible in virtually all instances of [t͡ʙ̥] but these lip movements were not substantial enough to form trills of the kind observable in Figure 2(a). In the example in Figure 2(b) the [t] is released and the lips are drawn together to form a voiceless bilabial stop, as is indicated by the narrow phonetic transcription below the spectrogram. This stop is released into a bilabial fricative which is followed by a small drop in intensity, indicating that there might be a second movement in which the lips move slightly toward one another. The degree of voicing varied somewhat, with the fricative portion being fully voiced in three of the 156 tokens, but with the voice onset time being near the start of the vowel in most cases. The duration and intensity of the friction also varied from speaker to speaker, but in all cases it was sufficient to distinguish this sound from the unaspirated stops in [toto].

All three Oro Win speakers produced large oscillatory movements of the lips, forming a clearly distinguishable bilabial trill on nearly every occasion. The mean rates of vibration and other statistics are shown in Table 3. These figures are much in accord with the purely bilabial trills observed in other languages. Ladefoged and Maddieson (1996) point out that:

> It is interesting to note that the bilabial trill releases of prenasalized stops ... have a similar rate of vibration to other trills, despite the fact that the lips have a larger mass than either the uvula or the tongue tip. Ladefoged, Cochran, and Disner (1977) report a mean rate of 29.3 Hz for prenasalized bilabial trills (five speakers), while Maddieson

(1989) reports a mean rate of 24.8 Hz (four speakers). The range is thus very similar to that observed in apical and uvular trills.

Table 3. Mean trill rates for the three Oro Win speakers.

Speaker	Approx. age	number of [t͡ʙ]	% trilled	Trill rate (Hz)	Std. Dev.
Male	55	19	68	29	3.9
Male	35	24	75	26	3.2
Female	35	27	85	23	2.9

It should be noted that the articulation of [t͡ʙ̥] in Chapakuran languages is not the same as that for any previously reported bilabial trill. Summing up the kinds of bilabial trills in the world's languages, Ladefoged and Maddieson (1996) note:

> Apart from a few exceptions which remain unexplained in Nias (Catford 1988), and the special case of Luquan Yi fricative vowels (see Chapter 9), all bilabial trills historically developed from a sequence of a prenasalized bilabial stop followed by a relatively high back rounded vowel, i.e. a sequence such as [mbu]. These segments remain prenasalized and contain a short oral stop phase which is released into a trill that occupies much of the anticipated duration of the following vowel.

The Chapakuran sounds occur only before the rounded vowels [o] and [u]; but there is no suggestion of nasalization or prenasalization. Moreover, the initial laminal dental stop release that sets the lips vibrating is a necessary component of these Chapakuran trills that is not found in any of the other bilabial trills. This is a unique sound involving a voiceless apico-dental plosive generating an airstream that sets the lips moving.

The evidence from our fieldwork indicates that [t͡ʙ̥] is a regularly produced and recognized sound in Chapakuran languages. It has to be taken into account in any phonological description of these languages. It does not have the status of extralinguistic sounds such as the clicks used as interjections in English. A novelist describing the dental click ǀ that is used to express disapproval in English might write 'He went tsk tsk', and hope that readers might read this as [hi wɛnt ǀ ǀ]. But it would not be possible to write 'he was tsk tsking' and expect readers to read this as [hi wəz ǀǀɪŋ]. The sound [t͡ʙ̥] acts like any other single consonant in Wari' and Oro Win. It can occur only before [u] and [o], but it can occur at the beginning of any syllable within a word. It forms regular CV sequences in these languages. There is no phonological reason to regard this sound as anything other than a single segment. This is a language in which there are no consonant clusters.

If [t͡ʙ̥] has the same status as any other consonant in Chapakuran languages, then how do we describe it in terms of features? Of course it has separate elements which can be specified: [t] is [–voice, + coronal, + laminal (or distributed), + stop (or interrupted)], and [ʙ̥] is [– voice, + labial, + trill]. But this does not account for the close relationship between the two elements. If the **t** and the [ʙ̥] are said to occupy a single timing slot, it is not clear that the [t] precedes the [ʙ̥] and that it has a special tongue shape. If we want our phonological descriptions to use a universal feature set and to be phonetically interpretable, then we need some special feature designation for the unique sound [t͡ʙ̥].

All the Chapakuran languages are severely endangered. It seems likely that [t͡ʙ̥] will not occur in any of the world's languages in the latter part of the next century. Of course this sound may develop again elsewhere. Some high elitist group in England may decide that toilets are [t͡ʙ̥at͡ʙ̥a], and the word may catch on and spread to other things. So we can never really tell what features will be needed for describing languages. In principle it is the complete set of human vocal sounds that can be integrated into the flow of speech, and that are sufficiently distinct from one another. This is, however, too cumbersome a notion to be of practical value for working linguists describing languages. More importantly, it also fails to recognize the peripheral nature of some sounds as opposed to others. There is a set (with fuzzy boundaries) of more common sounds that participate in a wide range of general linguistic processes and that are the foundation of many of the observed phonological characteristics of language. These are the sounds that we regularly need to describe in terms of features, or parameters and values as outlined by Ladefoged and Maddieson (1996). There are also the rare sounds that have been observed in only one or two languages, which are not easily describable in terms of values of parameters. The difference between these two loosely defined groups of sounds is one of the fascinating quirks of linguistics that deserves further investigation.

The International Phonetic Association, a group that often seems to stumble on good things by accident, has long had a way of dealing with this problem. In the IPA set of symbols there are sounds that are described in terms of two charts, one for consonants and one for vowels. Each of these charts uses categories that are the equivalent of feature descriptions — plosive, nasal, fricative, bilabial, front, etc. Below the charts there is a set of 'Other symbols' that do not fit neatly into any of the intersections of the major categories. The Chapakuran [t͡ʙ̥] is clearly one of these sounds. We must not overlook it; but we must also acknowledge that it is one of the 'Other symbols', and that it does not fit into the usual set of features or phonetic categories. We can never know what 'other sounds' have been lost in the past; nor can we ever know what other sounds will arise. Only through the close investigation of endangered and less well known languages will we be able to see the present extent of this problem.

References

Catford, J. C. 1988. Notes on the phonetics of Nias. Studies in Austronesian Linguistics, ed. by R. McGinn, pp. 151-172. Athens, Ohio: Ohio University Center for International Studies.

Everett, D, and Kern. B. 1996. Wari': The Pacaas Novos language of Western Brazil. London: Routledge.

Key, M. Ritchie 1975. Paralanguage and kinesics. Metuchen, N.J.: Scarecrow Press.

Ladefoged, P., & Maddieson, I. 1996. Sounds of the World's Languages. Oxford: Blackwells.

Ladefoged, P., Cochran, A., & Disner, S. 1977. Laterals and trills. Journal of the International Phonetic Association, 72, 46-54.

Maddieson, I. 1989. Aerodynamic constraints on sound change: The case of bilabial trills. UCLA WPP, 72, 91-115.

An acoustic study of the tongue root contrast in Degema vowels

Sean A. Fulop

1 Background

1.1 Introduction

Nearly all the languages of the world use degrees of both tongue height and front-back position of the tongue body as contrasting articulatory features of their vowel sounds. These, together with lip rounding, are sometimes referred to as the *major* vowel features (e.g. Ladefoged and Maddieson 1996). Many languages possess vowel inventories which make essential use of features other than these (a common example being nasalization). One of the most perplexing of the *minor* vowel features is Advanced Tongue Root (ATR). This feature has generated considerable interest since its use in a great many African languages was first elucidated in the 1960s. Its many enigmatic properties include the fact that it is largely restricted to the Niger-Congo and Nilo-Saharan language families of Africa, and that it is invariably employed in a phonological process of vowel harmony.

This thesis will discuss many facets of the feature ATR, presenting the results of an acoustic investigation of the Niger-Congo language Degema in support of several points. The results provide evidence for two acoustic correlates of the ATR contrast, namely the frequency of the first formant, and the intensities of the first and second formants relative to each other. The first of these is well established in previous work, but the latter has received limited recognition. We will further demonstrate the correspondence of this second correlate to the theoretically predicted damping of F_1 due to viscous air flow in a narrow pharyngeal constriction.

---

[0] Thanks to Peter Ladefoged for giving me access to precious fieldwork data, and for keeping me in line; to Bruce Hayes for clear direction; to Russ Schuh and Ian Maddieson for reading carefully. Special thanks to Ethelbert Kari, whose linguistic skill and knowledge of Degema made this project possible. Thanks also to Pat Keating for important input, Matt Gordon, Aaron Shryock, Richard Wright, Dan Silverman, and all those in the Phonetics Lab who were willing to chat and give advice.

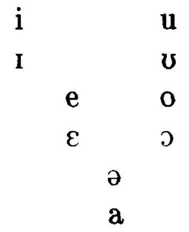


Figure 1: The ten vowels of a complete ATR-harmony system

1.2 African vowel harmony and the discovery of ATR

In some languages the vowel inventory is functionally divided into two subsets along a certain parameter. The phonological word can then become a domain over which the cooccurrence of vowels is restricted; vowels from set #1 may not appear in the same phonological word with vowels from set #2. This scenario is known as *vowel harmony*.

Ladefoged (1968) surveyed the vowel systems of a number of West African languages. He noted that previous work had uncovered the operation of vowel harmony in many of these languages, although the precise nature of this parameter had eluded investigators for many years.

Early attempts to specify the feature differentiating the two vowel varieties often equated the opposition with the so-called "tense/lax" feature of many Germanic languages, as exemplified by the English contrast *deed/did*. The auditory impressions reported by early investigators (e.g. Stewart 1967) indicated that the vowels of the "lax" set generally sound lower in tongue position, giving rise to the use of "tense/lax" terminology and to the transcription in Figure 1 for the 10-vowel systems. Here we use the symbols [i, e, ə, o, u] for the vowels of set #1 and [ɪ, ɛ, a, ɔ, ʊ] for the vowels of set #2. Such a system is reportedly found, for example, in some dialects of Akan (Twi-Fante) (Berry 1957), as well as in Abe (Stewart 1971); these are both Niger-Congo languages.

Akan examples from Stewart (1967) exemplify the vowel harmony:

wʊ́bɛ́nʊ́ḿʔ	you will drink it	wúbénúḿʔ	you will suck it
ɔɔbɛtʊ́ʔ	it (the hen) is going to lay	oobetú	he is going to pull it out
mɪkjɪrɛ́	I show	mitié	I listen

From this it is easy to see why Welmers (1946) analyzed such a system as involving 5 vowels plus a suprasegmental phoneme which altered their qualities (Ladefoged 1968).

Stewart (1967) characterizes vowel harmony in general as a kind of assimilation, wherein all the vowels of a given word agree in their setting of a particular parameter. It is precisely this view which made it clear that these harmony systems did not operate along familiar articulatory parameters. Early analyses often applied what Stewart calls the "Tongue Raising hypothesis," characterizing one set of vowels as being higher than the other. But this isn't true in an absolute sense; the "lax" vowel [ɪ] is usually higher than the "tense" vowel [e]. Even if they were relatively the same in height (as is sometimes the case), it is certainly true that when [ɪ] is brought into harmonic agreement with [e] by changing into [i], any height difference there may have been is actually increased. Stewart points out that since there is no assimilation based on height, tongue height cannot be the basis for the harmony; these facts also led him to call the phenomenon *cross-height* harmony (Stewart 1971).

Stewart then began searching for another articulatory correlate of the contrast. He noted that Hockett (1958, pp. 78–9) described the tense vowels of English as involving tensing of the muscles "above and in front of the glottis within the frame of the lower jaw;" he reports finding this very tensing action in the "tense" vowels of Akan, and refers to the phenomenon as "chin lowering" (Stewart 1967, p. 197). Stewart then refers to Pike's (1947) description of the action of pushing the tongue root forward, which states that the very same "chin lowering" effect results.

Stewart's search for a phonetic feature behind the vowel harmony came to its conclusion when he observed the cineradiographic tracings of Ladefoged (1968), taken of a single Igbo speaker saying four harmonic vowel pairs. Ladefoged commented that the tongue body appeared more retracted in one set of vowels than in the other; Stewart extrapolated this somewhat, and together with Pike's description of tongue root advancing and his own observations of chin lowering in Akan, the x-ray tracings convinced him that the tongue *root* was independently manipulated to facilitate the contrast between the two harmony sets. He also made a leap of faith in applying Ladefoged's evidence from Igbo to the other languages which concerned him at the time, certain as he was that the articulatory mechanism was essentially the same. Stewart came to call the two vowel sets Root-advanced (corresponding to Set #1 in Figure 1) and Root-unadvanced. He used this characterization to explain the variety of auditory and articulatory properties ascribed to these vowels.

Stewart had all along refused to accept the characterization of the Niger-Congo vowel harmony as involving the lax/tense feature of Germanic vowel systems. This was in spite of the considerable influence of prominent scholars; Jakobson and Halle (1962, p. 550) stated that the African languages which display such vowel harmony clearly demonstrate the autonomy of the "tense-lax" distinction. His reservations were certainly shared by other African language experts. In the earlier literature on these harmony systems, comments were often made about a noticeable difference in voice quality between the two vowel sets (e.g. Berry 1955). The consensus reported by Stewart is that the "tense" vowels of Figure 1 typically sound breathy and hollow, while the "lax" vowels sound choked or strangled. The incompatibility between these qualities and the typical tense/lax vowel contrast (as in English) was certainly a major reason for his rejection of the terms.

The tongue height difference, which had in turn led to the misunderstanding that this contrast was of the lax/tense nature, was explained as an epiphenomenon of the tongue root manipulation: the advancing of the tongue root would naturally raise the highest point of the tongue body somewhat. Also, the advancing of the root of the tongue "enlarges the throat cavity and thereby produces the breathy voice quality heard by Berry" (Stewart 1971, p. 199). Stewart's reasoning here is not obvious, but the important point is that he recognizes the possibility for a single linguistic gesture to produce a number of different acoustic reflexes. His new vowel feature soon would be called Advanced Tongue Root.

Stewart's identification of the advancing gesture as primary in general is shown by Painter (1971) to be somewhat arbitrary. Painter indicates that while Akan (studied by Stewart) has the advancing gesture as primary, in Anum, a Nilo-Saharan language with a 10-vowel harmony system, it is the retracting gesture that is primary. Identification of the primary gesture is rather subjective at this point, and depends upon the investigator's impression of which vowel set sounds furthest from the unmarked posture.

While most of the above discussion has concerned languages with a maximal (10-vowel) ATR harmony system, these are in fact relatively rare. The ATR contrast between the low vowels [ə] and [a] seems to be the most unstable. As a result, many ATR-harmony languages do not distinguish two different low vowels. Such languages which have only a single low vowel phoneme but which maintain advanced and retracted mid and high vowels are called *9-vowel* ATR languages. Most dialects of Akan, for example, are of this sort. Additionally, there are languages (e.g. Yoruba) which have collapsed the advanced and retracted high vowels, retaining a partially harmonizing 7-vowel system with only the mid vowels participating.

1.3 [ATR] and [tense/lax]

Stewart (1967) considers and ultimately rejects the equating of ATR with the tense/lax distinction found in many Germanic languages. His intuition was not without supporting evidence, but it nonetheless fell prey to skeptics. Halle and Stevens (1969) proposed that the tense and lax vowels of English are in fact distinguished by the feature ATR, and that the tense/lax terminology should no longer be used to refer to this familiar distinction. They made reference to x-ray tracings of American English vowel articulations which show a difference in pharyngeal aperture between tense and lax high vowels.

1.3.1 Evidence from factor analysis

It has since been demonstrated that finding one aspect of one linguistic contrast to be similar to another contrast is not sufficient evidence for supposing the two to be essentially the same. One must discover whether all important factors of the two contrasts in fact parallel each other before making such a claim.

With respect to vowel articulation mechanisms, a fruitful approach to uncovering the important aspects of a particular contrast is the exploratory statistical procedure known as *factor analysis.* Factor analysis is, in essence, a method for reducing the dimensionality of a multivariate data space to a minimum number of coordinates. These factors are arrived at by choosing a small number of coordinate axes within the data space and testing how well the data fit them. The best factors are then those which give the best fit to the data, using a least-squares criterion. In traditional two-way factor analysis by rotation, there are an infinite number of coordinate sets which will provide an equivalent solution for a given dimensionality, so while the scheme provides best-fitting factors, what those factors are is in some sense arbitrary.

An improvement on the general methodology of factor analysis was proposed by Cattell (1944), called the principle of *parallel proportional profiles.* Briefly, the method involves performing traditional factor analysis on more than one data set from similar populations assumed to show parallel behaviour, but with expected proportional differences along the variables being measured; thus the data sample itself is introduced as an additional coordinate, and the proportional differences of the variable weights across samples provides an important check on the reality of the factors. The fact that the data is well-described by a given set of factors is considered to be less arbitrary since "parallel" data sets are found to conform to the same factors. The intuition is that factors which are mere mathematical artifacts of a data space will not in general be preserved in other data sets where the variable contributions to the factors are altered.

A novel method of factor analysis incorporating this principle has been employed in investigations of vowel articulation mechanisms; English vowels were analyzed in this way by Harshman et al. (1977), who found that two factors were sufficient to describe the difference in sagittal tongue shape between tense and lax vowels, characterized by 18 aperture measurements at points along the length of the vocal tract. A more recent application of factor analysis by parallel proportional profiles was performed by Jackson (1988). He analyzed vowel articulations in several languages, including Akan. His methods were similar to those of Harshman et al., yet a 3-factor solution proved to be a significantly better fit for the Akan ATR distinction than the best 2-factor solution. This result is a significant one, bearing directly on the question of equating ATR with tense/lax. Jackson interpreted his results as showing that an additional factor of tongue-root position was necessary to describe the Akan ATR contrast, while the English tense/lax contrast can be fully characterized in terms of just two parameters of tongue body position.

1.3.2 Evidence from formant values

English tense vowels generally differ in both F_1 and F_2 from their lax counterparts. The contrast is often described in terms of the formant space as one of "centrality," with lax vowels appearing more central and tense vowels more peripheral. The effect on F_2 is assumed to result primarily from the change in front–back position of the constriction for the vowel; lax front vowels are shifted back slightly for a lower F_2, while back vowels are shifted forward for a higher F_2.

In contrast to this, [ATR] harmony counterparts do not show such a difference in F_2 values. [$-$ATR] vowels in general are not centralized, but are usually slightly lower in F_2 than their advanced counterparts. This is corroborated in a number of sources (including this one, below), and is also stated as a generalization by Ladefoged and Maddieson (1996). The upshot of the foregoing discussion is that [ATR] contrasts should not be equated with tense/lax contrasts; while the involvement of the tongue root in tense/lax articulations is similar to that found in [ATR] languages, the evidence from both factor analysis and formant frequency values demonstrates that the two types of vowel contrast must be kept distinct.

1.4 Articulatory and acoustic correlates of ATR

1.4.1 Pharyngeal expansion

The articulatory nature of the Advanced Tongue Root feature was further investigated by Lindau (1974, 1979), whose cineradiography of Akan speakers demonstrated that larynx lowering accompanied tongue root advancing in Akan vowels. This led to Lindau's suggestion that the articulatory feature would be better named Expanded, referring to the apparent fact that the primary gesture involves a change in size of the pharyngeal cavity. The cavity expansion is typically accomplished by both tongue root advancing and larynx lowering.

More recently, a magnetic resonance imaging study of Akan vowels was carried out by Tiede (1993). MRI data are more informative as to the total articulatory behaviour because the images form a series of cross-sections of the vocal tract along its entire length. The total visual information is thus three-dimensional, and is not limited to the sagittal plane as with x-rays. Tiede's data show a positive correlation between tongue root advancing and *transverse* expansion of the pharyngeal cavity, demonstrating that the expansion is not limited to the forward and backward movement of the tongue root. Tiede interprets this as evidence that the medial pharyngeal constrictors are involved directly in the manipulation of pharyngeal size.

1.4.2 First formant frequency shift

The only acoustic correlate of the ATR contrast that has long been established by instrumental measurement is the lowering of F_1 which accompanies tongue root advancing. This was first predicted by Halle and Stevens (1969), and later shown to be the case in Akan by Lindau (1978). This correlate is also evident in the formant plots of Jacobson (1978) for the nine vowels of DhoLuo, a Nilotic language. The occurrence of this frequency shift is not mysterious, and is expected from consideration of a mathematical model of the vocal tract as an acoustic system, which we will now discuss.

One well-developed treatment of the vocal tract as a physical system is one of a class of acoustic models called *lumped-parameter models.* The basic approach is standard in the solution of complex systems involving a number of enclosed cavities, or *tanks,* connected by *constrictions.* One tank with one open-ended constriction is known as a *Helmholtz resonator,* and its resonance frequency is given in Morse (1948, p. 235) as:

$$\nu_0 = \frac{1}{2\pi}\sqrt{\frac{c^2 A}{l_e V}} \quad \text{Hz} \tag{1}$$

where c is the speed of sound, A and l_e are the cross-sectional area and (effective) length of the constriction, and V is the volume of the tank. This expression is derived from the approximation that the air in the constriction is *mass controlled,* acting like a mass on the end of a spring, and the air in the tank is *stiffness controlled,* acting like a spring providing a restorative force. The idea, then, is to derive the equation using the volume velocity and pressure of the air as parameters which in approximation lump together the more complex characteristics of the actual sound field inside the cavities. When only the lowest resonances are of interest, this approach is reasonable (see for example Pierce 1989).

Following this principle, the vocal tract can be modeled as a double Helmholtz resonator. This is done in order to interpret the cavity-formant relations for vowels, and is in fact the source of the theoretical prediction relating the first formant to advanced tongue root. The model consists of two resonators like that described above, connected together by a constriction. This corresponds roughly to a vowel articulation with an approximation dividing the oral cavity from the pharynx. The resonances of each resonator as a separate unit will be determined by equation (1), but the two resonances in the coupled system will shift slightly (see Fant 1960, p. 286 for details). Such a treatment allows the now customary simplification of the relationship between the first two formants and articulatory posture: F_1 is determined by the coupled resonance of the back cavity (posterior to the constriction for the vowel) and, for back round vowels, F_2 is determined by the coupled resonance of the front cavity (anterior to the constriction).

This model is a gross simplification; modeling the front cavity as a Helmholtz resonator is often perceived as particularly inaccurate because there is no constriction at the lips unless there is lip rounding. For this reason, an alternative representation which is nearly as simple uses one Helmholtz resonator for the back cavity, but models the mouth cavity as a simple duct virtually closed by the oral constriction but open at the lips. That issue aside, we can see by using equation (1) as a guide that if the back cavity volume is increased (by advancing the tongue root, for instance) then the associated resonance (i.e. F_1) will drop in frequency. Hence we obtain the common generalization that expanding the pharynx will lower the first formant frequency.

Although many vowel articulations are relatively amenable to this view, a significant portion of the articulatory vowel space should not be viewed in this way. Dunn (1950, p. 746) is careful to point out that, for back vowels in particular, "the dependence of [the first two] formants on the two cavities cannot be separated for any of these vowels." Predictions of the Helmholtz resonator model about the movement of F_1 for these vowels should thus be regarded with caution.

1.4.3 Changes in spectral shape

Impressionistic descriptions of [ATR] vowel systems often include some statement about a difference in the overall sound or timbre of the voice in one set of vowels as against the other. Pike (1967, p. 130; 1947, pp. 21–2), in a general description of tongue root articulations, states that vowels made with an expanded pharynx have a "fuller" or "deeper" resonance, while vowels made with a constricted pharynx sound, not surprisingly, "choked up." The same impressions are given by Ward (1937), who considers the articulatory basis in the case of Abua, a Nigerian language, to be "wide pharynx with breathy somewhat hollow voice," as opposed to "pharyngal constriction." Stewart (1967, p. 199) goes one step further: "It seems, in fact, that breathy voice is the main auditory correlate of root advancing."

Jacobson (1978) considers the vowel systems of Western Nilotic languages as a group, about which Tucker (1966, p. 402) states: "An outstanding characteristic of these languages is the presence of both 'hard' and 'breathy' (or 'hollow') Voice Quality in the pronunciation of vowels, diphthongs, and semi-vowels." In a footnote, Tucker defines breathy as "pronounced with open pharynx, accompanied by a voice aspiration" and says that "non-breathy vowels are pronounced with varying degrees of pharyngal constriction" (Jacobson 1978, pp. 2–3). Tucker further states: "In S. Lwo there is a simple dichotomy of five 'hard' vowels against five 'hollow' vowels, in which the categories are distinguished by both tongue position and voice quality" (Jacobson 1978, pp. 2–3).

In any case, some manipulation of the vocal apparatus gives rise to a timbral change in a number of [ATR] languages. How this occurs is a matter of some debate, and the vocal tract models currently in use cannot be used to address the problem adequately. Indeed, a persistent problem exists in determining the precise nature of the actual spectral difference and measuring it, much less predicting it.

Measuring spectral shape It is often written, as mentioned above, that [+ATR] vowels sound breathy as against their retracted counterparts. One measure of breathy phonation that has been applied in the past (Ladefoged et al. 1988) is the difference in amplitude between F_0, the fundamental frequency, and H_2, the first harmonic above the fundamental (generally called the "second harmonic"). With breathy voicing, H_2 generally shows a markedly decreased amplitude in relation to F_0. Hess (1992) found no consistent behaviour of this measure in Akan vowels; there is no evidence that this correlates with [ATR] value at all in this language. Thus the prediction of Halle and Stevens (1969) that [+ATR] vowels would be articulated with a more lax or breathy phonation type due to the lowering of the larynx is not substantiated by Hess's results.

Another measure of spectral characteristics, correlated with the degree of vocal tract damping, is the bandwidth of each formant resonance. *Damping* refers to the acoustic impedance properties of the vocal tract that determine the degree to which the energy of the voice is dissipated. There is no general solution to the question of how each particular formant bandwidth relates to what kind of vocal tract damping, but the bandwidth of F_1 is usually the easiest to measure. This measure was found by Hess (1992) to correlate well with the [ATR] contrast; Akan [+ATR] vowels were found to have narrower F_1 bandwidths than [−ATR] vowels. Hess compared the [±ATR] pairs [e, ɪ] and [o, ʊ], which are not corresponding vowels in the harmony system. She chose to compare these pairs because they are extremely close in the formant frequency space, and so the results bear on the question of how they could be distinguished from each other, given that formant frequency values cannot be used.

Similar characteristics are reported by Ladefoged and Maddieson (1996) for Degema; these authors compared directly corresponding high vowels [i] and [ɪ]; their largely anecdotal evidence shows that not only are the formant bandwidths of the advanced vowel narrower, but also that the overall high frequency energy appears to be greater. These authors also describe their auditory impression of the Degema vowels, stating that the [+ATR] vowels sound "brighter because of the greater amount of energy in the higher part of the spectrum" (Ladefoged and Maddieson 1996, p. 301). They further speculate as to the mechanism behind the apparent difference in spectral balance, supposing that the [+ATR] vowels have narrower formant bandwidths as a result of "greater tension of the vocal tract walls and fewer acoustic losses in the region of the resonances."

These reports are surprising, in that they fly in the face of considerable earlier literature already discussed. Although none of the early reports include direct acoustic measurement, in light of the results just mentioned it is difficult to understand how so many experienced phoneticians could have come to describe the voice quality of retracted tongue root vowels as "harsh" or "strangled" as against the "dull" and "hollow" sound of the advanced vowels. Ordinarily, descriptions such as these have been assumed to correspond to precisely the opposite spectral attributes reported by both Hess and Ladefoged and Maddieson.

To find a way out of this morass, we must try to find out more precisely how the many measurable parameters of spectral shape are expected to correlate with articulatory behaviour. Unfortunately the literature in this area is often impressionistic and unilluminating; a considerable amount of this literature is reviewed by Laver (1980). He concludes that tense sounding voice quality is well correlated with decreased damping of the vocal tract, while lax sounding voice quality is correlated with increased damping of the vocal tract. As the current conflict of results regarding the spectral attributes of [ATR] vowels is impossible to resolve through such simple statements, we must consider the details underlying the generalization.

What makes a voice sound tense? We can characterize generally the different extremes of vocal timbre by using the terms *tense sounding voice* and *lax sounding voice*, in the hopes of being similar to but clearer than the terms *tense voice* and *lax voice* as employed by Laver

(1980). Laver's notions of tense and lax sounding voice make reference to general settings of tension throughout the vocal tract, and are not limited to phonation settings. According to Laver (1980, pp. 142–3), "tense and lax voices seem to be acoustically differentiated chiefly by the relative amounts of energy in the upper harmonics (van Dusen 1941), with tense voice having stronger upper harmonics than lax voice." In accord with this, Laver suggests that an important factor associated with the overall tension setting is the vocal tract damping. This kind of acoustic correlate (whether related to vocal tract damping or phonation manner) has often been called *spectral tilt,* in reference to the average rate of decline of the harmonic and/or formant amplitudes as frequency increases.

Measures of spectral tilt have most often been designed around the detection of differences in phonation type (breathy voice, creaky voice), as discussed in Ladefoged et al. (1988). In many ATR languages, however, the articulatory correlate(s) of the advanced/retracted tongue root distinction seem to be essentially pharyngeal; there is often little a priori reason to think that a phonation type distinction is involved.

According to past results on [ATR] distinctions the main *visible* articulatory parameter is pharyngeal cavity size; advanced tongue root vowels have an expanded pharynx and retracted tongue root vowels have a constricted pharynx. While tongue root retraction and larynx raising have both been observed to contribute to this, pharyngeal constriction can also be carried out by *isotonic* contraction of the superior, medial, and inferior constrictor muscles (Hardcastle 1976). Evidence for this action is provided by Tiede's (1993) MRI study, discussed above.

Hardcastle additionally suggests that the pharyngeal constrictors could perform *isometric* contraction, thereby increasing the muscular tension without further shortening them. This condition would arise if various muscles were employed in opposition to each other, holding the pharyngeal aperture constant by their mutual antagonism. Possible acoustic correlates of such tension have been mentioned by Kaplan (1960, p. 199):

> The texture, as well as the size and shape, of the pharynx and its apertures affects speech quality. A hard-surfaced resonator emphasizes the higher partials, or overtones, so that a pharynx tightly constricted by its muscles, takes on a metallic, strident, and tense tone. On the contrary, a soft surface, provided by relaxed throat muscles, increases the responsive range while damping the resonator. This in effect gives relative prominence to the fundamental and lower partials.

So then, the acoustic correlates of pharyngeal tension may be similar to those of the laryngeal or glottal tension which occurs in tense and creaky phonation. It seems reasonable that isometric tension in the pharyngeal constrictors would most likely appear in retracted tongue root vowels, since the pharyngeal muscles are already contracted in order to shrink the cavity.

Van den Berg (1955) concludes that pharyngeal tension is most strongly correlated with decreased formant damping in the low frequency range (below 1000 Hz). Insofar as F_1

bandwidth is a correlate of damping in this frequency range, this runs counter to the results of Hess (1992) which demonstrate that [+ATR] vowels show symptoms of reduced F_1 damping, and so should sound tense. Most investigators (with the notable exception of Ladefoged and Maddieson 1996) agree that in fact it is [−ATR] vowels which sound tense.

Perhaps, then, our understanding of the acoustic correlates of tension is not so clear as the literature discussed by Laver (1980) would have us believe. The conflict described here seems to shed doubt on the widely accepted notion that formant bandwidths are consistent indicators of vocal tract damping. But this is not tenable, as the fact that formant bandwidths are directly related to vocal tract damping is derived from fairly rigorous consideration of the vocal tract as an acoustic system (e.g. Fant 1960). Although these mathematical treatments can only provide approximate answers, it is unlikely that this conclusion is mistaken. More likely, then, is the possibility that we do not fully understand what makes a voice *sound* tense or lax, a perceptual factor that has not been fully investigated.

Perception of spectral tilt In choosing a measure of spectral shape which is to be used to draw linguistically relevant conclusions, one must take care to measure something which is directly perceived by language users, or which acts as a reliable index of something perceived. Many measures of spectral content are dubious in this respect. Two measures have already been discussed; another that has been established in previous research is the relative amplitude of a formant as against the fundamental frequency. Ladefoged (1983) measures the intensity difference between F_1 and F_0 to quantify differences in phonation type. Fulop (1994) measures the intensity difference between a higher formant (usually F_2) and F_0 to quantify overall tension differences in the vowels following fortis and lenis plosives. These measures involve comparing the intensity of F_0 with a higher spectral element; the relative intensity of the fundamental is well-established as a correlate of phonation type, but it is not clear that this factor is involved in [ATR] systems. It is also not clear that a difference along one of these measures guarantees a perceptible difference in phonation type. In other words, if a difference in these measures is found, we cannot be certain that this indicates a perceptible difference in the speech sound which can be used in a linguistic context.

One measure which seems to never have been employed in a linguistic investigation is the intensities of the formants relative to each other; i.e. the amplitude difference between two formants. Assman and Neary (1987, p. 520) report the finding of Kakusho et al. (1971) that "listeners can discriminate very small changes in the intensities of individual harmonics near the formant peaks of vowel sounds, but are poor at discriminating intensity changes in the spectral valleys." This result demonstrates that a listener's notion of "spectral shape" is different from that provided by a Fast Fourier Transform on our computers, which discriminate spectral differences equally well across the entire frequency range under study. The result also implies that the intensities of formants would be a good index of a listener's discrimination of spectral shapes; two differently shaped spectra are likely to be perceived by listeners as identical vowel sounds so long as the (relative) intensities of the harmonics within the formant bands are the same. In other words, the spectra may differ only in the "valleys" between formants, and the listener is unlikely to notice.

We thus make way for the suggestion that the relative formant amplitudes are a perceptually relevant measure of spectral tilt. A remaining question is whether spectral tilt itself is in any sense perceptually salient. Here, the results of Kohler and van Dommelen (1987) are relevant; in their perceptual experiment listeners were biased towards identifying German plosives as fortis (rather than lenis) when the plosive was presented within a "tense" voice quality frame showing "a less prominent first spectral peak in relation to the higher-order peaks and/or a less steep spectral tilt (p. 367)." Thus it appears that listeners do notice the difference between different amounts of spectral tilt in a speech signal, and that they probably make crucial reference to the relative formant intensities for this purpose. In the acoustic study which follows, the relative intensities of the first two formants are used as a measure of spectral shape for Degema vowels.

Comparing spectral shape measurements among vowels of different quality Recall that Hess (1992) compared [±ATR] vowel pairs which do not correspond to the supposed phonological vowel pairs; i.e. she compared the vowels [e] and [ɪ] rather than [i] and [ɪ]. This is because the phonological [±ATR] pairs are generally in different locations in the vowel formant space, [+ATR] vowels having a lower F_1. The pairs which were compared by Hess are more often found to acoustically overlap within the formant space, and so are of more similar "vowel quality" in the traditional sense of the term.

When formant amplitude and/or bandwidth are used as spectral shape measures, it is important to account for the fact (shown by Fant 1960) that these values are correlated with (and are at least grossly predictable from) the frequency of a formant, owing to the properties of the human vocal tract mechanism. Hess avoided the resulting difficulties by comparing the bandwidths of vowels with similar formant frequencies. If an investigator desires to compare spectral attributes among vowels of different formant frequencies (as would be the case in a comparison of phonologically corresponding [±ATR] vowels), s/he must demonstrate whether any evident spectral shape differences can be accounted for by their correlation with formant frequency differences.

2 Acoustics of ATR in Degema

2.1 Degema

Degema is a Niger-Congo language, more specifically classified as Edoid. The Degema are one of three groups of Edoid-speaking people in Nigeria's Rivers State; the three languages spoken in this small region comprise the Delta-Edoid subgroup. The two dialects of Degema are Usokun and Atala. These are not substantially disparate, showing only small differences in lexical tones and in the phonetic realization of one consonant (Elugbe 1989). Degema exhibits a 10 vowel harmony system involving the two mutually exclusive sets discussed above in more general terms:

[+ATR]	i, e, ə, o, u
[−ATR]	ɪ, ɛ, a, ɔ, ʊ

In a given word, all the vowels will be of one set or the other.

2.2 Data and methods

The speech of six male Degema speakers was recorded on a portable DAT recorder (thanks to the field work of P. Ladefoged and E. Kari). The utterances are one and two-word sentences. Two of the speakers provided eight tokens of each vowel, while the remaining speakers provided four tokens of each vowel. The vowels occurred in minimal sets (verb paradigms) between [m] and an alveolar stop (there are the same number with following [t] as with following [d]).

To analyze the data, FFT spectrograms were computed using a 150 Hz bandwidth filter and a Hamming window. LPC formant histories were superposed on these, computed with a filter order of 12 over a 10 ms window. This information was used to estimate the first three formant frequencies.

Additionally, the relative amplitudes of the formants were estimated using a narrow-band short-time FFT. The formant amplitudes in dB were measured as the energy in the most prominent harmonic clearly within each formant band. This technique is by nature approximate; the "true" amplitude can only be discovered this way if a harmonic occurs at the center frequency of the resonance. Otherwise, the harmonic amplitudes will be slightly less.

2.3 Results

2.3.1 F_1 value

Figures 2 and 3 show the vowels for each speaker plotted in an acoustic vowel space. The ellipses are computed to enclose two standard deviations of the token dispersion along the major and minor axes. Dark ellipses are used for [+ATR] vowels, and light grey for [−ATR] vowels. A few vowel groups were found by boxplot analysis to have probable outliers; the offending tokens have been removed from the plots to give tighter ellipses that are more accurately located.

As expected, the difference between the [+ATR] and [−ATR] vowels is often evident in the value of F_1; the retracted tongue root vowels frequently show a distinctly higher F_1

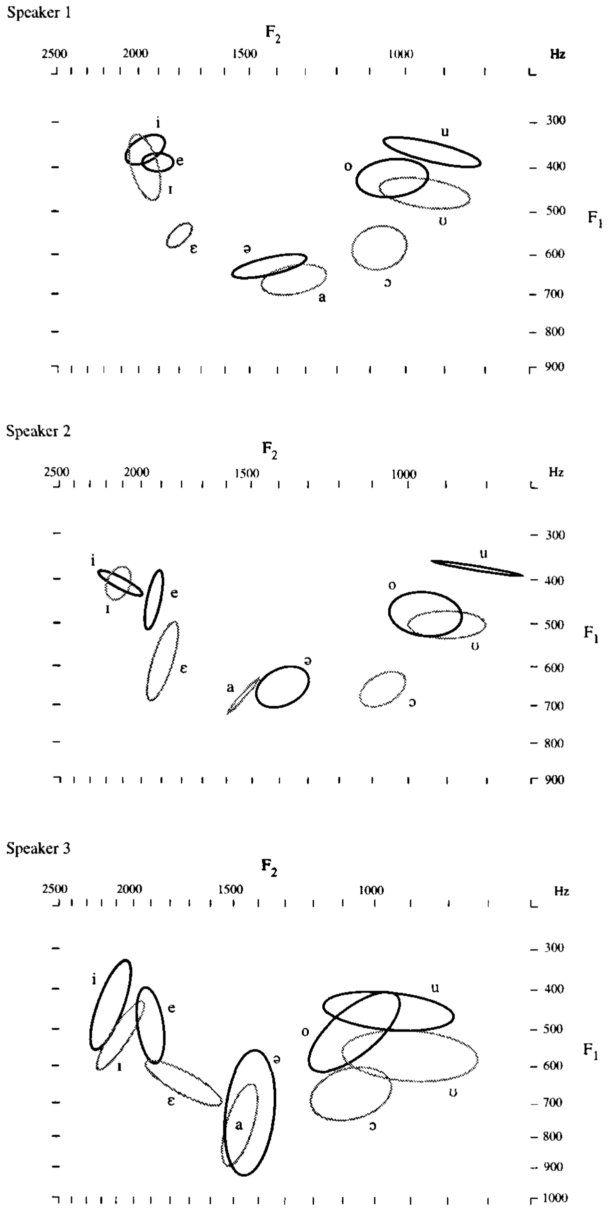


Figure 2: Conventional F_1–F_2 vowel formant plots for speakers 1–3

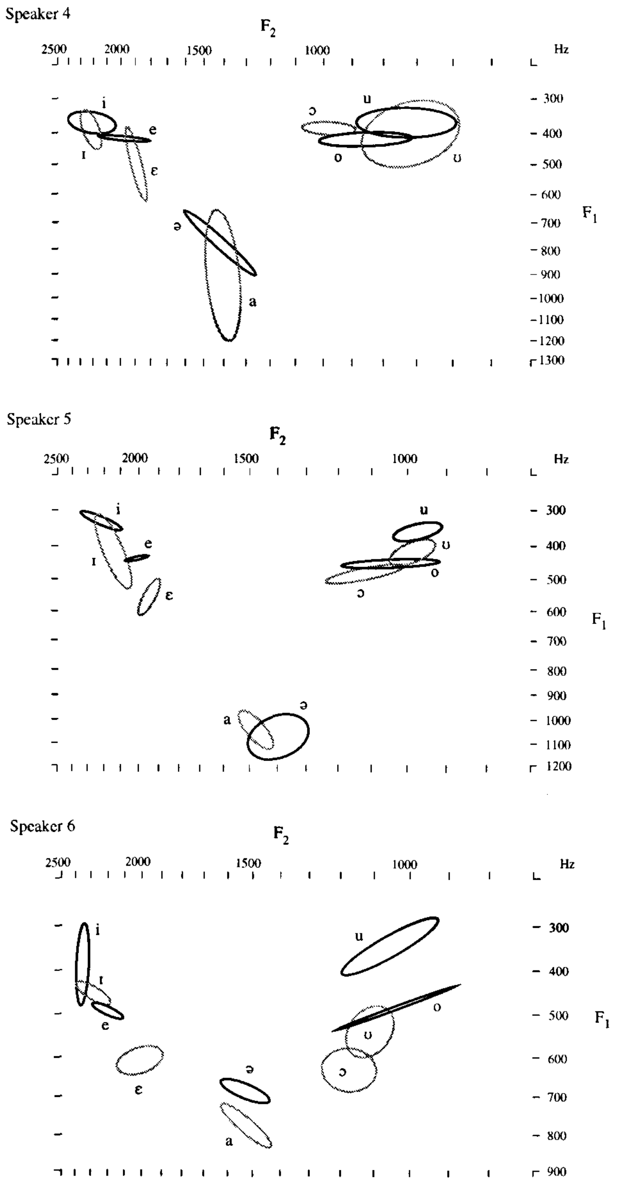


Figure 3: Conventional F_1–F_2 vowel formant plots for speakers 4–6

Table 1: F_1 difference: Statistical significance and effect sizes

	Vowel quality	a	e	i	o	u
Speaker 1	FP significance	3.4×10^{-5}	0	0.0012	0	0
	Effect size	0.89	1.0	0.85	1.0	1.0
Speaker 2	FP significance	0.49	0	0.90	0	4.8×10^{-7}
	Effect size	0.66	1.0	0.47	1.0	1.0
Speaker 3	FP significance	0.46	0	0.0013	0	0
	Effect size	0.62	1.0	0.84	1.0	1.0
Speaker 4	FP significance	0.23	0.23	0.37	0	0.50
	Effect size	0.75	0.78	0.72	0	0.66
Speaker 5	FP significance	0.30	4.8×10^{-7}	0.012	9.2×10^{-6}	0
	Effect size	0.28	1.0	0.88	0.97	1.0
Speaker 6	FP significance	0	0	7.4×10^{-6}	0	0
	Effect size	1.0	1.0	0.94	1.0	1.0

than their advanced tongue root counterparts. This is not always true of the contrastive low vowels [ə] and [a], nor of the high vowels [i] and [ɪ]. It is also not a robust effect in the vowel space of speaker #4. In fact, it is not obvious that speaker #4 distinguishes his vowels this way at all.

Table 1 gives the significance of the difference between the first formant frequency of the [+ATR] as against the [−ATR] vowels. The significance values are computed by two-sided Fligner-Policello tests on each pair of vowels; the reasons for using this procedure will be discussed in connection with further results below. The effect size gives us an indication of the degree to which the two vowels differ in their F_1 values, measured as the probability p that a randomly selected [+ATR] token will have a lower F_1 than a randomly selected [−ATR] token. For example, when the [+ATR] and [−ATR] vowels of each pair are perfectly distinguishable in this way and [+ATR] shows a lower F_1 than [−ATR] this probability has the value 1.0. If the two types of vowels are not distinguishable in this way (as when their ellipses in Figures 2 and 3 overlap), this probability has a value closer to 0.5. The one case where $p = 0$ results from the vowels being perfectly distinguishable, but with the expected direction of the difference reversed.

2.3.2 Relative formant amplitude

The bar chart in Figure 4 shows the grand mean difference $A_1 - A_2$ between the first two formant amplitudes for the six speakers. The relative intensity value in each case has been averaged across the five basic vowel qualities. For each speaker, the bar representing $A_1 - A_2$ is smaller for the [−ATR] vowels, indicating that F_2 has a relatively greater intensity.

So then, the second formant of the retracted tongue root vowels is in general a stronger

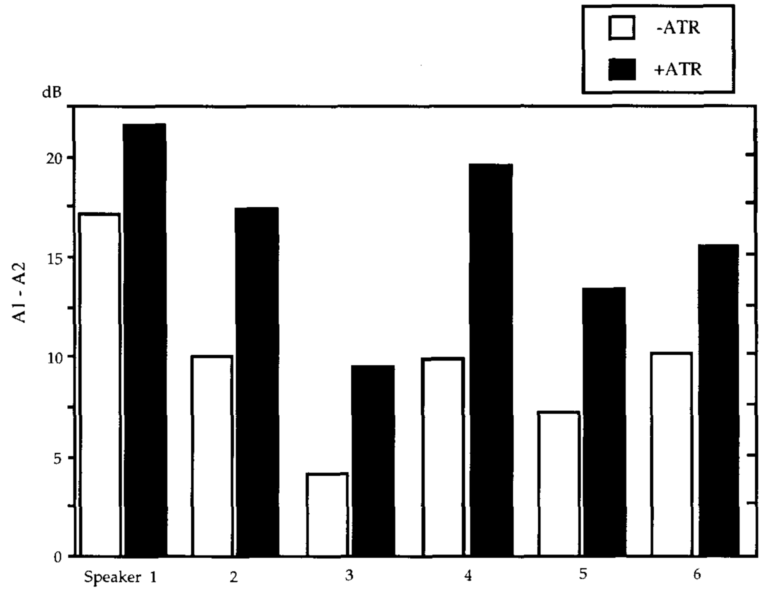


Figure 4: $A_1 - A_2$ for [±ATR] vowels, six speakers

contributor (in relation to F_1) to the spectral shape, with some statistical caveats discussed below. This may contribute to their having an audibly brighter timbre, as was discussed above. On listening to the Degema ATR contrast, it is my impression that the retracted vowels betray a distinctly brighter sound than their advanced counterparts. We will discuss below the probable relation between this acoustic correlate and the articulation of these vowels.

The results shown in Figure 4 present an overall picture of the formant intensity effect, but statistical analysis is required. This must proceed with two goals in mind. Firstly, we should establish the degree of significance of the effect; i.e. given a null hypothesis H_0 that there is no difference in relative formant intensity between [+ATR] and [−ATR] vowels, what is the chance of being mistaken in rejecting this and accepting the idea suggested by Figure 4? Rejecting H_0 when H_0 is in fact true is known as a *Type I error.* In essence we are seeking evidence of a difference between two groups of vowels, the [+ATR] as against the [−ATR].

We can seek such evidence through simple statistical means only so long as the distribution of measurements in each group does not depend on other factors. The groups compared in Figure 4 have been split by speaker (an obvious necessity), but still do not meet this requirement; each average measurement encompasses five different vowel pairs. Because formant intensities change with frequency, we have reason to expect that each vowel quality will have a different baseline formant intensity difference. This will cause each group to contain five clusters of data points. We must therefore split each group into five smaller groups of measurements according to basic vowel quality and compare phonologically corresponding vowels within each category.

Statistical analysis The challenge at this point is to make the most of statistically meager data; speakers 1 and 3 provided eight tokens of each vowel and the remainder provided only four tokens. While the best-known method for comparing two groups is Student's t-test, applying it here would not be the best choice for a number of reasons.[1] Student's t-test assumes a normal distribution of the data and compares the means of the two groups. This procedure is not reliable when there are departures from normality (which cannot be detected with small sample sizes) or with skewed distributions (as are present in this data) since the mean is not a resistant measure of location (i.e. it can be thrown off by a single datum). Two difficulties can plague t-test results in a case like this. The first is loss of control over Type I errors; in other words, false significance of small differences. The second is loss of *power* (control over Type II errors). When power is lost, we can obtain limited significance and be led to accept the null hypothesis, when in fact our alternative hypothesis is the correct one. The statistical analysis below shows evidence that both of these problems affected t-test results for the formant intensity data.

As an alternative to the t-test it is often prudent to consider using a measure of location other than the mean when two groups are being compared. The median, for example, is much more resistant; it is unaffected by a change in as much as 50% of the data. A problem still arises with small samples, however; it turns out that tests which explicitly compare measures of location require a sample size of at least 15 in order to confidently control Type I errors and power at 5% significance levels (Wilcox 1996). Of particular interest as an alternative here are methods of comparison based on ranks. The Mann-Whitney-Wilcoxon test (Mann and Whitney 1947), for example, allows us to test the hypothesis that two groups have identical distributions without relying on a direct estimate of some measure of location. The test is theoretically sensitive to differences between the medians, but no estimate of the median is used (Wilcox 1996). It also allows the computation of an equally important property revealed by the data, the *effect size*.

Let p be the probability that an observation randomly sampled from the first group is less than a randomly sampled observation from the second group. If two groups have identical distributions, $p = 1/2$; the Mann-Whitney-Wilcoxon procedure tests an alternative against the null hypothesis that $p = 1/2$. One difficulty with the procedure is that it assumes the two groups have equal shapes and therefore equal variances, a condition that is not met by our data. As a result, we will employ a related but improved procedure that eliminates this requirement, the Fligner-Policello test (Fligner and Policello II 1981). Again, the null

---

[1] Another procedure which is often employed is the Analysis of Variance (ANOVA). This applies an F-test to the problem of comparing the means of two or more groups. Since it actually contains Student's t-test as a special case (viz. the case of two groups, Wilcox 1996), there is never any reason to choose this method over the simpler t-test when two groups are being compared.

hypothesis is $H_0 : p = 1/2$. As p approaches 0 or 1, the two-tailed procedure is more likely to show a significant difference in distributions. According to Wilcox (1996), the Mann-Whitney-Wilcoxon test (and by extension the Fligner-Policello test) has much higher power than methods for comparing means like the t-test. The test was carried out using Minitab statistics software, and macro procedures written by Wilcox.

The discussion above and the decision to use the Fligner-Policello test has so far resolved only one of the two issues arising from the formant intensity data. We have determined values that indicate the significance of the effect, but this is meaningless without some indication of the importance of the effect (i.e. its size) and a corresponding interpretation of the significance. The value that is often used to report significance tells us the chance of being mistaken in accepting our alternative hypothesis. This a good indication of our confidence of direction in which the two groups are different. It is important to actually report this value, rather than just saying "the results are/are not significant."

Tukey (1991) argues strongly for a more comprehensive approach to significance results; he advocates taking hints from inconclusive data and reporting confidence intervals for all values. The Fligner-Policello test allows us to take Tukey's advice; we are able to report significance results and draw conclusions from them as such, and we are also provided with an excellent measure of the size of the effect, the value of p that is used as a test statistic. In fact, Cliff (1993) argues for this statistic as an excellent measure of effect size in general.

Table 2 shows the results of statistical analysis of 30 groups of [±ATR] vowel pairs (six speakers, five basic vowel qualities). The significance of one-tailed t-tests ($H_1 : \mu_1 < \mu_2$) and two-tailed FP-tests are reported, as well as the probability p that an observation from group 1 is less than an observation from group 2. The alternative hypothesis being considered in each case is that the [−ATR] vowels have a $A_1 - A_2$ value less than that of the [+ATR] vowels, as suggested by Figure 4. The FP significance values should be considered as the most accurate indication of our confidence in a difference; given the small sample sizes involved, we submit that any value around 0.1 or less should be taken seriously. This leaves a 10% chance of a Type I error; the usual 5% tolerance is extended because tremendous loss of power may result otherwise, i.e. we could run an unacceptable risk of rejecting results that in fact vindicate our alternative hypothesis. The figures show that the data are quite varied, although the situation looks fairly neat in Figure 4. There are also some cases where the t-test and FP-test tell different stories.

The reader should bear in mind that the two measures of significance are different; t-test significance gives the probability that it is mistaken to say "the mean of the [−ATR] vowels is less than the mean of the [+ATR] vowels." FP-test significance gives the probability that it is mistaken to say "the two distributions differ, and in particular the median of the [−ATR] vowels is less than the median of the [+ATR] vowels." The effect size gives us some idea of how important the difference is. If it is nearly 1, the [−ATR] group is almost entirely less than the [+ATR]; if it is nearly 0.5, the difference between the two groups is negligible even if significant ("significant" does not have its everyday meaning, and simply

Table 2: Statistical significance and effect sizes

	Vowel quality	a	e	i	o	u
Speaker 1	T-test significance	0.0085	0.0048	0.30	0	0.041
	FP significance	0.0011	0.011	0.63	0	0.092
	Effect size	0.86	0.81	0.57	1.00	0.73
Speaker 2	T-test significance	0.30	0.011	0.56	0.0040	0.0008
	FP significance	0.81	$*$	1.00	$*$	$*$
	Effect size	0.56	$*$	0.50	$*$	$*$
Speaker 3	T-test significance	0.47	0.0005	0.090	0.0099	0.020
	FP significance	0.92	$*$	0.26	0.0064	0.0047
	Effect size	0.52	$*$	0.67	0.81	0.84
Speaker 4	T-test significance	0.11	0.0003	0.16	0.10	0.0099
	FP significance	0.076	$*$	0.11	0.016	0
	Effect size	0.81	$*$	0.81	0.88	1.0
Speaker 5	T-test significance	0.76	0.11	0.031	0.013	0.030
	FP significance	0.23	0.61	7.3×10^{-6}	$*$	0.012
	Effect size	0.25	0.63	0.94	$*$	0.88
Speaker 6	T-test significance	0.0001	0.69	0.52	0.0023	0.015
	FP significance	0	0.44	0.80	0	0
	Effect size	1.0	0.31	0.56	1.0	1.0

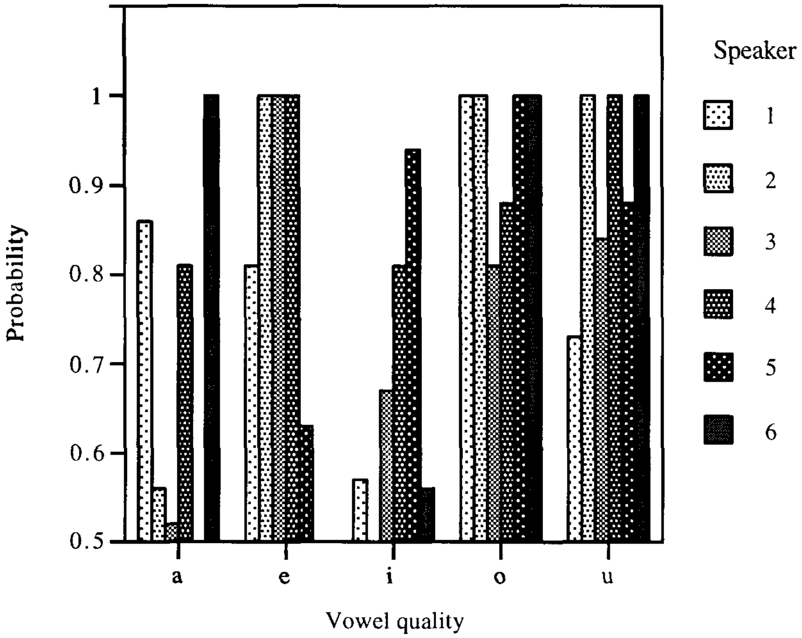


Figure 5: Size of the formant amplitude effect, measured as the probability p that a random observation from the [$-$ATR] tokens has a smaller $A_1 - A_2$ than a random observation from the [$+$ATR] tokens.[3]

indicates a certain degree of confidence in the difference); if it is nearly 0, we must conclude that the [$-$ATR] group is in fact greater than the [$+$ATR] group. The asterisks in the table mark points where the FP-test calculations returned a zero-value error, the result of the two groups being entirely different. The effect size in these cases can be assumed to equal 1, with the significance (chance of error) near 0. We suggest that, for the small samples involved, an effect size in the neighbourhood of 0.8 should be taken seriously.

The effect sizes from Table 2 have been graphed in Figure 5. This shows the statistical variability of the amplitude effect across both vowel quality and speaker. If the effect were manifested consistently in all cases, every bar in the graph would reach the top of the y-axis where $p = 1$.

We chose to perform the Fligner-Policello test because statistical research has shown there to be differences in performance which, under certain conditions, can save the experimenter from costly errors of both type I and type II (Wilcox 1996). Examples of this are evident in Table 2. For speaker 3, vowel quality [i], the t-test could have led us to accept our alternative hypothesis that the [$-$ATR] and [$+$ATR] vowels are distributed differently; the FP-test shows that there is a 26% chance that this would be mistaken (a type I error), and that the effect size is unacceptably small in any event. For some of the data from speaker 4, t-tests show

Table 3: F_1 frequency shift and $A_1 - A_2$ decrease

	Vowel quality	a	e	i	o	u
Speaker 1	F_1 increase (octaves)	0.072	0.55	0.11	0.47	0.31
	$A_1 - A_2$ decrease (dB)	2.9	7	2.2	8.9	3.6
Speaker 2	F_1 increase	0.026	0.41	−0.025	0.47	0.48
	$A_1 - A_2$ decrease	2.4	10.4	2.6	13.2	11.2
Speaker 3	F_1 increase	0.029	0.40	0.22	0.44	0.35
	$A_1 - A_2$ decrease	−1.3	9.3	3.9	4.2	4.3
Speaker 4	F_1 increase	0.20	0.26	0.13	−0.097	0.075
	$A_1 - A_2$ decrease	2.3	18.8	5.1	14.5	11.8
Speaker 5	F_1 increase	−0.047	0.38	0.42	0.11	0.23
	$A_1 - A_2$ decrease	−2.3	3.3	12.8	10.7	5.7
Speaker 6	F_1 increase	0.20	0.30	0.19	0.35	0.73
	$A_1 - A_2$ decrease	12.3	−5.6	1.7	11.2	5.5

only modest or poor significance levels while the FP-tests show better significance with good effect sizes; here the danger of several type II errors (wrongly rejecting a correct hypothesized difference) has been averted.

Correlation with formant frequency shift It was mentioned above that a formant's amplitude is correlated with its frequency. In fact, according to Fant (1960), we can expect the intensity of F_1 (for example) to decrease 12 dB for each octave of frequency increase. Within each [±ATR] vowel pair in Degema, our formant plots (Figures 2 and 3) show that F_1 is generally larger for the [−ATR] vowel, while F_2 remains fairly steady. This must partly account for the further result that F_1 so often contributes less to the intensity profile of the spectrum (i.e. that $A_1 - A_2$ is smaller for the [−ATR] vowels). A remaining question is, does the shift in frequency of F_1 *totally* account for the difference in $A_1 - A_2$, and thereby negate the importance of the results?

The calculated figures in Table 3 help to answer this question. Given that shifts in F_2 are small fractions of an octave, a perfect correlation between the F_1 increase and $A_1 - A_2$ decrease in [−ATR] vowels would be betrayed by a correspondence between the proportion of an octave covered by the F_1 shift and the proportion of $\sim$ 12 dB covered by the $A_1 - A_2$ shift. Table 3 gives the values whose correlation is sought, showing the medians for each speaker and basic vowel quality.

Table 4 shows that, as anticipated, the values are not totally uncorrelated. Since, a priori, a correlation between the shift in F_1 and the change in relative amplitude $A_1 - A_2$ is expected to be linear if present, the degree of correlation can be estimated using the standard correlation coefficient r:[4]

Table 4: Correlation coefficients for Table 3

Speaker	r
1	0.882
2	0.984
3	0.808
4	−0.128
5	0.565
6	0.024

Speakers #1, 2, and 3 show high correlation, with r close to 1 in each case. Thus a case might be made that, for these speakers, the shift in F_1 accounts completely for the change in the relative amplitude of the formants. One task then remaining would be the explanation of the much greater size of the effect than Fant's acoustic theory would predict; Speaker #2, for example, shows amplitude changes on the order of 12 dB with F_1 shifting only 1/2 octave or so. But Fant's calculations lead us to expect an amplitude change of only 6 dB in these cases.

The notion that there is no independent amplitude phenomenon here is also cast in doubt by the results for speakers #4, 5, and 6, who show no clear correlation between F_1 frequency shift and amplitude shift. This demonstrates that while the observed formant intensity change from [+ATR] to [−ATR] vowels is highly correlated with the concomitant F_1 increase for three of the speakers, for the other three speakers the formant intensity effect is not well correlated with F_1 increase.

Interpreting the value of r is a tricky business; unfortunately, the correlation coefficient provides little more than a hint of any underlying connection between two variables. The value of r is constrained to lie between -1 and 1; negative values indicate that the best linear relationship has a negative slope. The problem comes in interpreting the magnitude of r, which is influenced by the following three factors. First, the closer the points are clustered around a straight line, the higher r tends to be (Wilcox 1996). A value of 1 or -1 indicates that the data points lie perfectly along a line. Second, if r is not equal to 1 or -1, rotating the points (i.e. changing the slope of the line) will alter r (Wilcox 1996). Third, r is not resistant; a single point can have a large influence on its value and thus give a misleading

---

[4] Occasionally, the statistical accuracy of r is also reported using a confidence interval for the population correlation coefficient ρ that it estimates. This practice has not been followed here, as contemporary judgement is that no satisfactory method for computing such a confidence interval exists, particulary when samples are small (e.g. Wilcox 1996).

picture of how the bulk of the observations are related to one another (Wilcox 1996).

As if the preceding were not enough, r actually tells us nothing about the causal relationship between the two variables. If we firmly establish that r is very nearly 1 (as is the case for speakers #1, 2, and 3), then it might be that shifts in F_1 frequency *entirely cause* the concomitant changes in relative formant amplitude. Unfortunately, the correlation coefficient is not a tool that can be used in establishing such a result (Wilcox 1996). The theoretical expectation is that the variables correlated here will be causally linked in general; the speakers for whom the correlation is poor actually provide an interesting puzzle since this is at odds with general understanding of vocal tract acoustics.

3 Explaining the results

As mentioned, the formant intensity results show that F_2 is frequently a relatively greater contributor to the spectra of retracted tongue root vowels in Degema. Switching perspective, they also show that F_1 is frequently a lesser contributor to the spectral intensity of these same [−ATR] vowels. Stated in this way, the result can be seen to agree with that of Hess (1992), who found that Akan [−ATR] vowels showed a larger F_1 bandwidth. Acoustic theory tells us that this will invariably produce a decreased F_1 intensity, just as in the Degema vowels. We are unable to report direct bandwidth measurements for the Degema vowels, as this is extremely prone to measurement error and may thus require a larger sample in each case.

Earlier discussion has addressed the fact that formant bandwidth is directly related to damping in the acoustic system; the greater the bandwidth, the more the damping. What we did not discuss, however, is the way in which one formant can be damped without damping the others, and indeed this provides us with a sensible interpretation of the present results.

3.1 Constriction damping and cavity resonance

Much literature (including that reviewed above) considers "vocal tract damping" as a singular presence which affects the speech spectrum uniformly. Statements have been made to the effect that the presence of damping decreases the intensity of upper harmonics and that this makes the voice sound tense.

By modeling the vocal tract configuration of the vowel /i/ as a series of connected acoustic cylinders, and performing calculations on an electrical analog (Dunn 1950), van den Berg (1955) was able to distinguish three different contributions to the resistance (damping) of the supraglottal vocal tract. These are:

1. Radiation resistance, due to the coupling of the mouth opening with the air outside, which affects primarily F_2 and higher formants.

2. Friction damping, due to the viscosity of the air within any approximation involving

the tongue or pharynx. This is present to an appreciable degree only in close vowels, and affects primarily F_1.

3. Cavity wall coupling; the vibration of the vocal tract walls in sympathy with that of the air consumes energy from the acoustic signal. This also mainly affects F_1.

There are also damping effects associated with the degree of opening of the glottis and the mode of vibration of the vocal folds.

Regarding our results for Degema formant intensity, the key contributor is, we believe, friction damping due to the air viscosity within a narrow constriction inside the vocal tract. Mason (1948, pp. 118) gives an approximate expression for the characteristic impedance of a constriction placed in a series of cylindrical sections:

$$Z_0 \approx \frac{\rho c}{O}\left[1 + \frac{1}{r}\sqrt{\frac{\mu}{2\omega\rho}} - i\frac{1}{r}\sqrt{\frac{\mu}{2\omega\rho}}\right] \tag{2}$$

where O is the cross-sectional area of the constriction, ρ is the density of air, c is the speed of sound in air, ω is the angular frequency of the sound waves, r is the radius of the constriction (or what it would be if it were circular), and μ is the coefficient of viscosity for the air in the constriction. This expression gives, roughly, the ability of a constriction to consume the energy supplied by a driven oscillation (i.e. the damping factor); the crucial point is that losses owing to air viscosity are not neglected. The characteristic impedance in a tube neglecting losses is simply

$$Z_0 = \frac{\rho c}{O};$$

this does not account for the flow conditions in a narrow constriction, and is thus only suitable to model fairly wide tubes. With the cross-sectional area in the denominator, this "simple" impedance increases as r^2 decreases. Mason's constriction impedance, however, contains an additional term which increases as r^3 decreases; accounting for losses makes quite a difference for small openings.

A small constriction used as the radiation orifice of a cavity will provide considerable damping to the resonances within. Recall our earlier discussion of the relation of the various formant frequencies to the vocal tract cavities present during vowel articulation. There we reviewed the oft-espoused approximation (due to the modeling of the vocal tract as a pair of tanks connected by a constriction) that the "back cavity" is the source of F_1. This is at least approximately true for any vowel articulated with a fair constriction separating a sizable oral cavity from a larger pharyngeal cavity. It must, however, be disregarded for other configurations. In the applicable articulations, though, it is clear that a narrow constriction separating the front and back cavities would be a source of Mason's friction damping. This friction damping would primarily affect the resonance frequencies of the tank behind, since sound at this frequency is due to vibration in this tank. Vowels which fit the double-Helmholtz resonator configuration (and its variations) in general produce F_1 in the somewhat larger back cavity and F_2 in the front, so the constriction damping would primarily affect F_1 without much affecting F_2. This would be manifested as a broadening of the F_1 bandwidth, with an accompanying decrease in A_1 and a decrease in the relative formant intensity $A_1 - A_2$.

3.2 Articulatory models of Degema vowels

To get some idea whether constriction damping plays a part in the Degema ATR contrast, we attempted to find out more about vowel articulation in Degema using just the acoustic records. A certain amount of articulatory information can be extrapolated from the formant frequency values of a vowel through articulatory modeling. By applying a computer simulation of the vocal tract, modeled as a series of interconnected cylinders, vocal tract area functions (i.e. maps of the changing cross-sectional area from glottis to mouth) can be obtained for any set of formant frequencies. The accuracy of the corresponding area functions depends largely on the number of cylinders used in the model.

Two problems plague the procedure that was used in the present work, limiting our confidence in the results. First, the vocal tract model is fixed in overall size, although the dimensions along its length can be altered. Modeling a specific speaker using a generic vocal tract template can thus be likened to "putting one person's tongue in another person's mouth" (Ladefoged et al. 1978, p. 1034).

Another problem is that an area function corresponding to a set of formant frequencies (particularly if this is limited to the first two formants) is not unique; compensatory forms of articulation can cause entirely different vocal tract shapes to produce the same formant frequencies (Båvegård and Fant 1995). In a vocal tract model which is physiologically realistic this problem is partly alleviated, since a great many area functions which would produce a given set of formants cannot correspond to any actual tongue shapes. This is owing to the basic anatomical factors restricting the relative positioning of the different parts of the tongue. With these caveats in mind, modeling the Degema vowel formants will provide a useful qualitative illustration of the kinds of tongue postures that are likely to be involved.

3.2.1 Method

The program VocalTracts (written by P. Ladefoged) was used to accomplish the desired modeling. Using this program, vocal tract shapes were developed by matching their first two computed formants against the mean F_1 and F_2 for each of the five advanced tongue root vowels of two of the speakers. This is essentially a trial and error procedure, guided by presumptions about the articulations involved. In addition, the lip apertures of speaker #2 producing each of the ten vowels in isolation were measured from a videotape. These values were used as a guide in the modeling procedure. A vocal tract shape was considered representative when its computed F_1 and F_2 matched those of the real vowels to within a 10

Hz margin. This is because in practice, the model may be unable to adapt to a speaker's set of formant frequencies and no solution can be found unless some tolerance is allowed (Båvegård and Fant 1995, p. 55).

After modeling the [+ATR] vowels in this fashion, each vocal tract shape was adjusted, primarily by retracting the tongue root, until the computed formants matched those of the corresponding [−ATR] vowel in the actual speech. To minimize the number of degrees of freedom, larynx height adjustment was not employed as a means of expanding the pharynx. The articulatory hypothesis implicit in this approach is that the difference between [+ATR] and [−ATR] is small insofar as the more familiar aspects of vowels are concerned, and that most of the difference lies in the tongue root/pharynx region.

3.2.2 Degema tongue shapes

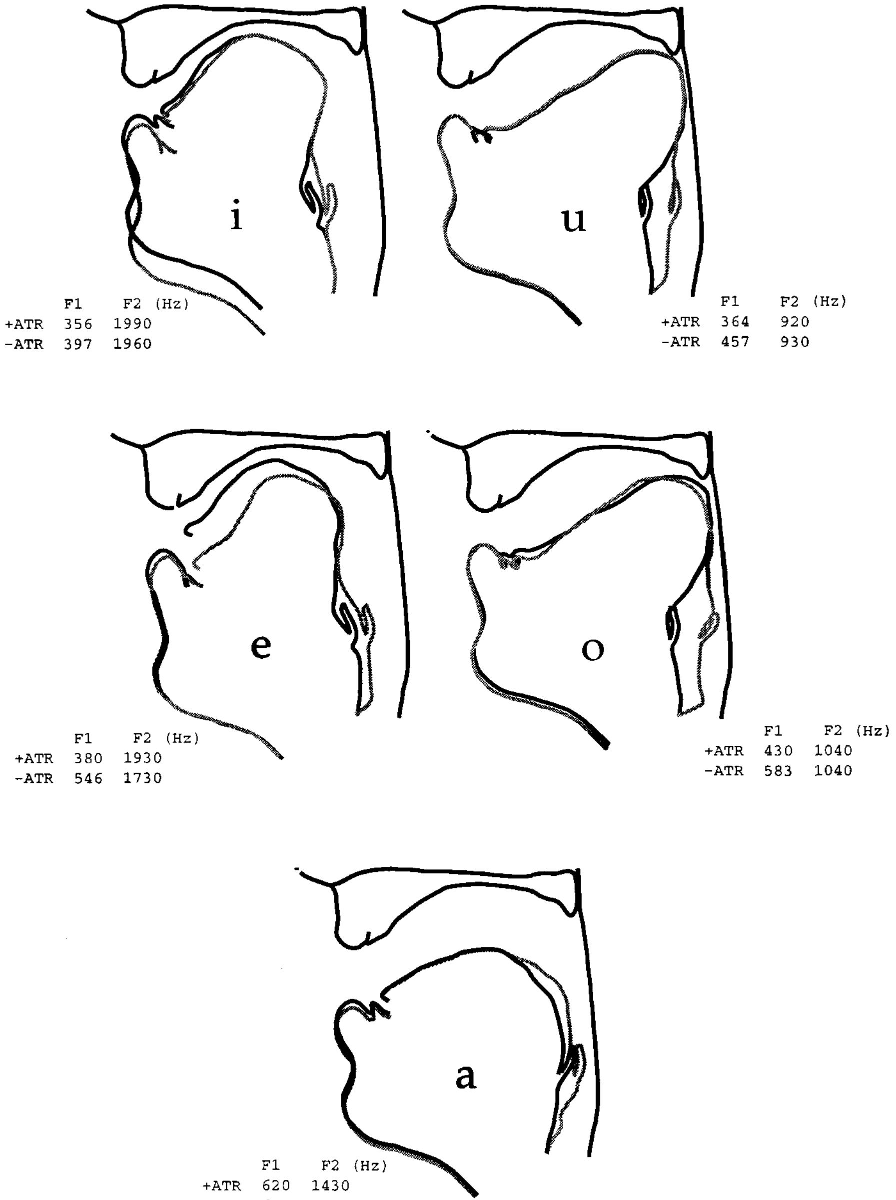
The resulting tongue shapes provided by the model are shown in Figures 6 and 7; [+ATR] vowels are again shown with black lines, as against the grey lines for the [−ATR] vowels. The formant frequencies computed for these shapes are also shown.

It must be emphasized that these pictures cannot be taken as seriously as x-ray tracings, or any other direct articulatory information. They do not precisely correspond to the speakers' actual articulation, but they can nonetheless be viewed as an indication of the likely articulatory behaviour in each case.

Interpretation Despite their approximate nature, the derived tongue shapes show remarkably well the effect sought. The retracted vowels in general demonstrate a marked narrowing in the pharyngeal region; this narrowing tends to increase the length of the constricted portion considerably, leading to a lengthened region over which Mason's friction damping could be expected to act. An increase in the degree of friction damping in the channel leading out from the back cavity is predicted to increase the bandwidth of any formants directly associated with the resonances of that cavity. In the case of vowels fitting the traditional two-cavity mould, this will generally mean a decrease in the relative intensity of F_1.

3.3 Pharyngeal tension

In earlier discussion it was suggested, following Hardcastle (1976), that isometric tension of the pharyngeal constrictors is likely to accompany pharyngeal constriction. By increasing the tension of the throat walls in this manner, the vocal tract damping resulting from coupled vibration of the walls would be reduced. We also mentioned van den Berg's (1955) indication that this would primarily affect resonances below 1000 Hz. Given that F_1 is in this frequency range, the effects of pharyngeal tension in [−ATR] vowels should be opposite to our observation of *increased* damping of F_1 relative to F_2.



	F1	F2 (Hz)
+ATR	356	1990
-ATR	397	1960

	F1	F2 (Hz)
+ATR	364	920
-ATR	457	930

	F1	F2 (Hz)
+ATR	380	1930
-ATR	546	1730

	F1	F2 (Hz)
+ATR	430	1040
-ATR	583	1040

	F1	F2 (Hz)
+ATR	620	1430
-ATR	677	1320

Figure 6: Modeled vocal tract shapes for speaker 1

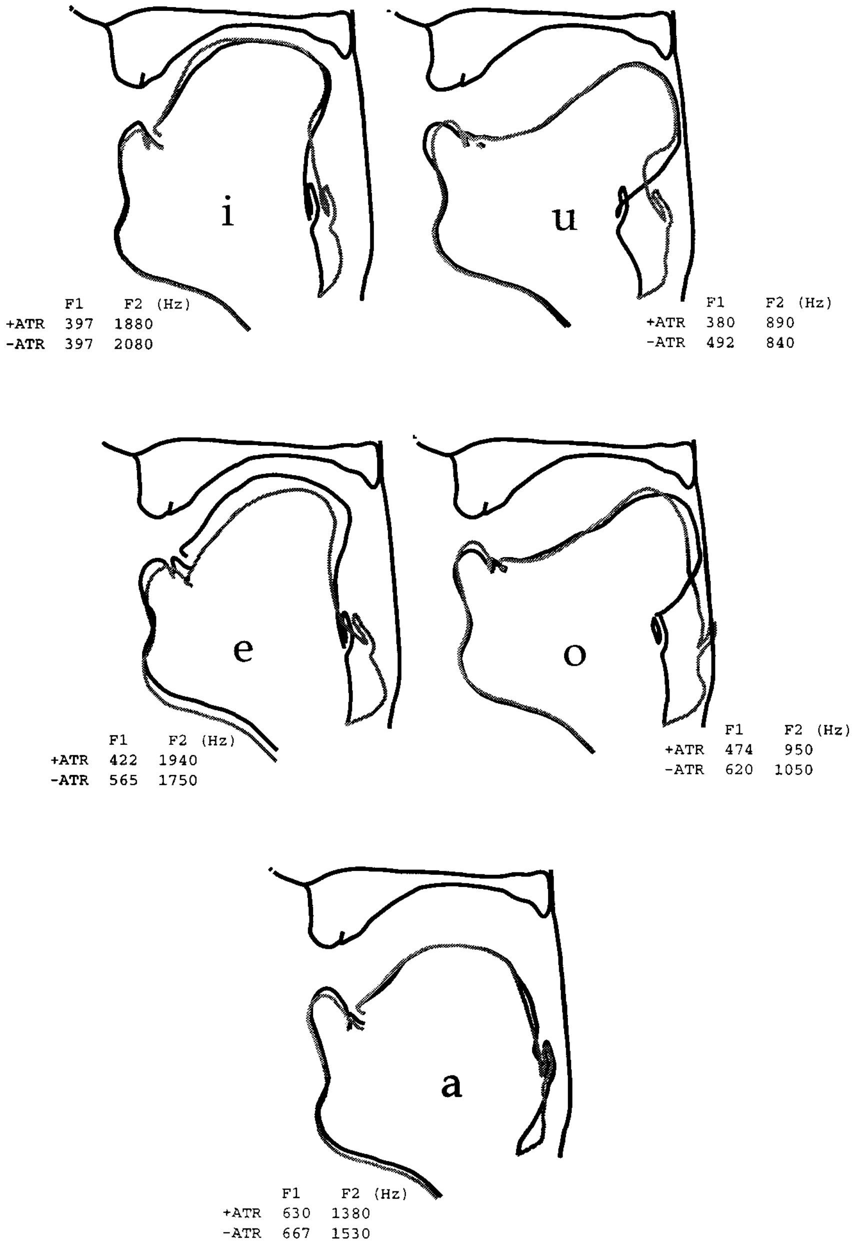


Figure 7: Modeled vocal tract shapes for speaker 2

This can be partly explained by the fact that the vocal tract wall coupling does not contribute as much to the damping as do the losses due to viscous flow in a constriction. Here, some calculations by Stevens (forthcoming) are relevant. He derives values for the contribution to the formant bandwidths B_x (in Hz) by a number of different sources of damping in the vocal mechanism; more dominant sources of resistance have a larger contribution. For $F_1 = 300$ Hz, produced by a Helmholtz resonator with opening area 0.32 cm^2, Stevens gives $B_w = 28$ Hz (contributed by coupling with the vocal tract walls). His formula for the contribution by airflow in a constriction gives $B_c = 37$ Hz.

Perhaps more importantly, the damping characteristics of the vocal tract walls cannot be changed by the speaker over as wide a range as the constriction damping. So, the pharyngeal wall coupling factor will remain much steadier than the constriction damping factor during pharyngeal constriction of the sort illustrated above. Additionally, the damping contributed by the vocal tract walls is also larger for smaller constrictions, whether the walls are stiff or slack. This amounts to saying that any *decrease* in F_1 damping caused by increased pharyngeal tension will be effectively swamped by the *increase* in damping owing to a marked narrowing of the pharynx.

3.4 Perceptual implications

Earlier, we discussed the probable perceptual relevance of our measure of spectral shape (i.e. the intensity difference between F_1 and F_2). Taking a second look at the Degema formant plots in Figures 2 and 3, the reader may notice that the ten vowels could not possibly be distinguished by formant frequency alone. One may also note that the five [+ATR] vowels are generally well-separated in the F_1–F_2 space, themselves forming a vowel sub-inventory which is easily categorized in this way. The same holds true for the [−ATR] vowels. The perplexing question is: can the speakers distinguish all ten vowels properly? We cannot address this issue at length here, but the implication at this point is clear; if Degema speakers can distinguish all ten vowels, they must make some use of the formant frequency and intensity factors (i.e. the perceptually salient aspect of the spectral shape) that the present investigation has described.

Our statistical analysis of the intensity factor has shown it to be somewhat shaky, acting as more of a tendency than a constant presence. This may help explain the fact that ATR contrasts of this sort are invariably involved in a phonological process of vowel harmony, wherein the linguistic system applies a significant constraint to the distribution of the ten vowels in actual utterances. Degema speakers know that all the vowels of a given phonological word must belong to one set or the other, and this knowledge may facilitate easier classification of the vowels. The listener may thus use the spectral information from more than one vowel to complete the classification task.

4 Concluding remarks

This thesis has attempted to cover considerable ground, tracing some of the history of the linguistic investigation of ATR contrasts in African languages, and discussing Stewart's original work at some length. The general nature of the ATR distinction has been shown to be unlike a tense/lax distinction; Lindau's characterization of tongue-root advancement as pharyngeal expansion seems to be a better generalization.

The expected acoustic correlates of pharyngeal expansion/contraction have been discussed, through consideration of the acoustic properties of the vocal tract and the relevant actions within. The expected effects, based on both mathematical facts and past empirical results, are the lowering of F_1 for [+ATR] vowels and a possible difference in vowel spectral shape between the two tongue root postures. Our discussion of the tense or lax *sound* of the voice led to the suggestion that this could be linked to the listener's perception of spectral shape by a perceptually relevant measure, namely $A_1 - A_2$ (the intensity difference between the first two formants).

In an investigation into the acoustic properties of the ATR contrast in Degema, the expected difference in F_1 was largely confirmed, but conventional formant plots showed that the F_1–F_2 acoustic space is not sufficient to permit the classification of the ten vowels. The formant intensity measure $A_1 - A_2$ was shown to distinguish [+ATR] from [−ATR] in a great many cases; this fact led to further discussion of the acoustic properties of the articulations involved.

To facilitate this, vocal tract models corresponding to the vowel formants of two Degema speakers were constructed via computer simulation. The influence of friction damping due to air viscosity within a narrow constriction was shown to be a decisive factor in the determination of the theoretically expected damping of the first formant. The approximate tongue shapes provided by simulations of Degema vowels were then shown to illustrate the influence of friction damping on the actual acoustic results for these same vowels.

It is thus suggested that the increased damping of F_1 due to constriction narrowing is an important contributor to the overall sound of the [−ATR] vowels. This change is no doubt detectable by the Degema speakers, who may be listening for a change in relative formant intensities as was measured here. They may also listen for the F_1 bandwidth directly; indications are that both factors should be perceptually salient. The presence of these acoustic correlates is thought to contribute to the Degema's remarkable ability to distinguish the ten vowels shown in Figures 2 and 3. The statistical unsteadiness of the measured effects may be compensated for somewhat by the additional constraints on vowel distribution provided by the ATR harmony system.

References

Assman, Peter F., and Terrance M. Nearey. 1987. Perception of front vowels: The role of harmonics in the first formant region. *Journal of the Acoustical Society of America* 81(2):520–534.

Båvegård, Mats, and Gunnar Fant. 1995. From formant frequencies to VT area function parameters. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1995(4):55–66. Stockholm: Royal Institute of Technology.

van den Berg, J. 1955. Calculations on a model of the vocal tract for the vowel /i/ (meat) and on the larynx. *Journal of the Acoustical Society of America* 27:332–337.

Berry, J. 1955. Some notes on the phonology of the Nzema and Ahanta dialects. *Bull. School of Oriental and African Studies* 17:160–165.

Berry, J. 1957. Vowel harmony in Twi. *Bull. School of Oriental and African Studies* 19:124–130.

Cattell, R. B. 1944. 'Parallel proportional profiles' and other principles for determining the choice of factors by rotation. *Psychometrika* 9:267–283.

Cliff, N. 1993. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin* 114:494–509.

Dunn, H. K. 1950. The calculation of vowel resonance and an electrical vocal tract. *Journal of the Acoustical Society of America* 22:740–753.

van Dusen, C. R. 1941. A laboratory study of the metallic voice. *Journal of Speech Disorders* 6:137–140.

Elugbe, Ben O. 1989. *Comparative Edoid: Phonology and Lexicon.* University of Port Harcourt Press.

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production.* The Hague: Mouton. Reissued 1970.

Fligner, M. A., and G. E. Policello II. 1981. Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association* 76:162–168.

Fulop, Sean A. 1994. Acoustic manifestations of tension in plosives. Master's thesis, The University of Calgary.

Halle, M., and K. Stevens. 1969. On the feature Advanced Tongue Root. *Quarterly Progress Report, MIT Research Laboratory of Electronics* 94:209–215.

Hardcastle, W. J. 1976. *Physiology of Speech Production.* London: Academic Press.

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

Hess, Susan. 1992. Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan. *Journal of Phonetics* 20(4):475–492.

Hockett, C. F. 1958. *A Course in Modern Linguistics.* New York: Macmillan.

Jackson, Michel T. T. 1988. Phonetic theory and cross-linguistic variation in vowel articulation. *UCLA Working Papers in Phonetics* (71). PhD dissertation.

Jacobson, Leon C. 1978. DhoLuo vowel harmony. *UCLA Working Papers in Phonetics* (43). PhD dissertation.

Jakobson, R., and M. Halle. 1962. Tenseness and laxness. In Roman Jakobson (Ed.), *Selected Writings of Roman Jakobson*, Vol. 1, 550–555. 's-Gravenhage: Mouton.

Kakusho, O., H. Hirato, K. Kato, and T. Kobayashi. 1971. Some experiments of vowel perception by harmonic synthesizer. *Acustica* 53:132–142.

Kaplan, H. M. 1960. *Anatomy and Physiology of Speech.* New York: McGraw-Hill.

Kohler, Klaus J., and W. A. van Dommelen. 1987. The effects of voice quality on the perception of lenis/fortis stops. *Journal of Phonetics* 15:365–381.

Ladefoged, Peter. 1968. *A Phonetic Study of West African Languages.* Cambridge: Cambridge University Press. Second edition.

Ladefoged, Peter. 1983. The linguistic use of different phonation types. In D. Bless and J. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, 351–360. San Diego: College-Hill Press.

Ladefoged, Peter, R. Harshman, L. Goldstein, and L. Rice. 1978. Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America* 64(4):1027–1035.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell.

Ladefoged, Peter, Ian Maddieson, and Michel Jackson. 1988. Investigating phonation types in different languages. In O. Fujimura (Ed.), *Vocal Physiology: Voice Production, Mechanisms, and Functions.* New York: Raven Press.

Laver, John. 1980. *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Lindau, Mona. 1974. The feature Advanced Tongue Root. In E. Voeltz (Ed.), *Third Annual Conference on African Linguistics*, 127–134. Bloomington: Indiana University.

Lindau, Mona. 1978. Vowel features. *Language* 54(3):541–563.

Lindau, Mona. 1979. The feature Expanded. *Journal of Phonetics* 7:163–176.

Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18:50–60.

Mason, W. P. 1948. *Electromechanical Transducers and Wave Filters.* New York: D. van Nostrand.

Morse, Philip M. 1948. *Vibration and Sound.* Acoustical Society of America.

Painter, Colin. 1971. Vowel harmony in Anum. *Phonetica* 23(4):239–248.

Pierce, Allan D. 1989. *Acoustics: An Introduction to its Physical Principles.* Acoustical Society of America.

Pike, K. L. 1947. *Phonemics.* Ann Arbor: University of Michigan Press.

Pike, K. L. 1967. Tongue-root position in practical phonetics. *Phonetica* 17(3):129–140.

Stevens, Kenneth N. n.d. Basic acoustics of vocal-tract resonators. From Chapter 4 of a book on acoustic phonetics, in preparation.

Stewart, J. M. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16(4):185–204.

Stewart, J. M. 1971. Niger-Congo, Kwa. In Thomas A. Sebeok (Ed.), *Linguistics in Sub-Saharan Africa,* Vol. 7 of *Current Trends in Linguistics*, 179–212. The Hague: Mouton.

Tiede, Mark K. 1993. An MRI-based study of pharyngeal volume contrasts in Akan. *Haskins Laboratories Status Report on Speech Research* SR-113:107–130.

Tucker, A. N., and M. A. Bryan. 1966. *Linguistic analyses: The non-Bantu languages of north-eastern Africa.* London: Oxford University Press.

Tukey, J. W. 1991. The philosophy of multiple comparisons. *Statistical Science* 6:100–116.

Ward, I. C. 1937. Phonetic phenomena in African languages. *Archiv für die gesamte Phonetik* 1.

Welmers, William E. 1946. A descriptive grammar of Fanti. *Language* 22(3). Supplement.

Wilcox, Rand R. 1996. *Statistics for the Social Sciences.* San Diego: Academic Press.

Phonetic structures of Banawá, an endangered language

Peter Ladefoged and Jenny Ladefoged
Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, CA 90095-1543
and
Daniel Everett
Linguistics Department, University of Pittsburgh, Pittsburgh, PN 15260

This paper describes the phonetic characteristics of Banawá, an endangered language spoken in Brazil. The qualities of the Banawá vowels are described in terms of their formant frequencies. The places of articulation of each consonant and the VOT and the manners of articulation are documented. The structure of syllables and words is delimited, and the location of stressed syllables is described and verified experimentally.

Nobody knows how exactly many languages there are in the world today, but there is little doubt that there are more now than there will be in our children's time. The best estimates of the numbers of languages and speakers are those in the Ethnologue (Grimes, 1992), which lists nearly 7,000 distinct languages. Over half of the languages listed are spoken by under 10,000 people, about a quarter of them are spoken by under 1,000 people, and about 10% by under 100. It will be very difficult for languages spoken by small numbers of people to survive.

One of these endangered languages is Banawá, an Arawan language, spoken by about 75 speakers deep in the rain forest in northern Brazil (see Figure 1). The group has been known to non-indigenous people for a comparatively short time, as they were first contacted by missionaries only 30 years ago. Since that time they have acquired many Brazilian characteristics; they have already learned to play football, Brazil's national pastime, which they eagerly watch on television when they manage to get out of the forest. They will probably become more completely assimilated into Brazilian culture in the near future, and their language may not be spoken in a generation or two. Irrespective of one's views on cultural change, there is an obvious scientific necessity to document the phonetic structures of the language while speakers are still available. This paper will provide some of this documentation through projects sponsored by the National Science Foundation to whom we are very grateful.

When describing the phonetic structures of any language we would like to describe all and only the linguistically relevant phonetic facts. Unfortunately, the resources that we have available for the description of a single language do not permit this. This account of Banawá is limited in two ways. Firstly we will describe only the major phonetic structures — the quality of the vowels in limited contexts, the principal consonantal properties, and some aspects of the suprasegmental features. This is all that it is possible to do within the limits of the present NSF project, if we are to consider a number of endangered languages for which even this information is not available. Secondly it is based on the speech of only five male speakers. We were unable to collect data from a larger group including female speakers. However, the limitation to only five speakers may not be as unsatisfactory as it first appears. There are probably less than 30 adult male speakers of Banawá, so a sample of five speakers still represents a larger percentage of the adult male population than is used as a basis for most phonetic descriptions.

Figure 1. The location of the approximately 75 speakers of Banawá.

The first necessity for an account of the phonetic structures of any language is a good phonological description of that language. There are two previous publications on the language (Buller, Buller and Everett, 1993; Everett 1995) which, supplemented by the third author's knowledge of the language, enabled us to determine appropriate material to record. The data for this study consists principally of recordings of 6 male speakers made in July 1995. The recordings of one of the speakers proved unusable for quantitative work, as he spoke so quietly and shyly that we eventually had to abandon attempts to make reliable analyses. Two of the speakers were recorded at the SIL Center in Porto Velho, where they worked extensively with us in preparing the word lists and other materials. The other speakers were recorded in the Banawá settlement, which is in the jungle about 120 miles due North of Porto Velho. All the recordings were made on a DAT recorder, using a close-talking, noise-canceling microphone for each speaker individually. The frequency response was substantially flat throughout the audible range, and the signal/noise ratio was greater than 45 dB.

Vowels

Banawá has four vowels, **i**, **e**, **a**, **u**, illustrated by the words in Table 1. We analyzed two tokens of each of the vowels in the words in Table 1, as spoken by each speaker (a total of 5 speakers with 2 tokens of 16 words = 160 vowels), using a Kay CSL system. Formant frequencies were determined from observations of the formant histories throughout the word, and superimposed FFT and LPC spectra made at the most steady state portions of each vowel.

Table 1. Words illustrating the vowels of Banawá in stressed syllables after **b** and **t**, and in unstressed syllables after **b** and **f**.

Stressed		**Unstressed**	
bita	mosquito	**ibi**	each other
befa	other	**ibe**	a strip
bata	to pick	**iba**	to put/place
bufa	put on water	**ibufa**	to dump into water
tifa	drink water	**tafi**	eating
tefe	food (m.)	**tafe**	food (f.)
tafa	to eat	**tafa**	to eat
tufa	to block in	**tafu**	to eat

Even with good computer analysis facilities, the determination of the formants — the resonances of the vocal tract, as opposed to the peaks in the observed spectra — cannot be made entirely algorithmically. There are often cases when LPC spectra either do not show a sharp peak or provide a spectral peak that is clearly not a resonance of the vocal tract. FFT spectra may be of no further assistance. If the number of points in the transform is such that individual harmonics are shown, then the formant resonances are not each defined by a single peak; if the number of points in the transform is less, then the time interval may reflect only a particular part of a single glottal pulse. All these factors lead to the possibility of error in the determination of the formant frequencies from the acoustic data.

One way of checking on the reliability of the measurement procedures is to compare the first two formants in the two tokens of each word, as was first done by Peterson and Barney (1952). There is a confounding factor in this procedure in that comparison of the two sets of measurements shows both differences in the tokens produced, and differences due to the accuracy of the measurement techniques; the speaker might have used a different pronunciation, or there might have been measurement error. Variation between tokens could be avoided by comparing two sets of measurements of the same token. However, errors in determining the true resonances of the vocal tract are more likely to arise when considering different tokens which may have slight differences in the glottal source (a common and unnoticed variation in the production of the same vowel in a language without phonologically contrastive phonation types). The FFT and LPC procedures are affected by differences in the glottal source function even when the formant resonances remain the same. We want to have a valid representation of each vowel type as produced by each speaker separately, irrespective of any phonation type difference or any other non-contextual variation that might have been made. Accordingly we can conveniently combine the two sources of variation (within speaker variation and measurement error) and check the reliability of our representation by comparing two tokens of each vowel as produced by each speaker. (When considering within speaker variability it would have been nice if we had had more than two repetitions of each vowel by each speaker. But we do not; and, as shown by Johnson, Ladefoged and Lindau (1995), between speaker variability, which we consider later, is much greater than within speaker variability.)

The results for the stressed vowels are as shown in Figure 2. It may be seen that there is a good correlation between the two sets of measurements. The higher values of F1 have some differences, but the measurements of F2 are very similar in the two repetitions of the same word.

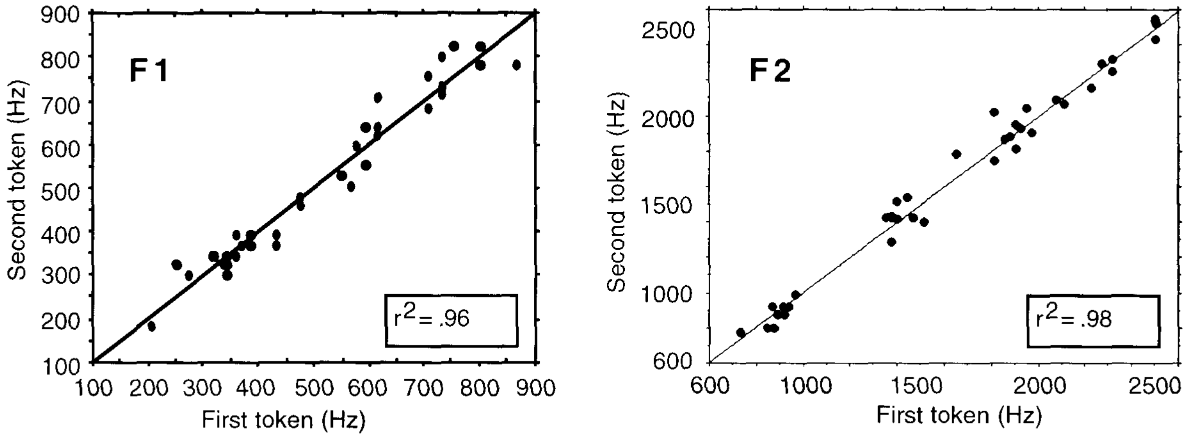


Figure 2. The correlation between two measurements of each of F1 and F2, one measurement in the first token of each of the words containing stressed vowels, and the other in the second token of the same word.

The same procedure was used in a comparison of the unstressed vowels, as shown in Figure 3. The F1 differences are slightly greater, but the F2 differences are very much the same as in the case of the stressed vowels. For both the stressed and the unstressed vowels, the high correlations indicate that the measurements of the formant frequencies are reliable.

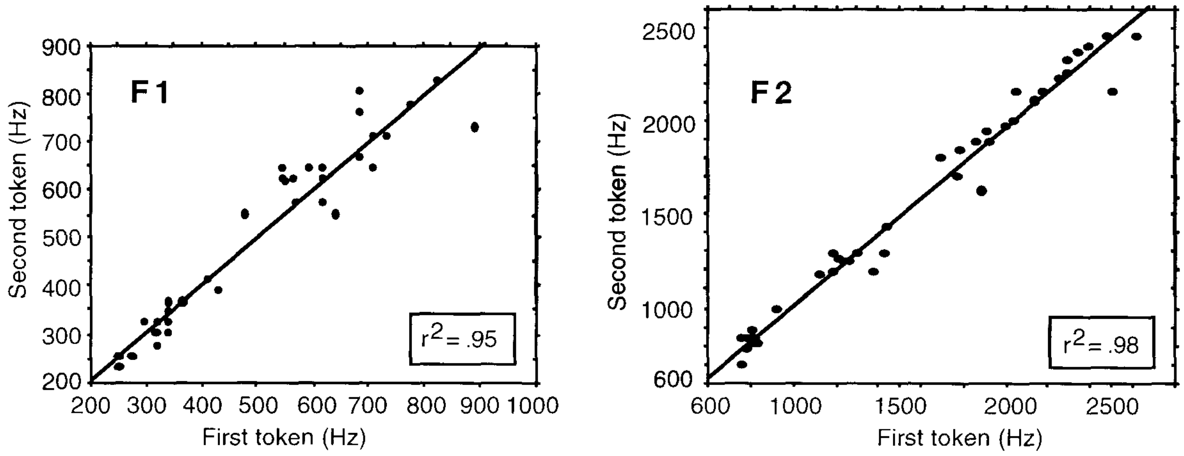


Figure 3. The correlation between two measurements of each of F1 and F2, the one measurement in the first token of each of the words containing unstressed vowels, and the other in the second token of the same word.

Figure 4 shows the first two formant frequencies of the vowels as produced by the five speakers saying two tokens of each of the words in Table 1 illustrating the stressed vowels (20 points for each vowel, some of which are overlapping). The values on the scales are in Hz, but the distances on the scales are arranged so that equal distances represent equal bark intervals of each formant. The ellipse in this figure encloses four vowels of speaker 4 which are significantly different in their F1 values from those of F1 in the same words as spoken by the other 4 speakers ($p < .01$ in an ANOVA of F1 by speaker by vowel). This speaker has an aberrant vowel, with a higher F1 (a more open vowel) than that of other speakers.

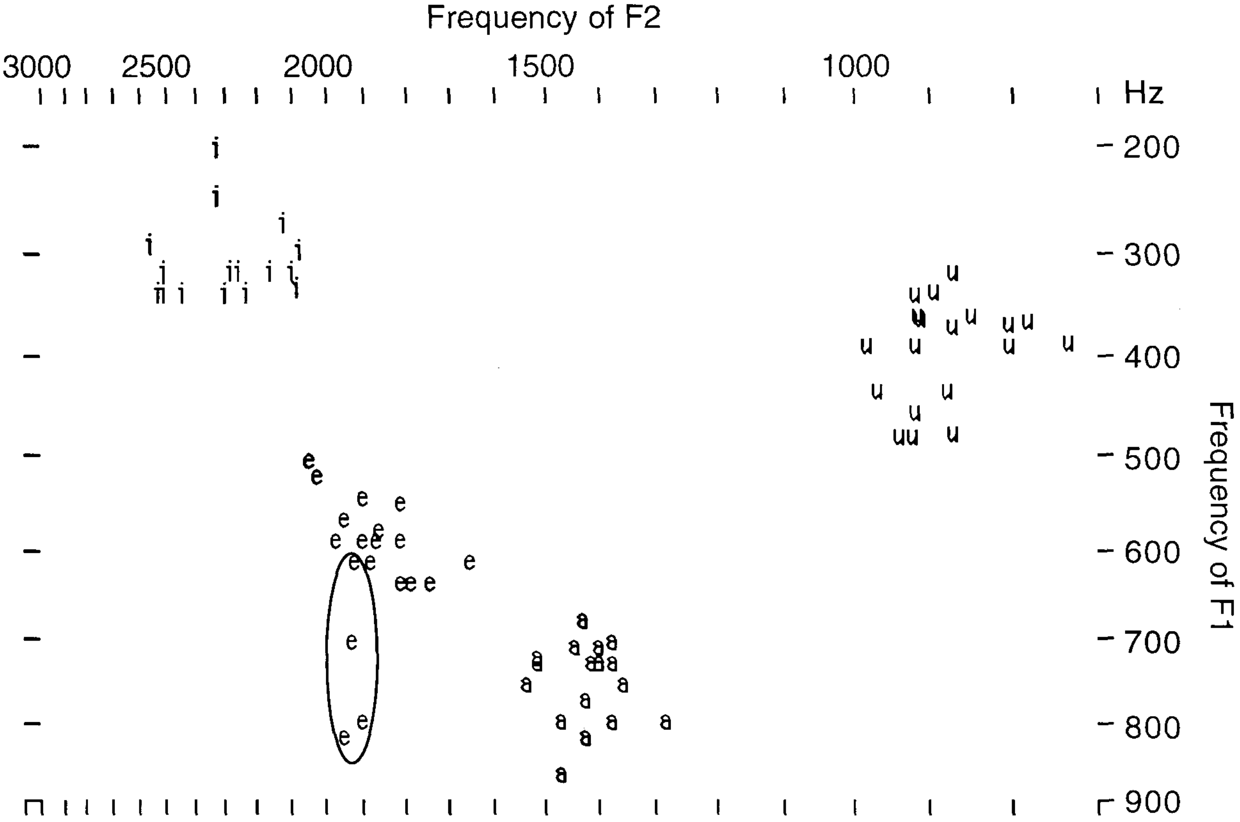


Figure 4. The first two formant frequencies of the stressed vowels in Table 1, as spoken by 5 speakers. The ellipse encloses the four vowels of speaker 4 which are discussed in the text. See text for discussion of the scales and axes.

Figure 5 shows the variation in the first two formant frequencies of individual tokens in Figure 4, except that speaker 4's **e** vowels have been omitted from the calculations. The first step in plotting this graph was to determine the principal components of the variation in the F1-F2 space shown in Figure 4, considering vowels after labials and vowels after coronals separately. Ellipses were centered on the means for the vowels in each of these contexts. The axes of the ellipses were angled along the principal components, and the radii were made equivalent to two standard deviations of the mean variation in these dimensions. Assuming our five speakers constitute a representative sample of the approximately 30 adult male speakers of Banawá, we can say that 95% of them, i.e. about 26 or 27 Banawá adult males, have vowels with formant frequencies that lie within the ellipses shown in Figure 5, with some possible extra variation for the vowel **e**, where one of our speakers had a different vowel quality.

The differences between vowels after bilabials and vowels after coronals are shown in Table 2. As can be seen from both Figure 5 and Table 2, formant frequencies are generally slightly lower after a bilabial consonant than after a coronal consonant. An analysis of variance shows that the lowering of F2 in **u** is very significant ($p < .001$), but that none of the other differences are even probably significant ($p > .05$).

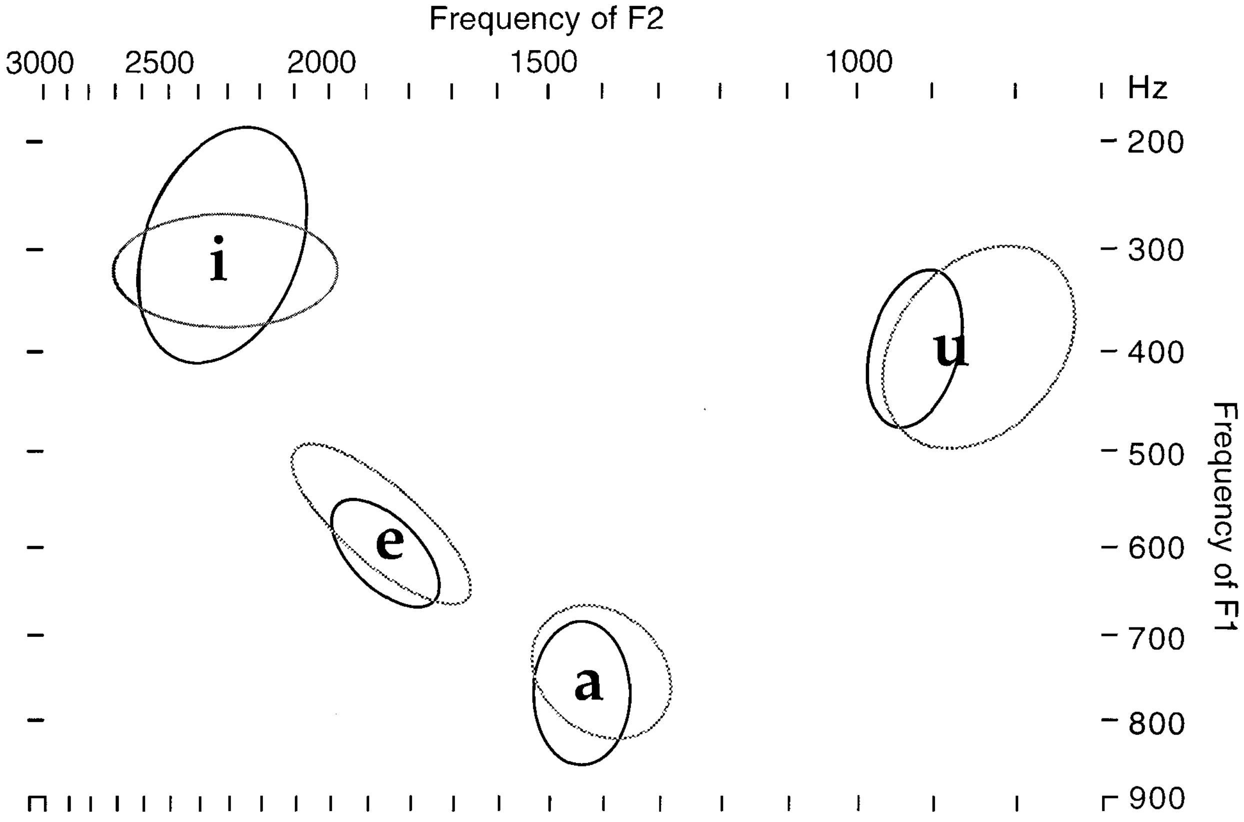


Figure 5. Formant plots of the Banawá vowels in Table 1, as produced by five speakers. The ellipses drawn with solid lines show areas containing all points within two standard deviations of the mean for vowels in stressed syllables after a coronal consonant. The dashed ellipses show the same range for vowels in stressed syllables after a bilabial consonant. The vowel symbols are placed at the grand mean for each vowel, irrespective of context.

Table 2. Differences (Hz) in formant frequencies after bilabials as compared to vowels after coronals.

	i	**e**	**a**	**u**
F1	23.9	-33.0	-27.5	-5.2
F2	-5.0	11.4	-36.6	-46.9**
F3	-56.5	-130.4	95.0	-18.3

The plots in Figure 5 show that the three vowels **i**, **e**, **a**, differ in vowel height from one another, taking F1 as a measure of the feature Height. There is also an apparent difference in height between **i** and **u**. An analysis of variance indicates that when all 5 speakers are taken together, this difference is significant ($p < .01$). This is an interesting point in that work in progress on other endangered languages with a limited number of vowels is showing that the high back vowel is often not as high as the front vowel. In Banawá the high back vowel has been written in two different ways. Everett transcribes it as **u** (Everett 1995, 1996; and in Buller, Buller and Everett, 1993), but Buller and Buller in their field notes (p.c.) transcribe it as **o**. The ANOVA might be taken as indicating that the Bullers' transcription is preferable; however, the same analysis shows that there is an interaction between speakers and vowels, and for two of the five speakers there is no statistical difference in Height between **i** and **u**. In Banawá the difference in Height between **i** and **u** seems to be speaker dependent; in this paper we will continue to use **u**, the more usual symbol for a high back vowel. The statistics for all five speakers (except for **e** in the case of speaker 4) for the first three formants for the four vowels are shown in Table 3.

The ellipses in Figure 5 show that feature values cannot be given context free specifications. Thus [front] (one of the possible values of the feature Back) means one thing when it occurs in the context of [high] (one of the possible values of the feature Height), and another when it occurs with [low]. If we regard the implementation of the feature Back as being dependent on F2, then [front] has a value of 2320 Hz when it is in conjunction with [high] for the vowel **i**, and 1869 Hz for the vowel **e** when it is in conjunction with [low]. Another way of specifying the implementation of the feature Back is to regard it as a function of the distance between the first two formants (the old Jakobsonian definition of Diffuse). In that case the feature specification [front] would have a value of 2320 – 311 = 2009 Hz for a [high] vowel, but this same feature specification would have a value of 1869 - 591 = 1278 Hz for a [low] vowel. Whatever measure one chooses it appears that the phonetic interpretation of Back depends on whether a high vowel or a low vowel is being described. The context dependent nature of feature specifications has long been noted by phoneticians (Ladefoged 1972). We will consider another example in this paper when we discuss differences in voicing.

Table 3. Statistics for mean formant frequencies for all five speakers (except for **e** in the case of speaker 4) for the first three formants for the four vowels.

F1:	Count:	Mean:	Std. Dev.:	Std. Error:
i	20	311	47.1	10.5
e	16	591	41.0	10.2
a	20	758	46.6	10.4
u	20	396	48.6	10.9
F2				
i	20	2320	164.0	36.7
e	16	1869	101.1	25.2
a	20	1422	58.9	13.2
u	20	883	63.8	14.3
F3				
i	20	2822	181.8	40.6
e	16	2741	178.5	44.6
a	20	2593	279.3	62.4
u	20	2685	164.1	36.7

The tables and figures show that these vowels are more in line with the theory proposed by Lindblom (1990), requiring vowels to be what he would call adequately dispersed in the auditory vowel space, rather than with the articulatory/acoustic notion of quantal vowels proposed by Stevens (1989). The low vowel **a** is too far forward (has a too high F2) to be equated with the quantal vowel **ɑ**, which Stevens defined as one in which F1 is maximal and F2 minimal so that they are close together. He notes that this requires "a *backed* and low tongue position" (Stevens 1989:14, emphasis added). Banawá, like many other languages with five or fewer vowels, has a low central rather than a low back vowel.

Consonants

The consonants of Banawá are as shown in Table 4, and illustrated in Tables 5 and 6. The more specific symbols **ɸ** and **ɾ** are used in the chart in Table 4; elsewhere, so that the examples are more readable, the more general symbols **f** and **r** are used. Table 5 shows two examples of each consonant in word initial position before **a** and **i**, and Table 6 has two examples of each consonant in word medial position between **a_a** and between **i_i**.

Table 4. Banawá consonants.

	Bilabial	Dental	Palatal	Velar	Labial-velar	Glottal
Stop	**b**	**t d**	**ɟ**	**k**		
Nasal	**m**	**n**				
Tap		**ɾ**				
Fricative	**ɸ**	**s**				
Approximant					**w**	**h**

Table 5. Examples of each consonant in word initial position before **a** and **i**.

baka	(name)	**bisi**	to pinch
tafa	eat	**tisi**	fall
daka	(name)	**disi**	wasp
ɟaka	walk	**ɟiri**	to shock
kaka	toucan	**kisi**	descend
maka	snake	**misi**	up
naka	sticky	**nisa**	down
faki	twist	**fisi**	monkey
saka	jab, pierce	**sisi**	a few
rawi	write	**risa**	down on (f.)
waka	break	**wisi**	to cut
haku	spider	**hisi**	sniff

Table 6. Examples of each consonant in word medial position between **a_a** (or **u a**) and **i_i**.

baba	(name)	**kibi**	full
bata	rotten	**kiti**	strong
bada	(name)	**bidi**	small
baɟa	palm fronds	**diɟi**	to wobble
baka	(name)	**kiki**	to look to the side
bama	catfish	**kimi**	corn
baa ana	she hits	**kini**	green
bafa	Bafa (river)	**kifi**	to cross a bridge
basa	to put a stick up high	**kisi**	to descend
baa ara	another woman	**kiri**	I am itching
baa uwa	I hit	**tiwi**	did you see?
haha	to laugh	**kihi**	potato

The voiced stops **b**, **d**, **ɟ** are voiced throughout (although, as we will see, the voicing in **ɟ** is often not fully evident). The voiceless stops **t** and **k** are virtually unaspirated. Measurements were made of two tokens of each of the words in Table 7. The mean Voice Onset Times (VOTs) for **t** and **k** in the different phonetic contexts provided by these words are shown in Figure 6. An analysis of variance showed that there was no difference ($p > .05$) between the measurements for initial and medial occurrences, and these measurements have been pooled. Accordingly each column represents the mean of 6 measurements, except in two cases when faults in recording the first two speakers resulted in only four valid measurements of the first word, **tisi**, 'fall'.

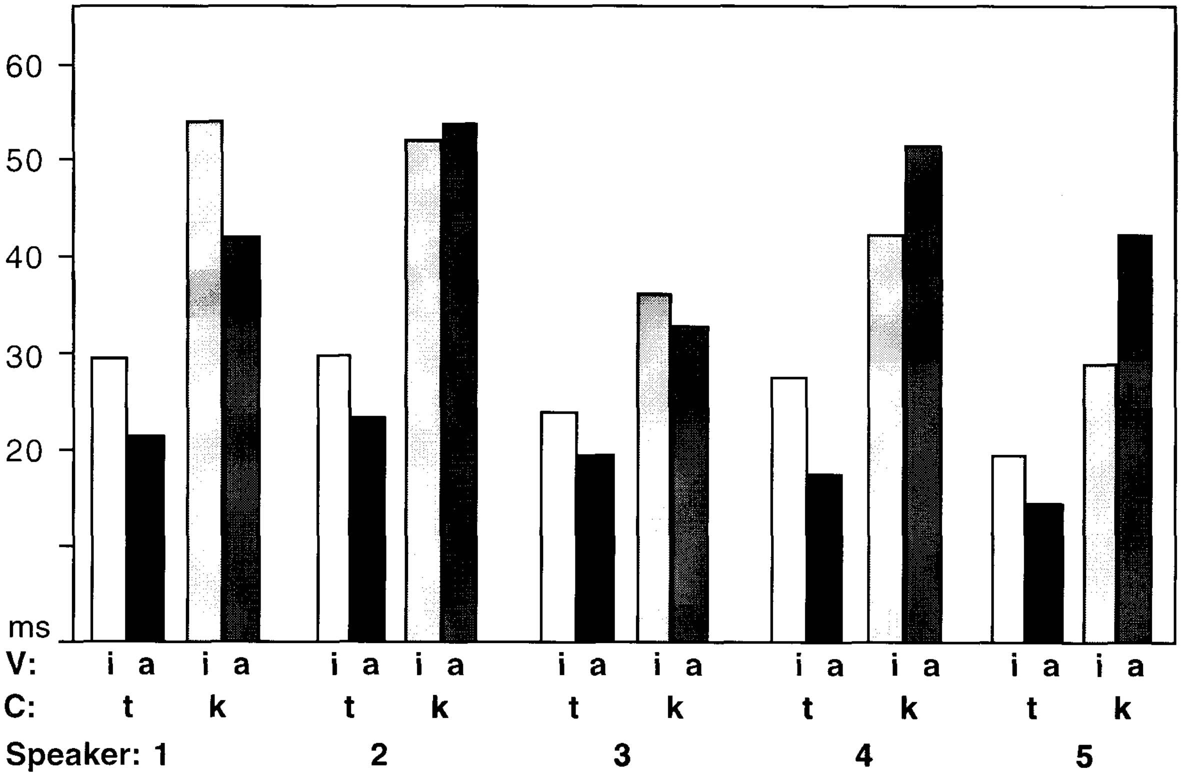


Figure 6. The Voice Onset Times for the two voiceless stops **t** and **k** as produced by the five speakers before the vowels **i** and **a**.

Table 7. The words used for VOT measurements.

	t		**k**	
INITIAL	**tisi**	fall	**kisi**	descend
	tima	up stream	**kimi**	corn
	tafa	eat	**kaka**	toucan
	tama	vine	**kama**	come
MEDIAL	**kiti**	strong	**kiki**	to look to the side
	bata	rotten	**baka**	(name)

The first point to note about Figure 6 is that the VOT for **t** (the white and black columns) is much less than that for **k** (the shaded columns). An analysis of variance shows that for each speaker this is a significant difference ($p < .01$). The mean VOT for velars is 44 ms, whereas that for dentals is 22 ms. There is also a smaller but still significant difference ($p < .01$) in the VOT of **t** before **i** (the white columns) and before **a** (the black columns). There is no significant difference in the VOT of **k** before different vowels, three speakers having a longer VOT before **a**, and the other two before **i**.

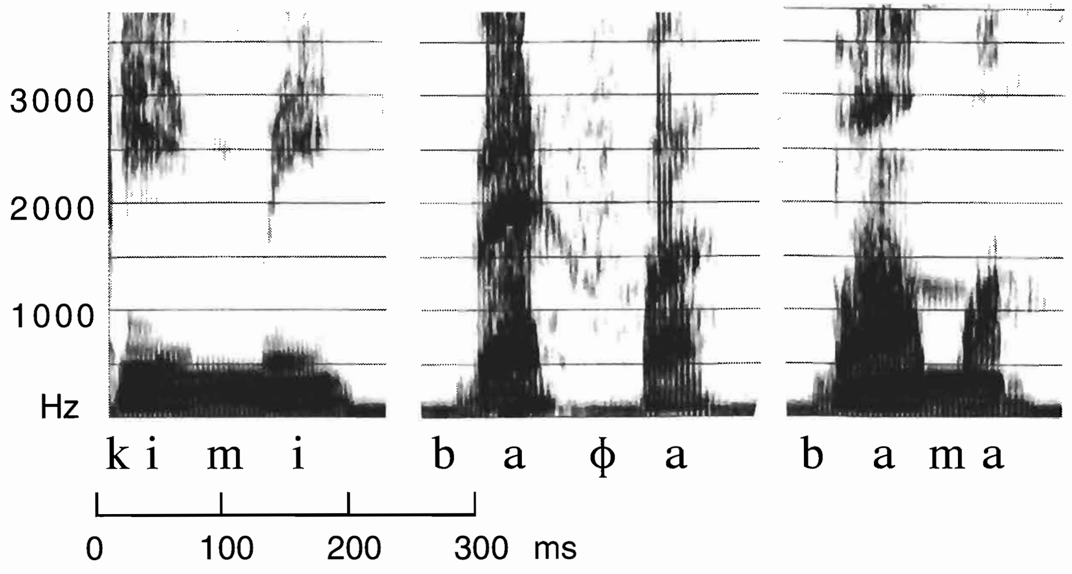


Figure 7. Spectrograms of **kimi**, 'corn', **bafa**, 'Bafa' (name of a river), and **bama,** 'catfish'.

The dependence of VOT on place of articulation (and, to a lesser extent, on the quality of the following vowel) has been noted for many languages, going back at least to Fischer-Jørgensen (1954). This may be another example of the interdependence of features. One has to know whether one is talking about dental or velar stops before one can say how aspirated a stop might be. In other words, the value of the feature Voicing cannot be stated without taking into account the co-occurring value of the feature Place. This may be partly an artifact of our form of description. The best phonetic exponent of the feature Voicing may not be VOT but a measure of the difference in timing between two gestures, one for the stop, and one for the glottal adjustments. If this measure could be stated correctly, at least some of the differences in VOT between velar and dental stops might be seen to be due to concomitant aerodynamic and articulatory differences that we do not yet fully understand. However, some of the differences between places of articulation are language specific. In their extensive study, Lisker and Abramson (1964) report that Dutch unaspirated **t** has a mean VOT of 15 ms and **k** of 25 ms, a difference of 10 ms, whereas Cantonese **t** has a mean VOT of 14 ms and **k** of 34 ms, a difference of 20 ms. In Banawá the mean VOT for dentals is 22 ms and that for velars is 44 ms, a difference of 22 ms. We have to know both what language is being described and which place of articulation is involved before we can assign a phonetic value to the feature Voicing.

All our speakers used a bilabial fricative **ɸ** rather than **f**. As noted, we have used the more common symbol, in accordance with the principle of using a simple phonemic transcription whenever possible in order to make the examples easier to read. We use the more exotic symbol only when specifically referring to the bilabial character of this fricative. An aspect of all the bilabial sounds, particularly noticeable in **m** and **ɸ**, but also evident in **b**, is that the release of the articulation has a **w**-like component, so that these sounds might have been transcribed **mʷ**, **ɸʷ**, **bʷ**. This component, which can be clearly seen in the spectrograms in Figure 7, may not be a labial velar of the kind regularly symbolized by **w**. The lowering of the formants (notably F2) may be simply due to the protrusion of the lips that occurs during the course of these sounds.

All the consonants **t**, **d**, **n**, **s** have laminal articulations. Three of our speakers used interdental articulations for **t**, **d**, **n**. Some of our speakers had rather poor dentition, so it was possible to observe that the fricative **s** was usually made a little further back (it was never interdental); but it was always further forward than the typical English alveolar **s**. The tap which we have symbolized **ɾ** has a dental articulation in which the tongue tip is raised from the floor of the mouth, moving slightly forward as it strikes the anterior part of the alveolar ridge. Spectrograms of this sound in initial and medial position are shown in Figure 8.

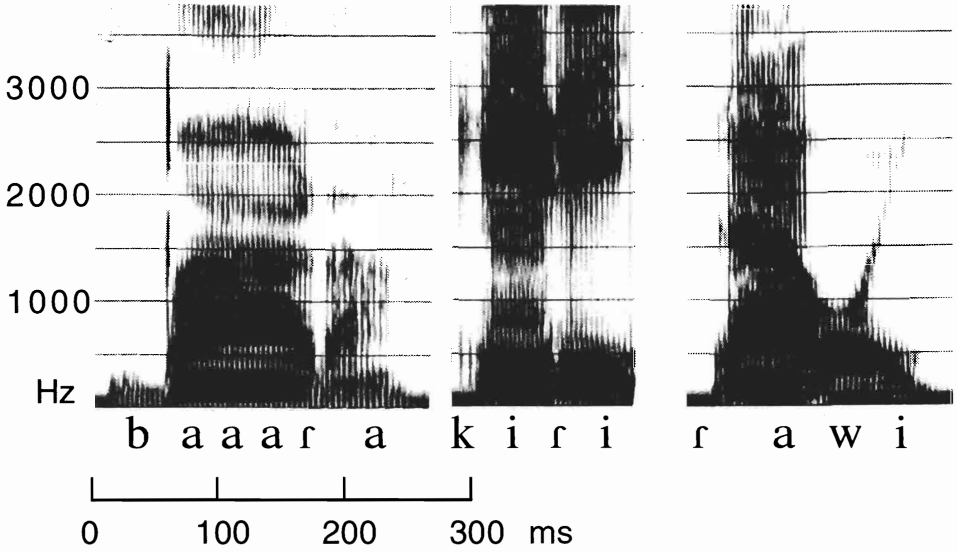


Figure 8. Spectrograms of **baaɾa**, ‘another woman’, **kiɾi**, ‘I am itching’, **ɾawi**, ‘write’.

The voiceless approximant **h** may be accompanied by nasalization of the adjacent vowels, but this is by no means certain. The nasalization of vowels adjacent to **h** (and **ʔ** in other languages) appears to be an areal feature of languages spoken in this region. In the case of Banawá, all our speakers had vowels in words such as **haha** ‘to laugh‘ and **kihi** ‘potato’ that were clearly breathy (and therefore had auditory characteristics in common with nasalized vowels), but spectrographic analysis did not reveal any evidence of a nasal pole/zero pair in any of these vowels. Unfortunately we were unable to obtain any aerodynamic or other physiological records that would throw more light on this problem.

Previous accounts of Banawá (Everett 1995, 1996; Buller, Buller and Everett, 1993) suggested that the consonants **ɟ** and **w** are not part of the phonological inventory of Banawá, but are simply the result of glide formation when underlying **i** and **u** are syllable initial before a vowel. This is somewhat reminiscent of a proposal by Jakobson, Fant and Halle (1951) in which they suggested that English words such as ‘yield’ and ‘woo’ could be transcribed as /iiild/ and /uuu/. This is possible in English because there is no contrast between word initial or final **j**, **w** and **i**, **u** respectively. There are good phonological reasons for these identifications in Banawá. Morphological structures and interlexical forms can cause /i/ to vary between **i** and **ɟ**, and /u/ to vary between **u** and **w**. But none of our Banawá speakers had a glide **j**; for all of them the consonant series included a palatal stop **ɟ**, as exemplified in the spectrograms in Figure 9.

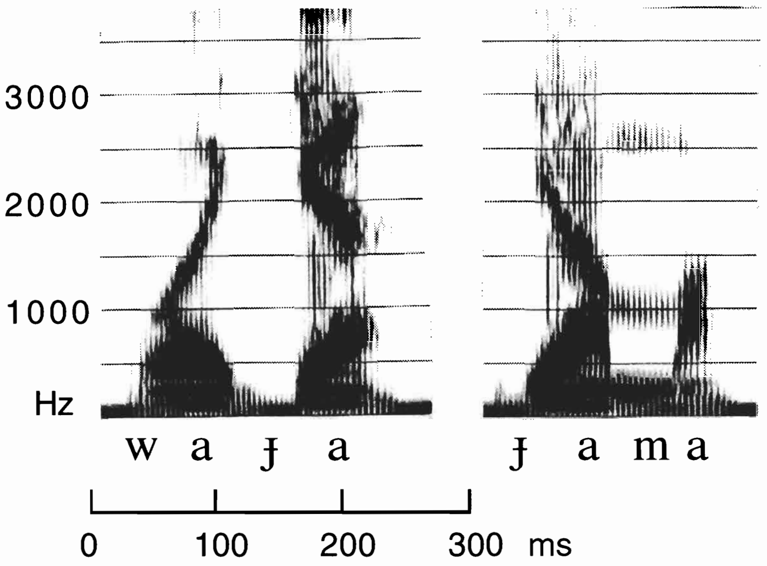


Figure 9. Spectrograms of **baɟa**, ‘palm fronds and **ɟaka**, ‘walk’.

Syllable and word structure

The most familiar building blocks of words and syllables are the consonant and vowel units denoted by C and V. The description of many languages, including Banawá, also requires a unit of timing, the mora, associated with each vowel, V. Syllables in Banawá can be described in terms of the following statements:

(1) Syllables have one or two moras.

Every syllable has at least one vowel, but there are no syllables with longer strings of vowels such as CVVV, which would have three moras.

(2) There are no final consonants.

(3) There are no consonant clusters.

If there is a consonantal onset to a syllable, it consists of only a single consonant.

As far as words are concerned we cannot say how long they could be, but we can say:

(4) All Banawá words have at least two moras.

Words can have one or more syllables, but if there is only one syllable it must have two vowels.

Table 8. Banawá words containing sequences of vowels. All syllable boundaries are shown by a period [**.**].

.bii.	fan
.faa.	water
.kai.ɟa.ra.	to take pride in oneself
.ka.wa.rie.	to cook
.ka.wa.ri.sei.	rafter
.me.na.nau.	they play
.ra.bi.kai.	sick
.u.wa.ria.	one

Some typical Banawá words are given in Table 8. Syllable boundaries within these words have been marked, showing how many syllables there are, and making it apparent that syllables can be CV or CVV, with VV denoting a vowel sequence of two moras within a single syllable. On most occasions this sequence is very similar to the sequence of vowels that occurs in a diphthong in traditional descriptions. All VV sequences in Banawá form diphthongs, except for those in monosyllabic words, where like vowels can follow one another. We can therefore say that there are no vowel length contrasts in Banawá. The sequences of identical vowels that do occur in monosyllabic words do not contrast with short vowels in these words, as short vowels constitute only a single mora and a monosyllabic word, like all Banawá words, must have at least two moras, in accordance with (4).

Banawá syllables can contain nearly all the possible sequences of unlike vowels. Of the 12 possible sequences, **ie**, **ia**, **iu**, **ei**, **ea**, **eu**, **ai**, **ae**, **au**, **ui**, **ue**, **ua**, only two, **ea** and **ae**, do not occur. One way of describing this situation is by (5):

(5) Except for monosyllabic words, adjacent vowels in the same word cannot have the same sonority.

This would still forbid long vowels, as like vowels obviously have the same sonority. If we also define sonority to make it clear that there is no difference in sonority between **e** and **a**, then (5) would constrain Banawá so that it would not have the sequences **ea** and **ae**. Our definition of sonority would also have to make it plain that **i** has more sonority than **u**, so that sequences such as **iu** and **ui** are allowed, and that both **i** and **u** have less sonority than **e** and **a**, allowing the other possible sequences of vowels.

There are, however, two objections to this proposal. Firstly, if we define sonority in terms of acoustic intensity, then the use of a ternary scale in which **(e = a) > i > u** requires special motivation. Why should there be a division between **i** and **u**, but not between **e** and **a**? Secondly, it provides little in the way of an explanation of the observed facts. Why should a difference in sonority (whatever that is) be required for a sequence of vowels? There is another possible explanation for the lack of sequences such as **ea** and **ae**, which appear to be generally disfavored among the world's languages. It may be that the two elements in each of these diphthongs do not differ sufficiently in quality (a much more salient property than sonority). A related possibility is that **ea** and **ae** are not sufficiently contrastive with **ia** and **ai** respectively, and that vowel systems tend to have only the pairs that contrast most in vowel quality. We could rephrase (5) so that it would provide an alternative explanatory account of the facts:

(5') Except for monosyllabic words, adjacent non-high vowels cannot occur in the same word.

Thus the sequences of [high] vowels **iu** and **ui** are permitted, but the sequences **ea** and **ae**, which contain the vowels that are most alike in quality are not permitted.

We must, however, note that this statement is not a fully satisfactory account of the facts, as it does not mention the systemic pressure to avoid having the similar pair **ea** and **ia**, and the other similar pair **ae** and **ai**. We want to describe the phonetic structure of a language in a way that explains why it is natural for things to be as they are. When we are trying to describe the phonetic structure in terms of a simple set of statements, we have a problem. A single constraint may account for the facts, but there may be two or more reasons for the observed phenomena. There are nearly always systemic (paradigmatic) forces constraining the possibilities that can occur at a given point in the structure, and syntagmatic forces constraining possible sequences. Thus a language may avoid having **ae** and **ai** because they are too similar, and also fail to have sequences such as **ae** because the two vowels are too alike as indicated by (5'). In an ideal explanatory description we would have both types of statements, those giving the syntagmatic constraints ensuring that diphthongs consist of elements with sufficiently different qualities, and those giving the paradigmatic constraints ensuring that all the diphthongs that can occur at a given point in the system are sufficiently different from one another. However, general principles of simplicity do not encourage us to describe the situation in terms of two formal statements when one will do. Recent work by Flemming (1995) has discussed the necessity of considering paradigmatic pressures. But at the moment we do not have a way of combining statements about syntagmatic and paradigmatic pressures into a single theory applicable to the description of the phonetic structures of a language.

Suprasegmental structure

Banawá has a very simple suprasegmental structure. Neither tone nor length play any role in distinguishing words, and stress is entirely predictable. Stress in Banawá occurs on the first mora in a word, and on every other mora after the first. The only exceptions to this rule are words with three or more moras beginning with a vowel. In such words the onsetless first vowel is not counted as a mora in the stress assignment process. A formal statement of the stress rule would embody two additional constraints:

(6) An initial vowel in a word cannot form a mora in words with three or more moras.
(7) Stress occurs on alternate moras within a word, beginning with the first mora.

These constraints are exemplified by the words in Table 9. Stress falls on the first vowel of the two syllable words in (a) and (b). It falls on the first and third vowels in the three and four syllable words in (c) and (d), and on the first, third and fifth vowels in the five syllable word in (e). Note that stress falls on the first vowel of the two mora word in (a), but in the longer words in (f) and (g), the initial vowel is skipped, so that stress falls on the second vowel in (f) and on the second and fourth vowels in (g).

Table 9. Words exemplifying Banawá stress placement. Syllable boundaries are shown by a period [.].

a.	**.'u.wi.**	cry
b.	**.'bi.ta.**	mosquito
c.	**.'wa.ra.'bu.**	ear
d.	**.'wa.na.'kuri.**	spider
e.	**.'re.re'u.ka.'na.**	crank
f.	**.u.'wia.**	go out (as of a fire)
g.	**.u.'wa.ri'a.**	one

With the exception of (e), (f) and (g), each of the words in Table 9 has only one mora in each syllable. We can see that stress is assigned by reference to moras rather than syllables by considering the additional words that have syllables with two moras given in Table 10.

Table 10. Further words exemplifying Banawá stress placement. Syllable boundaries are shown by a period [.].

a.	**.'kai.'ɟa.ra.**	to take pride in oneself
b.	**.'reu.'ka.na.**	to stir
c.	**.'su.ki'a.**	dark

In (a) **'kai** must be a single syllable. Part of the evidence for this is that there cannot be two syllables, **'ka.i** because **i** cannot form a syllable by itself; all word internal syllables have to have an initial consonant. In so far as it provides additional confirmation, we can also note that the **'kai** syllable in (a) sounds much like the English syllable [**kaɪ**] as in 'kite'. The second syllable, **'ɟa**, is stressed, so stress cannot be said to be on alternate syllables; with alternate syllable stress the syllable **ra** rather than the syllable **ɟa** would have been stressed. All this shows that stress assignment must be mora based rather than syllable based. The same arguments apply in (b). We must interpret **'reu** as a single syllable, as **u** cannot be a separate syllable without an onset. As the following syllable **ka** is stressed, stress must be on alternate moras not alternate syllables.

Similarly, the two vowels at the end of (c) must belong to the same syllable. In this case the stress falls on the second mora in this last syllable of the word.

These observations on Banawá stress were verified by tests with three of our six Banawá speakers. Before we can discuss this work we must be fully explicit concerning the nature of stress. All aspects of speech are the results of muscular actions, and stress is no exception. In our view stressed syllables are produced by increased activity of the respiratory muscles, typically the internal intercostal muscles (Ladefoged 1967). One of the few arguments against this notion is that of Adams (1979), who makes it quite clear that she disagrees with Ladefoged (1967). Adams recorded emg activity from the internal intercostals and says "It should be emphasized ... that in no case ...were localized burst of internal intercostal activity found to correlate with the incidence of stressed syllables." (Adams 1979:117) It is certainly true that one cannot see bursts of internal intercostal activity occurring on stressed syllables in Adams' published data. But this does not mean that such activity did not occur in the utterances she recorded. Her data reflects the activity of a large number of motor units firing almost simultaneously, so that the action potentials are superimposed on one another in the record. In the studies summarized by Ladefoged (1967) fewer units were recorded, so that the separate bursts of activity are more apparent. One cannot see bursts of internal intercostal activity in Adams's records because of the density of the activity and the time scale of the recordings. But that does not mean that they are not there. One cannot see the formant structure in the waveforms that accompanied her emg data; but no one would doubt that the formants are there and are varying in accordance with the sounds being produced. It just takes some signal processing of the waveform to make the variations evident. In addition there are many other studies supporting the notion that stressed syllables involve extra respiratory effort (though not necessarily of the internal intercostal muscles). Stressed syllables have a higher subglottal pressure (Ladefoged 1967). There is no way to explain peaks of subglottal pressure in different places in noun verb pairs (recorded within frames) such as "an overflow" and "to overflow" without postulating controlled variations in respiratory activity that can be correlated with the positions of the stressed syllables.

If we consider stress to be a motor gesture, then one way of determining where stressed syllables occur is to get speakers to tap on something while they speak. This activity is a very basic gesture that can be easily coordinated with other movements. Speakers can often beat time on the stresses, or nod their heads, or make some other gesture such as tapping on a table. It is difficult to get speakers to tap only on the unstressed syllables, or on every occurrence of some other phonetic feature, such as whenever the tongue touches the alveolar ridge, or every time the lips close. But speakers usually have no difficulty in tapping on stressed syllables. At least in a language that has lexical stress, as long as a word is said with its normal rhythm, beats will occur on stressed, and only on stressed, syllables.

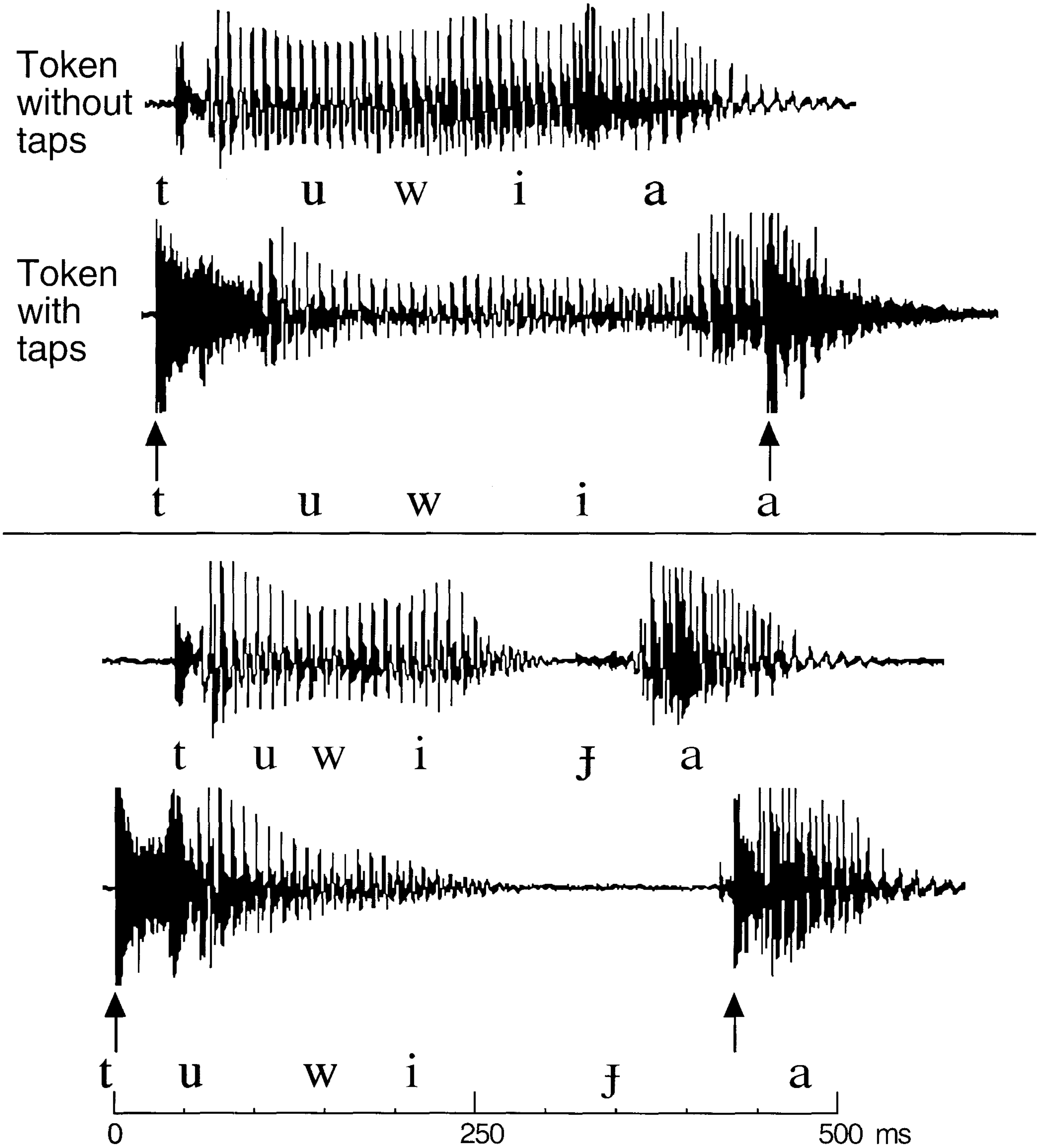


Figure 10. Two tokens of each of two Banawá phrases **'tu.wi'a**, 'to allow not to accompany' and **'tu.wi.'ɟa,** 'poorly done'. In each case the second token is accompanied by taps on a glass near the microphone. The arrows mark the onsets of these taps.

We instructed our main Banawá speakers to tap on stressed syllables in Portuguese, a language with which they were familiar, demonstrating on Portuguese examples what was required, and getting them to practice in that language first. We then asked them to say a number of Banawá phrases, each time saying the phrase twice in a natural way and repeating it a third time while tapping on appropriate syllables. Occasionally, when asked to tap in this way, the speakers responded by trying to tap on every syllable; but when they did this they altered the rhythm of the sentence they were producing, saying it syllable by syllable. We would then ask them to say it at a regular speed, and to tap on just the principal parts. They responded by producing the required sentence, and repeating it with taps on the stressed syllables with no changes in rhythm. We used our two main speakers to explain the task to a third speaker who was not so familiar with Portuguese. Unfortunately circumstances did not permit us trying this task with other speakers.

Figure 10 exemplifies this process. The upper part of the figure is the waveform of the phrase **'tu.wi'a**, 'to allow not to accompany'. Below is another token of this phrase, said while tapping on a glass which produced a sharp ring (and overloaded the recording). The lower part of the figure is the phrase **'tu.wi.'ɟa**, 'poorly done'. This part of the figure also shows one token without taps and below it one token with taps. In each case, the token accompanied by taps was said slightly more forcibly, and therefore slightly more slowly. In the first phrase the second tap occurs on the last vowel in this two syllable word, which is the third mora in the word. In the second phrase the tap occurs as the palatal stop is released, making this phrase have three syllables, with stresses on the first and last. This same pattern was evident in all three of the speakers who were recorded while producing taps.

In all the Banawá phrases we tested in this way, speakers tapped on the mora predicted by the account of the location of stresses given above. The phrases used in this part of the study were not controlled for segmental content in ways that would have enabled us to use acoustic measures such as duration and intensity as correlates of stress. We hope that a later study of additional material will be able to investigate this topic. On the basis of our present data the constraint based description of stress placement seems to provide a good indication of speakers' behavior.

Conclusion

This is a far from complete account of the phonetic characteristics of Banawá. We have described the qualities of the Banawá vowels, noting the slight allophonic differences between vowels in a coronal and labial context; but many other small allophonic differences could have been noted. We have described the places of articulation of each of the consonants, and given an account of the variations in VOT. The manner of articulation has also been documented for most of the consonants. But little has been said about allophonic variations. The structure of syllables and words has been delimited, and the location of stressed syllables verified; but there has been no analysis of intonational features. Much more work remains to be done. But we have been able to present the major phonetic characteristics of this endangered language.

References

Adams, C.: English speech rhythm and the foreign learner. (The Hague, Mouton. 1979.)

Buller, E.; Buller, B.; Everett, D.: Stress placement, syllable structure, and minimality in Banawá. Int. J. Amer. Ling. 59: 280-293 (1993).

Everett, D.: Prosodic levels and constraints in Banawá and Surwaha. (Rutgers Optimality Archive, Rutgers 1996.)

Everett, D.: Sistemas prosodicas da familia Arawa; in Estudos Fonologicos de Linguas Indigenas Brasileiras (Editora da Universidade Federal do Rio de Janeiro 1995).
Fischer-Jørgensen, Eli. Acoustic analysis of stop consonants. Misc. Phon. 2: 42-59 (1954).
Flemming, E.: Auditory features in phonology. UCLA Ph.D. Dissertation (1995).
Grimes, B.: Ethnologue. (Dallas: Summer Institute of Linguistics. 1992.)
Johnson, K.; Ladefoged, P.; Lindau, M.: Individual differences in vowel production. J. Acoust. Soc. Amer 94: 701-714 (1993).
Ladefoged, P.: Phonetic prerequisites for a distinctive feature theory; in Papers in Linguistics and Phonetics in Memory of Pierre Delattre. (The Hague, Mouton. 1972.)
Lindblom, B.: Models of phonetic variation and selection. PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm) 11: 65-100 (1990).
Lisker, L.; Abramson A.: Cross-language study of voicing in initial stops: Acoustical measurements. Word 20: 384-422 (1964).
Peterson, G.; Barney, H.: Control methods used in a study of the vowels. J. Acoust. Soc. Amer 24: 175-184 (1952).
Stevens, K. N.: On the quantal nature of speech. J. Phonet. 17: 3-46 (1989).

Acknowledgements

Many thanks to our ever helpful Banawá consultants. We are also grateful to friends at the SIL Center at Porto Velho, Brazil, for their hospitality and help. Helpful comments on earlier versions of this paper were made by Robert Kirschner, Donca Steriade, Richard Wright and two anonymous reviewers. This research was supported by NSF Grant 9511118 to Peter Ladefoged and Ian Maddieson, amd NSF Grant 9631322 to Daniel Everett.

Wari' phonetic structures

Margaret R. MacEachern, Barbara Kern and Peter Ladefoged

The Wari' language is spoken along the Pacaas Novas river in Western Rondonia, Brazil (see Figure 1). Everett and Kern (in press) state that there are about 1300 speakers of the language; a 1986 Summer Institute of Linguistics survey places the number at between 990 and 1147 (Grimes 1988). Voegelin and Voegelin 1978 place Wari' (Pacas Novas) in their Arawakan: Chapacura-Wanhaman: Guapore: Madeira grouping, along with the extinct languages Jaru and Tora(z), and extant Urupa (Txapacura). Grimes 1988 categorizes Wari' (Oro Wari, Pakaasnovos, Jaru, Uomo, Pakaanovas, Pacaas-Novos, Pakaanova, Pacahanovo) as a Chapacura: Madeira language, and describes it as isolated. (Note also that Grimes lists Jaru as an alternate name for Wari', rather than as a distinct language.) Everett and Kern (in press) is the only major publication on this language. Ladefoged and Everett (in press), commenting on a single sound of Wari', observe that another Chapacuran language, Oro Win, was spoken in an adjacent area; there remain only five speakers of this language in Brazil, and perhaps a few more in Bolivia. Ladefoged and Everett note that the only other living Chapacuran language is Moré, which is spoken by a single elderly speaker.

Figure 1. The location of the Wari' language.

The data for this study consist principally of recordings of 6 male and 6 female speakers made in Guajara Mirim in July 1995. The words and phrases were elicited by the second author, who is a fluent speaker of Wari'. The recordings were made on a DAT recorder, using a close-talking, noise-canceling microphone for each speaker individually. The frequency response was substantially flat throughout the audible range, and the signal/noise ratio was nearly always greater than 45 dB. (The exceptions were due to occasional uncontrollable outside noise, and the fact that two of the women spoke very quietly; nevertheless, the recordings analyzed always had a signal/noise ratio better than 40 dB.) All recordings were made at 48,000 samples/second; these were then down-sampled to 8,000 samples/second for acoustic analysis, so as to provide greater precision in the frequency scale (see Ladefoged 1996:177 for an explanation of this technique).

Vowels

In this section, we discuss several topics: the qualities of stressed vowels, the existence of contextually-determined allophones of stressed vowels, the qualities of stressless vowels, and the occurrence of vowel sequences in the language.

The vowel inventory of Wari' is large for a lowland Amazonian language (Everett and Kern [in press]), and Wari' vowels are unevenly distributed in the vowel space. Wari' has five vowels, **i, e, a, o, y**, plus an additional vowel, **ø**, which is comparatively uncommon and occurs only after a limited range of consonants. (The front rounded vowel **y** is transcribed **ü** in Everett and Kern [in press].) This inventory is exceptional in two respects. First, a study of several large typological surveys (Ruhlen 1975, Crothers 1978, and Maddieson 1984) indicates that the vowel inventory of Wari' has not been reported for any other language. Crothers 1978 reports the most common six-vowel inventories to be {**i, e, ɛ, u, o, ɔ**} and {**i, ɛ, ɨ, a, u, ɔ**}. Typologically, the existence of even one front rounded vowel in a six-vowel system is unexpected; the existence of two is remarkable.

Second, the Wari' vowel inventory seems to ignore the principle of dispersion argued for by Crothers 1978, Lindblom 1986, 1990, and others. This principle is differently formulated by different authors, but essentially it says that vowels will be widely and evenly distributed in the vowel space. For example, Lindblom 1990 provides a model for determining the optimal vowel system, given an inventory of a particular size. Lindblom's model projects {**i, a, u, ɛ, ɔ, ʉ**} as the favored six-vowel inventory; again, note the lack of front rounded vowels.

It is true that the vowel **ø** is limited in its occurrence and distribution within the language; however, even the remaining five vowels, if considered as a group, are unexpected on typological and theoretical grounds. This is largely because of the inclusion of **y**.

Stressed vowels

Table 1 gives examples of the five vowels of Wari' that can occur after each of the possible syllable onsets. Wherever possible, the vowel is in a final open syllable, where it is stressed. Asterisks denote places in which a particular vowel does not occur after a certain consonant in stressed position. Table 2 exemplifies the sixth Wari' vowel, **ø**, which occurs only after **t, k, ʔ, m, h, j**. Many (but not all) of these words appear to have an onomatopoeic basis.

Formant charts showing the qualities of these vowels are given in Figures 2 and 3. The formant values were recorded with the aid of linear predictive coding (LPC) and fast Fourier transform (FFT) spectra from a Kay Elemetrics Computer Speech Laboratory (CSL). Formant measurements were taken in the middle of the steady-state portion of the vowel, or simply in the middle of the vowel itself, if no steady-state portion was present. The filter order was set to 10 poles for analyzing the speech of the female consultants, and 12 poles for analyzing the speech of the male consultants. In these and all other charts in this paper, the vowels are plotted using the UCLA program PlotFormants; the axes on these charts represent Hz on a Bark scale. Figures 2 and 3 show F1 vs. F2; the points plotted represent the mean formant values for that vowel for each speaker. Ellipses around the mean are drawn with axes on the first two principal components, enclosing all points within two standard deviations of the mean.

Table 1. Vowels in final, stressed, syllables of words.

	i	**e**	**a**	**o**	**y**
p	**nopi** bee (sp.)	**ʔan pe** to put down	**papa** stingray	**pok pok** to boil	**kopy** my manioc
t	*	**kote ka** his father	**pita** fish (sp.)	**kawotoʔ** clay pot	*
k	**koki** its thigh	**ʃikeʔ** shell corn	**koka** fish (sp.)	**koko** basket type	**koky** my blood
kʷ	**tokʷi** its seed	**tokʷe** Brazil nut	**tokʷa** corn drink	*	*
ʔ	**ʔiʔ** to tear	**maʔe** OK (fem.)	**kaʔa** bird (sp.)	**toʔo** feminine name	**ʔaʔy** feminine name
m	**komi** its water	**teme** ant (sp.)	**hyma** lizard	**kamo** bird (sp.)	**ʃymy** my heart
n	**pe ni** to be separate	**mene** my thing	**wina** my head	**wino** cashew	**myny** my belly
ʃ	**kaʃiʔ** to be sick	**ʔoʃe** fruit (sp.)	**kaʃaʔ** single man's bed	**naʃoʔ** fish (sp.)	**kyʃy** my machete
h	**hihi** owl (sp.)	**hehe** to hesitate	**toha** to shine	**maho** vulture	**oɾo kohy** cedar tree
ʍ	*	**ʍet** to approach	**ʍap** to be fast	*	*
ɾ	**naɾi** to be related	**kʷeɾe** my body	**ʔaɾa** to do	**noɾo** to look	**nyɾy** to blow
w	**tawi** honey	**towe** to be fat	**kawa** toy arrow	**kowo** frog (sp.)	**ʔawy** toucan
j	**maji** let's go	**ʔje ʔjeʔ** hawk (sp.)	**mija** to be much	**wajo** hawk (sp.)	**kyjy** bird (sp.)

Table 2. Words illustrating the sixth Wari' vowel, **ø**.

t	**k**	**ʔ**	**m**	**j**	**h**
tokoɾom mao tø	**køk**	**koʔøk**	**kamø**	**jøk**	**ʔøhø**
masculine name	fish (sp.)	fish (sp.)	water rodent (sp.)	to push	to cough

In the typical case, the speaker means were calculated on the basis of six tokens (two tokens of three words) for the vowels **a**, **e**, **i**, **o**, **y** and twelve tokens (two tokens of six words) for the vowel **ø**. (The words used for measurements of **a**, **e**, **i**, **o**, **y** are those in Table 1 that illustrate the vowel after **p**, **k**, and **ʔ**; the words used for measurements of **ø** are those in Table 2.) However, one female speaker provided only one token of each word, and there were also 9 unanalyzable or mispronounced tokens from other speakers. A word was considered to have been mispronounced if its vowel was located in the F1—F2 space at a point more than three standard deviations from the mean of other tokens of this vowel as spoken by this speaker. This happened on 6 occasions, presumably because the speaker misunderstood the word that was required.

The distribution of vowels in the Wari' vowel space is surprising. Both **y** and **ø** are definitely front vowels. As a result, instead of being evenly distributed, four of the six vowels cluster about the upper front regions of the vowel space. A symmetric distribution would have required the high rounded vowel to be back rather than front. Clearly, the Wari' data we have presented here are at variance both with the findings of crosslinguistic surveys such as those referred to above, and with the theoretical models proposed to account for such data.

Crowded inventories raise the problem of maintaining distinctions among vowels of different qualities. Figures 2 and 3 show that **a** and **o** are distinct from **i**, **y**, **e**, and **ø**, for all speakers. **ø** is also distinct from **i** for all speakers. However, for male speakers, **i** and **y** overlap, as do **e** and **ø**; **y** also overlaps with **ø** and (very slightly) with **e**. Among the female speakers, **y** overlaps with **ø**, **e** and **i.** There is also a slight overlap between **e** and **i.**

The vowel **y** was observed during recording to involve lip-rounding, which has a significant effect on F3 in high front vowels. Figures 4 and 5 show F1 vs. F3 plots for male and female speakers, respectively. The tokens shown are the same as those reported above, except that there was one additional token for which it was not possible to locate a value of F3. It may be seen that **i** and **y** are clearly distinguished by the location of F3. Note that there is still a considerable amount of overlap between **e** and **ø** for the male speakers. This may be related to the low functional load borne by the **ø** vowel. (We have omitted **o** from consideration in these F3 plots, as this vowel is well distinguished from the other five vowels in terms of its F1 and F2 values).

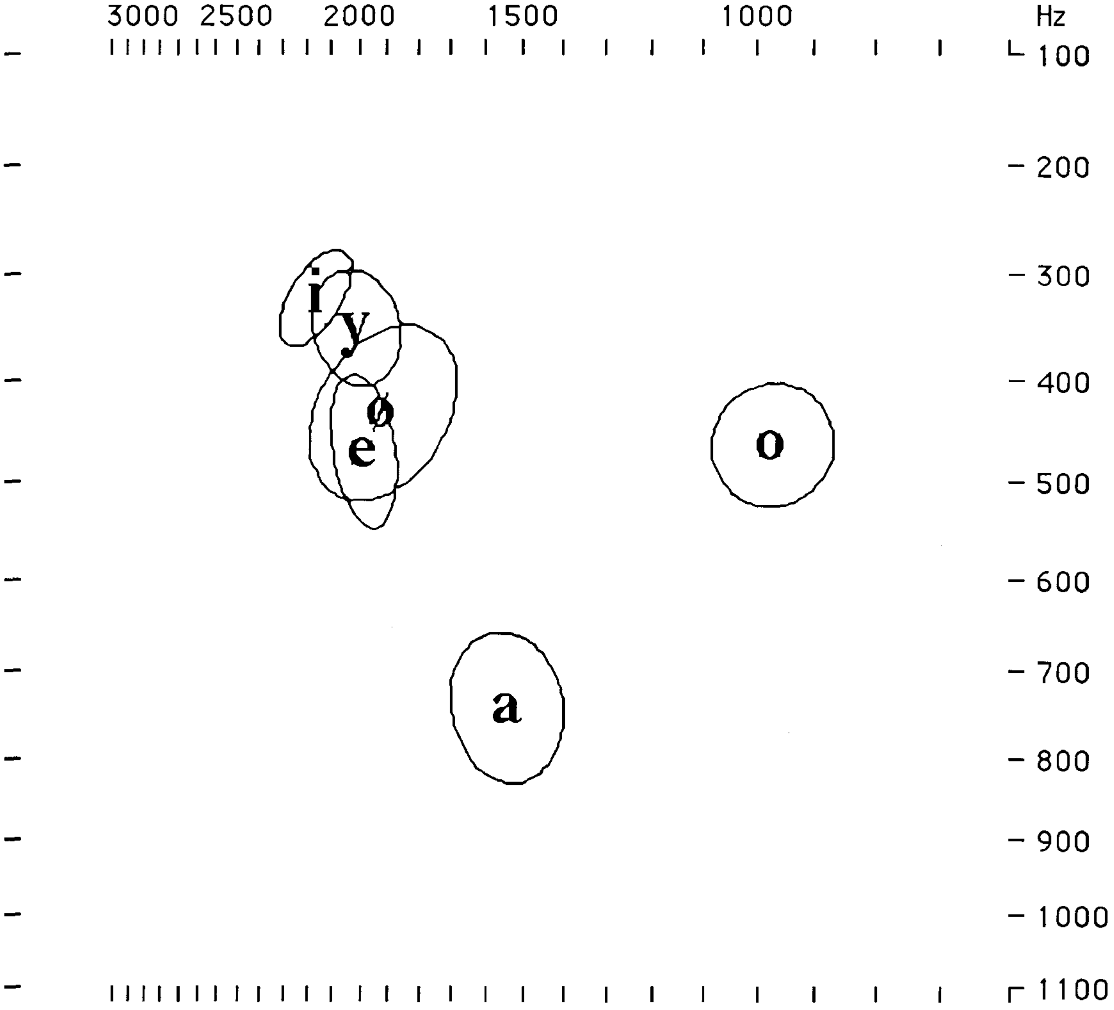


Figure 2. F1 vs. F2 of stressed vowels as produced by 6 male speakers. See text for a description of the scaling procedures.

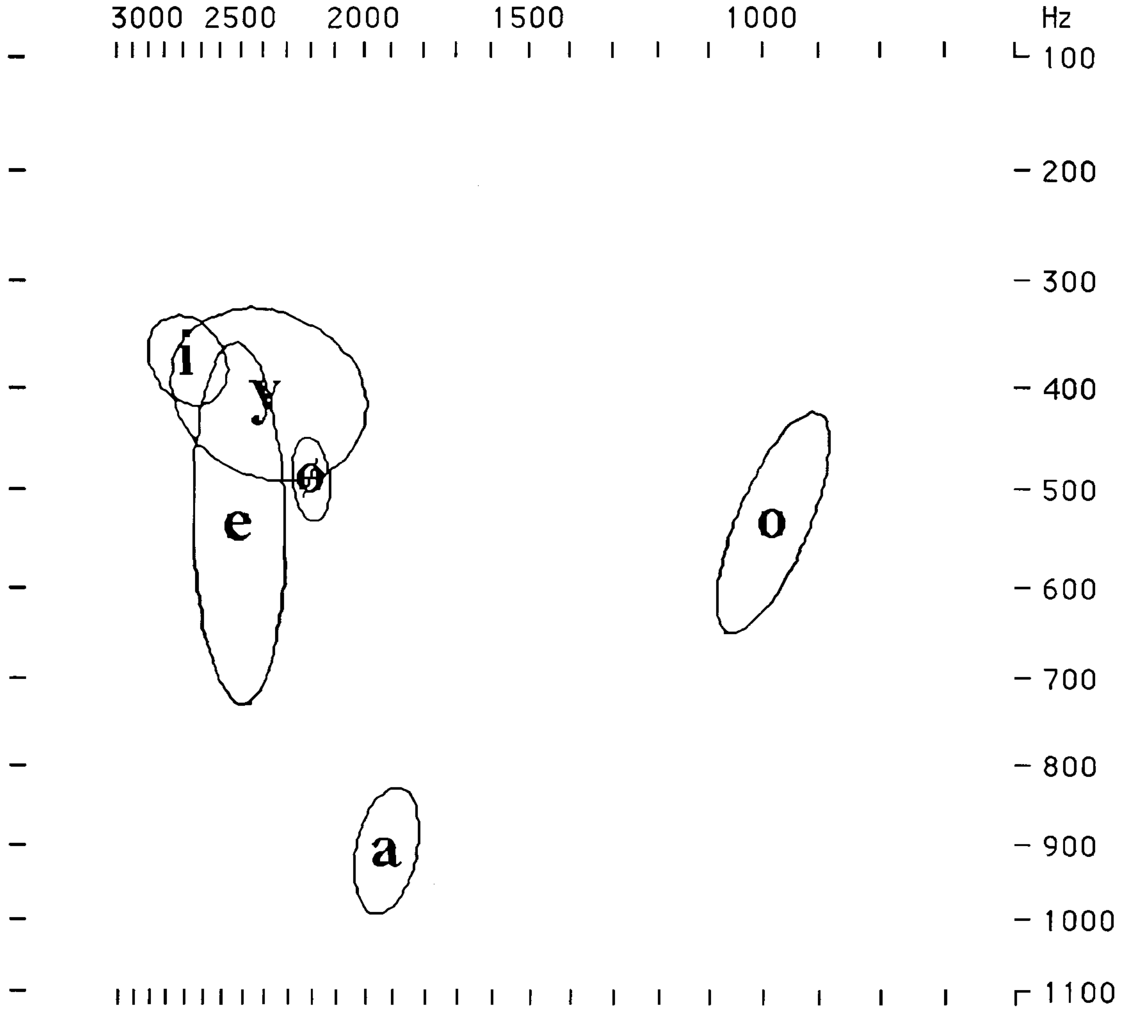


Figure 3. F1 vs. F2 of stressed vowels as produced by 6 female speakers.

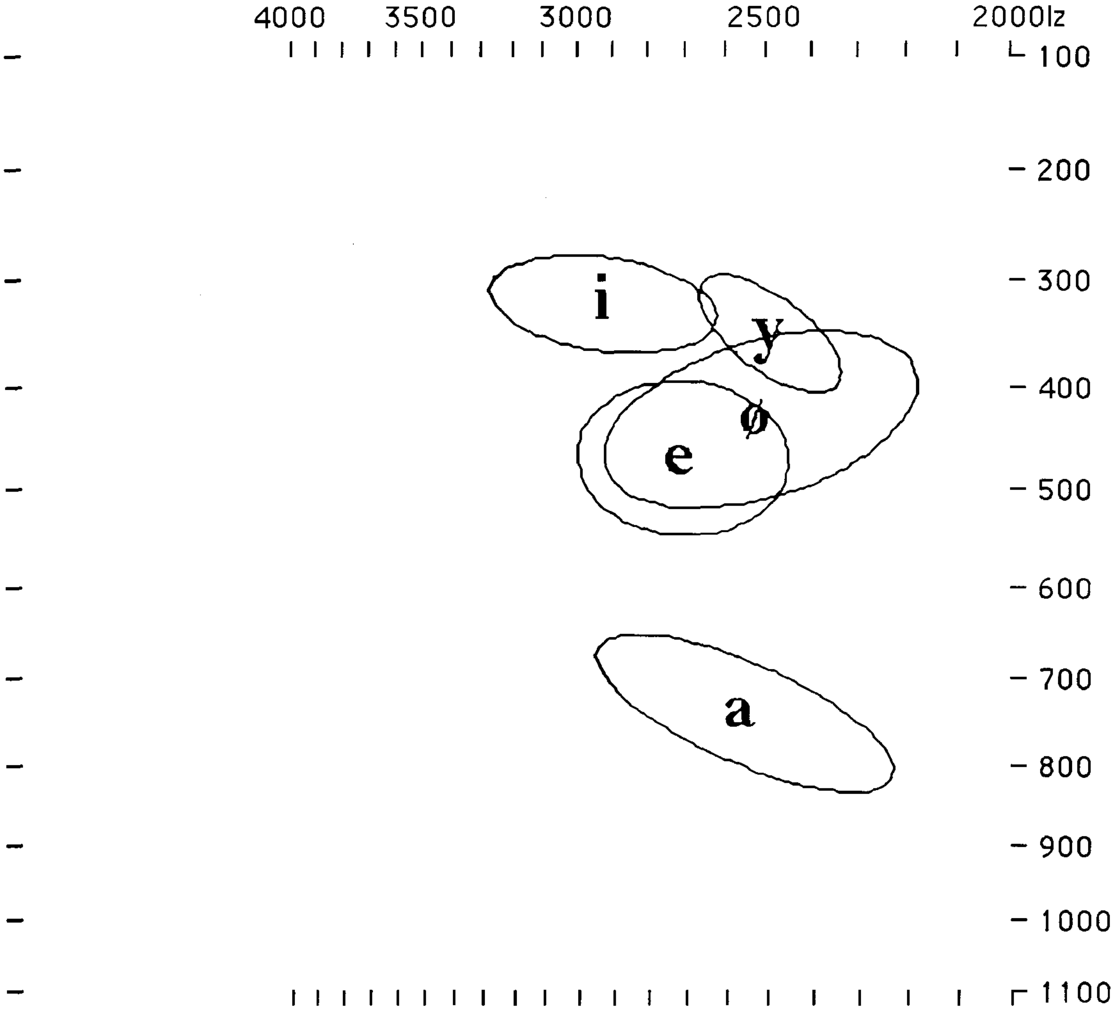


Figure 4. F1 vs. F3 of stressed vowels as produced by 6 male speakers.

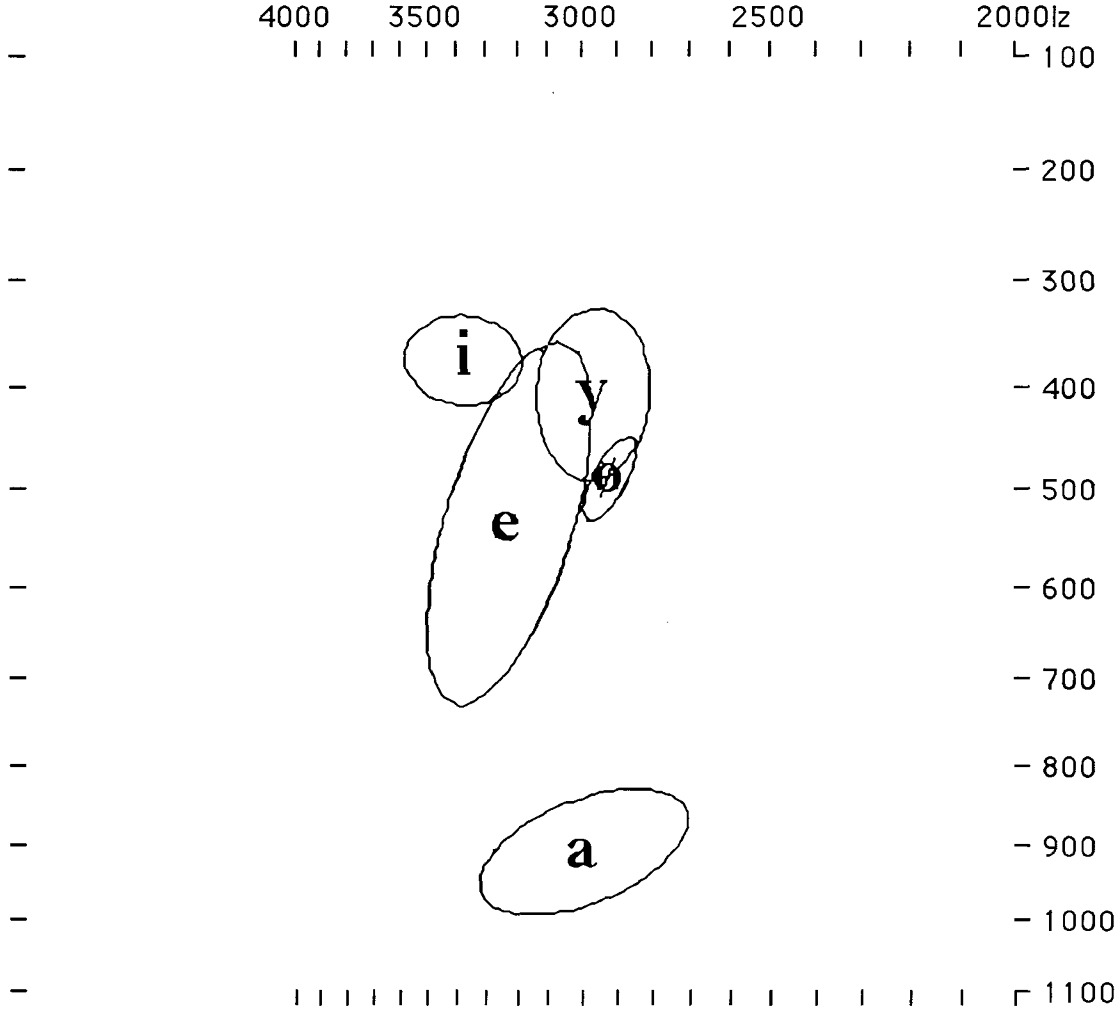


Figure 5. F1 vs. F3 of stressed vowels as produced by 6 female speakers.

The amount of variation in vowel quality was found to differ greatly across speakers. The most conservative of our twelve consultants kept all stressed vowels distinct, even on an F1 vs. F2 chart, as illustrated in Figure 6. The speaker who showed the most variation provided several sets of overlapping vowels. When the values were plotted on an F1 vs. F2 plot, **i** and **ø** intersected, as did **ø**, **y**, and **e**, as shown in Figure 7. In Figures 6 and 7, each point plotted represents a single measurement. All values (six tokens for each of the vowels **a**, **e**, **i**, **o**, and **y**, and twelve tokens for the vowel **ø**) are present for the speaker represented by Figure 6; one token of **ø** is missing for the speaker represented by Figure 7.

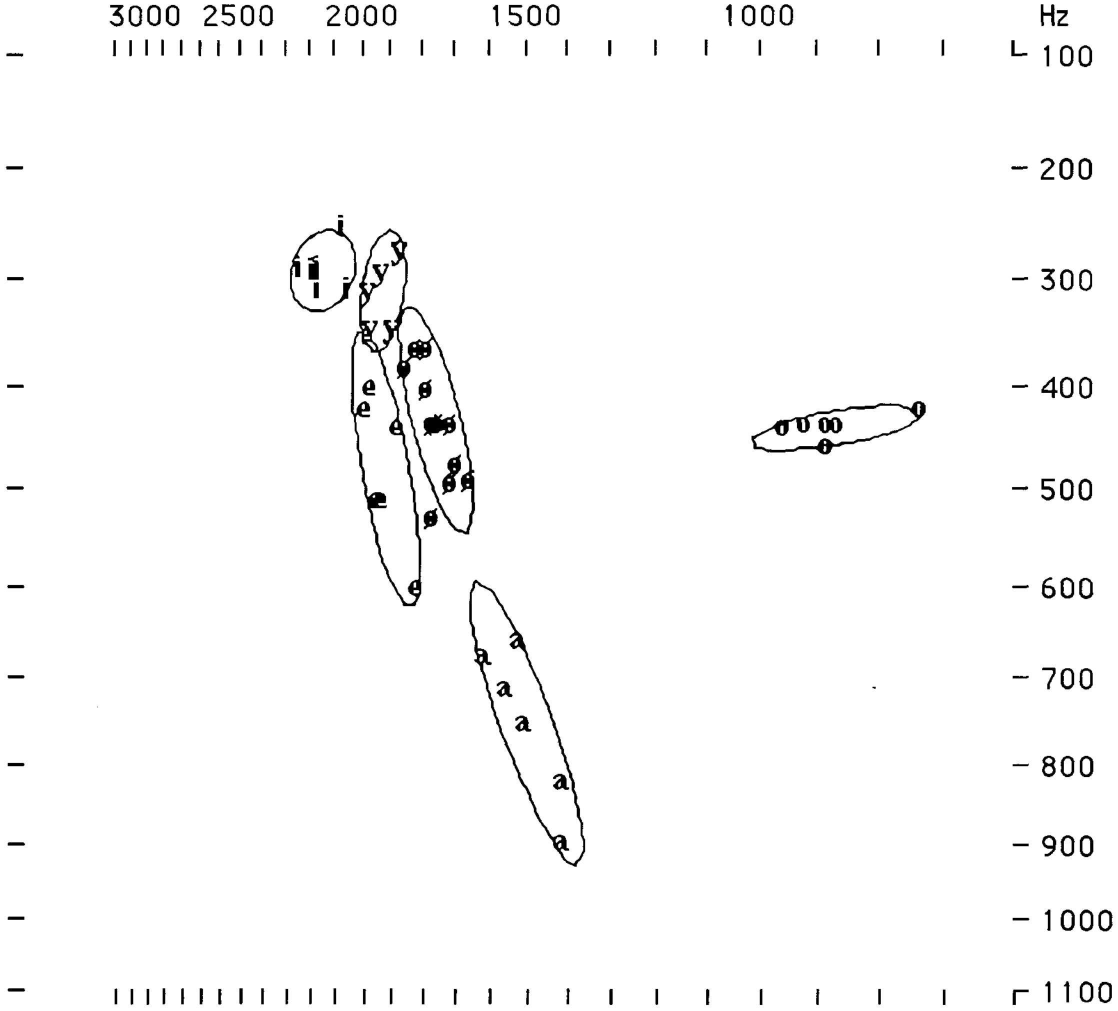


Figure 6. F1 vs. F2 of Wari' vowels in stressed syllables as produced by one (male) speaker.

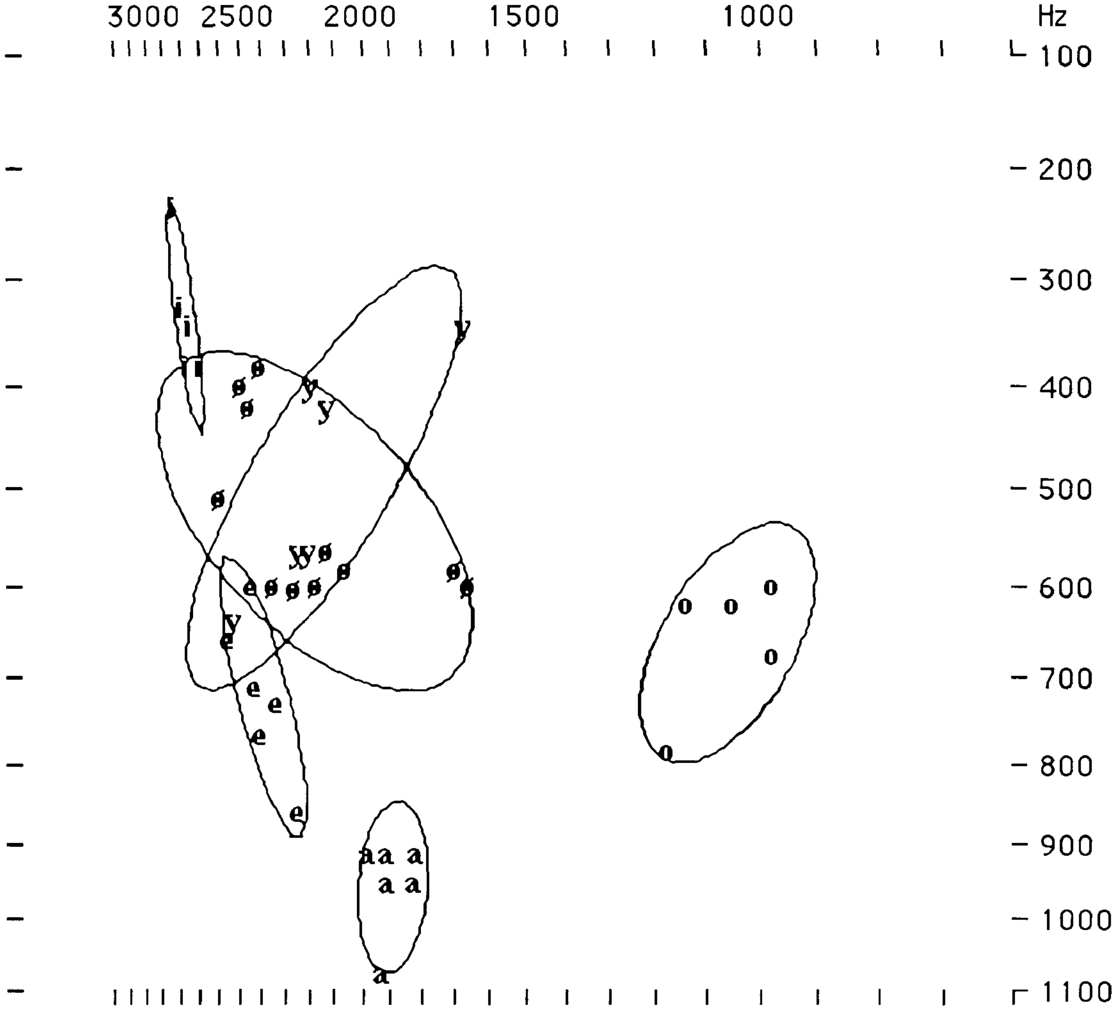


Figure 7. F1 vs. F2 of Wari' vowels in stressed syllables as produced by one (female) speaker.

There are also large differences in the amount of variation in the production of different vowels pooled across speakers (see Figures 2-5). We will now compare the relative amounts of variation in the realization of stressed vowels of different qualities, both within and across speakers, in a quantitative manner. One way of describing this variation is to consider the area of the ellipses. (The reader will recall that these ellipses include all points within two standard deviations of the mean.) Tables 3 and 4 provide rectangles describing this variation; the long and short sides of the rectangles represent the principle axes of the ellipses.

Table 3. Variation in F1 and F2 values of stressed vowels produced by male speakers.

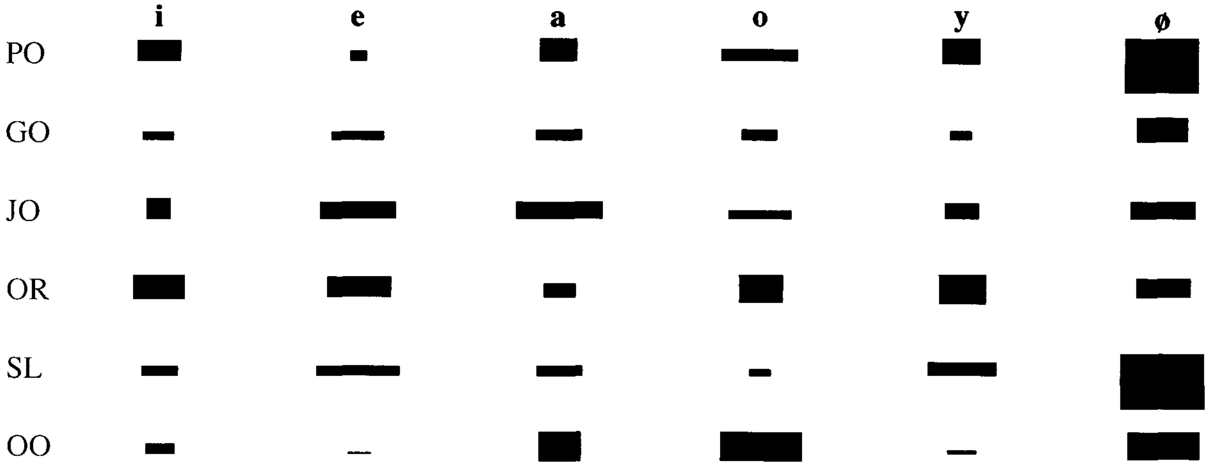
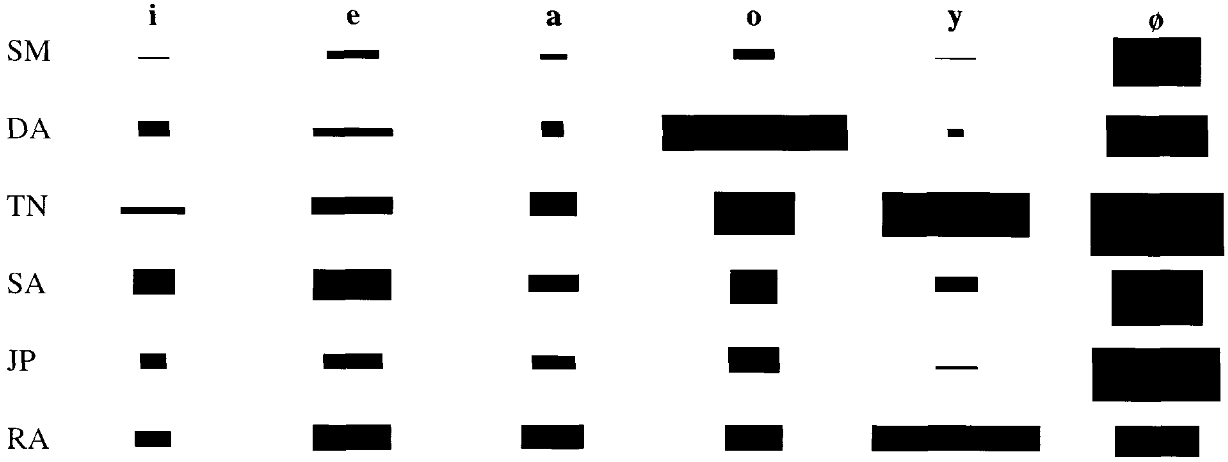


Table 4. Variation in F1 and F2 values of stressed vowels produced by female speakers.



The vowel ɸ stands out as being the most variable of the vowels of Wari'.

Allophones of stressed vowels

Our auditory impression was that some Wari' stressed vowels have centering glides in open syllables. In order to test this hypothesis, first and second formant measurements were made at the midpoints and ends of stressed vowels in 62 tokens. The tokens were produced by three male and three female speakers, randomly chosen, and included representations of each of the six stressed vowels. The direction of formant movement of centralized vowels differs depending on where the uncentralized vowel is located in the vowel space. For the purposes of this test, we use simple formant indices: we assume that raising is inversely proportional to F1, and fronting is inversely proportional to F2.

Centralized **a** will have a lower F1 frequency; the vowel is already central, so we would not expect F2 movement. Centralized **i** and **y** will have both higher F1 and lower F2, as the vowels are lowering and backing. Centralized **e** and **ø** will have just a lowered F2. Centralized **o** will have just a higher F2.

Paired samples t-tests were conducted on the values of F1 at the midpoints and endpoints of **a**, **i**, and **y**. In addition, paired samples t-tests were conducted on values of F2 at the midpoints and endpoints of **o**, **i**, **y**, **e**, and **ø**. Each vowel quality was represented by from 7 to 13 tokens. The results of these tests give some support to the hypothesis that stressed vowels center in the front-back dimension in open syllables; they do not support the hypothesis that these vowels center in the height dimension.

The difference in F1 measurements for **a** was probably significant, but in the wrong direction: we found mean formant values of 842 Hz at the midpoint, versus 902 Hz at the endpoint (t=2.20, p<.05). The vowel **a** appears to be *lowering* somewhat in open stressed syllables, rather than raising. The difference in F1 measurements for **i** and **y** was nonsignificant, whether the data was considered in the aggregate or separated by vowel quality (in all tests, p>.05). We conclude that the high front vowels do not manifest lowering glides in open syllables.

The only F2 measurements that were even probably significant were those for **i**, in which the mean was 2516 Hz at the midpoint versus 2407 Hz at the endpoint (t=2.23, p < .05). The midpoint mean of F2 for **o** was 967 Hz, while the endpoint mean was 1091 Hz. This difference is in the expected direction, but it was not statistically significant (p>.05). The F2 measurements for **y**, **e**, and **ø** also showed nonsignificant differences (in all cases, p>.05). However, as for **o**, the midpoint and endpoint means all showed movement in the centering (in this case, backing) direction: mean midpoint F2 for **y** = 2245 Hz; endpoint F2 = 2156 Hz; mean midpoint F2 for **e** = 2305 Hz; endpoint F2 = 2248 Hz; mean midpoint F2 for **ø** = 2249 Hz; endpoint F2 = 2238 Hz. Finally, the aggregate test showed a significant difference in mean F2 values when the data sets for **i**, **y**, **e**, and **ø** were combined (2333 Hz at midpoint versus 2260 Hz, t=2.02, p<.005). We conclude that it is possible that all front and back vowels undergo centering glides in stressed open syllables, but more data would be needed to establish this statistically.

Another stressed vowel for which allophones have been proposed is **e**. Everett and Kern (in press) suggest that this vowel is somewhat higher and more front (i.e. **ɪ**) before nasals and somewhat lower (i.e. **ɛ**) before oral stops than it is in open syllables and before **ʔ** (where it is **e**). The words in Table 5 illustrate these contexts. Formant charts are shown in Figures 8 and 9. In the typical case, each speaker provided two tokens of each word representing what we might call the **m** tokens (illustrating phonological **e** before an **m**), the **k** tokens (**e** before **k**), and the **e** tokens (**e** in open syllables and before glottal stop). One female speaker, however, provided only one token of each word, and one token from a female speaker was unanalyzable.

Table 5. Stressed **e** in various environments. The symbol at the top of each column indicates the quality of the allophone proposed by Everett and Kern (in press) for this context.

ɪ	e	e	e	ɛ
komem	**ʔan pe**	**maʔe**	**ʃikeʔ**	**ʃek**
deer (sp.)	to put down	OK (fem.)	shell corn	day

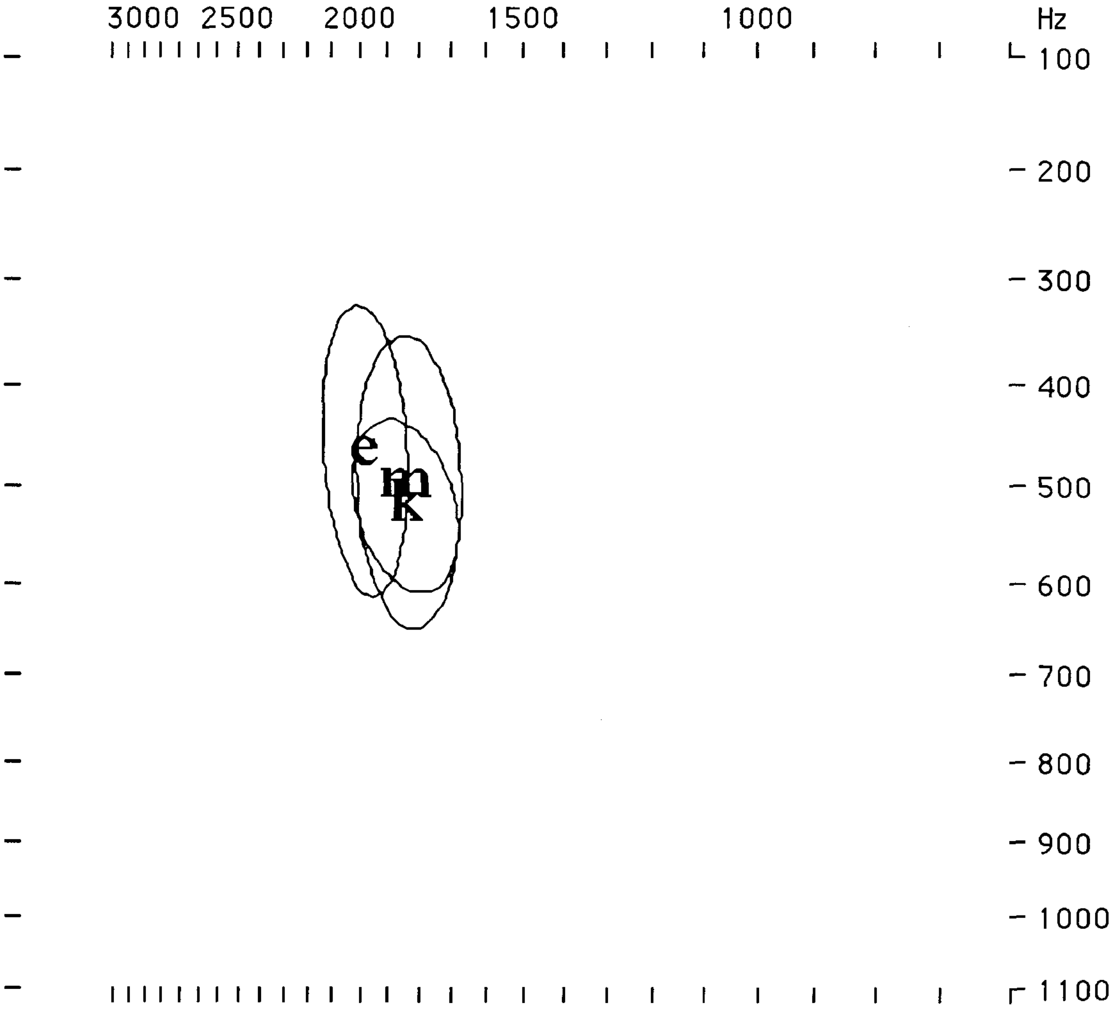


Figure 8. Variation in **e** as produced by 6 male speakers. **e** represents the vowel in open syllables and before glottal stop; **m** represents the production of the vowel before **m**, and **k** represents the vowel before **k**.

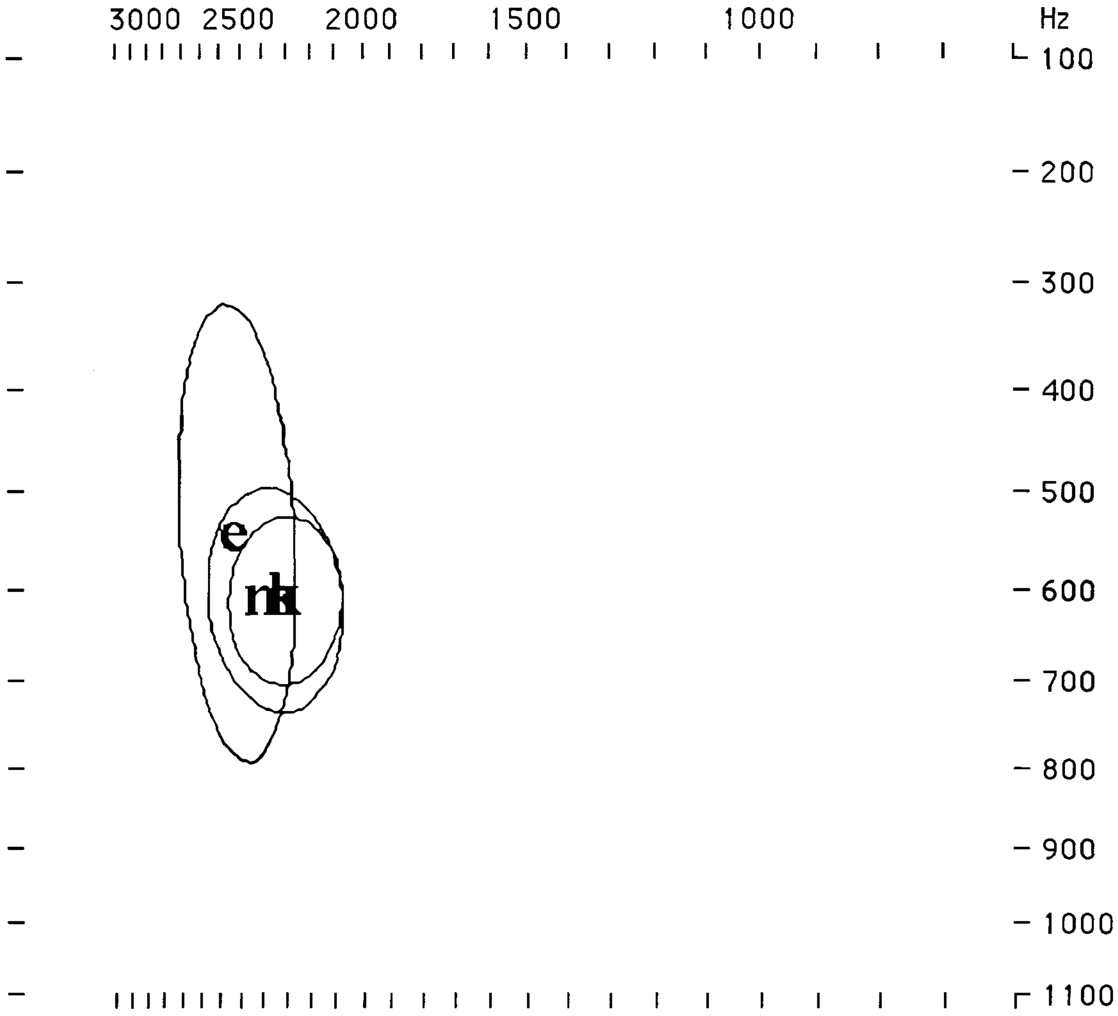


Figure 9. Variation in **e** as produced by 6 female speakers.

For the male speakers, two-samples t-tests assuming unequal variances showed that **e** is significantly higher and fronter in open syllables and before **ʔ** (mean F1 466 Hz) than it is before oral stops (mean F1 521 Hz, t= 2.04, p<.005). The F1 of the **e** tokens was not significantly lower than that of the **m** tokens (p>.05), but the difference between the means was in the expected direction (the mean F1 of the **m** tokens was 498 Hz). The F2 values for the **e** tokens (mean=1980 Hz) were very significantly higher than those of both the **k** tokens (mean=1844 Hz; t=2.12, p<.001) and the **m** tokens (mean=1833 Hz, t=2.12, p<.001).

For the female speakers, the results are similar. The differences between the F1 of the **e** tokens (mean=546 Hz) and that of both the **k** tokens (mean=613 Hz, t=2.02) and the **m** tokens (mean=613 Hz, t=2.03) were only probably significant (p<.05). The F2 values for the **e** tokens (mean=2525 Hz) were very significantly higher than those of both the **k** tokens (mean=2307 Hz, t=2.09, p<.001) and the **m** tokens (mean=2351 Hz; t=2.12, p<.005).

No significant differences were found between the **k** and **m** tokens, either for F1 or F2 values. Thus, we can say that lowered, backed allophones of **e** tend to be produced before nasals and oral stops. Everett and Kern's observation on the allophonic realization of **e** is borne out by the analyses performed here, although not in its smallest details.

Stressless vowels

Unstressed vowels (in the first syllable of each word) are illustrated in Table 6. Formant charts showing the qualities of these vowels are given in Figures 10 and 11. In the typical case, each speaker contributed two values per vowel (one from each token of one word). A total of 17 values (10 for women, 7 for men) were missing in this analysis: one female speaker provided only one token of each word, several vowels were so faint as to be unanalyzable, and two tokens involved mispronunciations.

Table 6. Words illustrating stressless vowels.

i	**e**	**a**	**o**	**y**
pikiɾim	**keɾek**	**maɾam**	**kokoɾok**	**kykyɾyp**
to rock	to see	soft	to shake	grub worm (sp.)

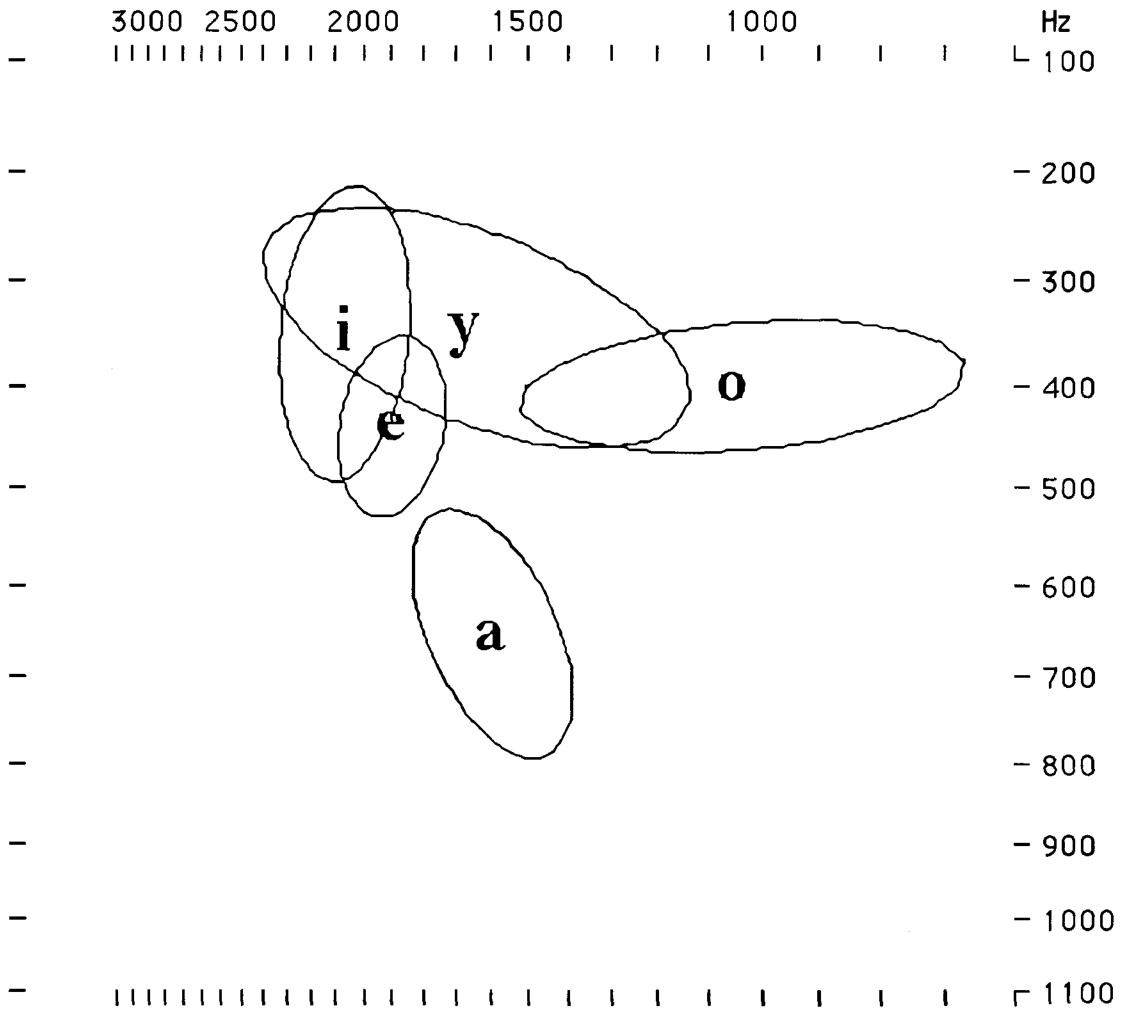


Figure 10. F1 vs. F2 of stressless vowels as produced by 6 male speakers.

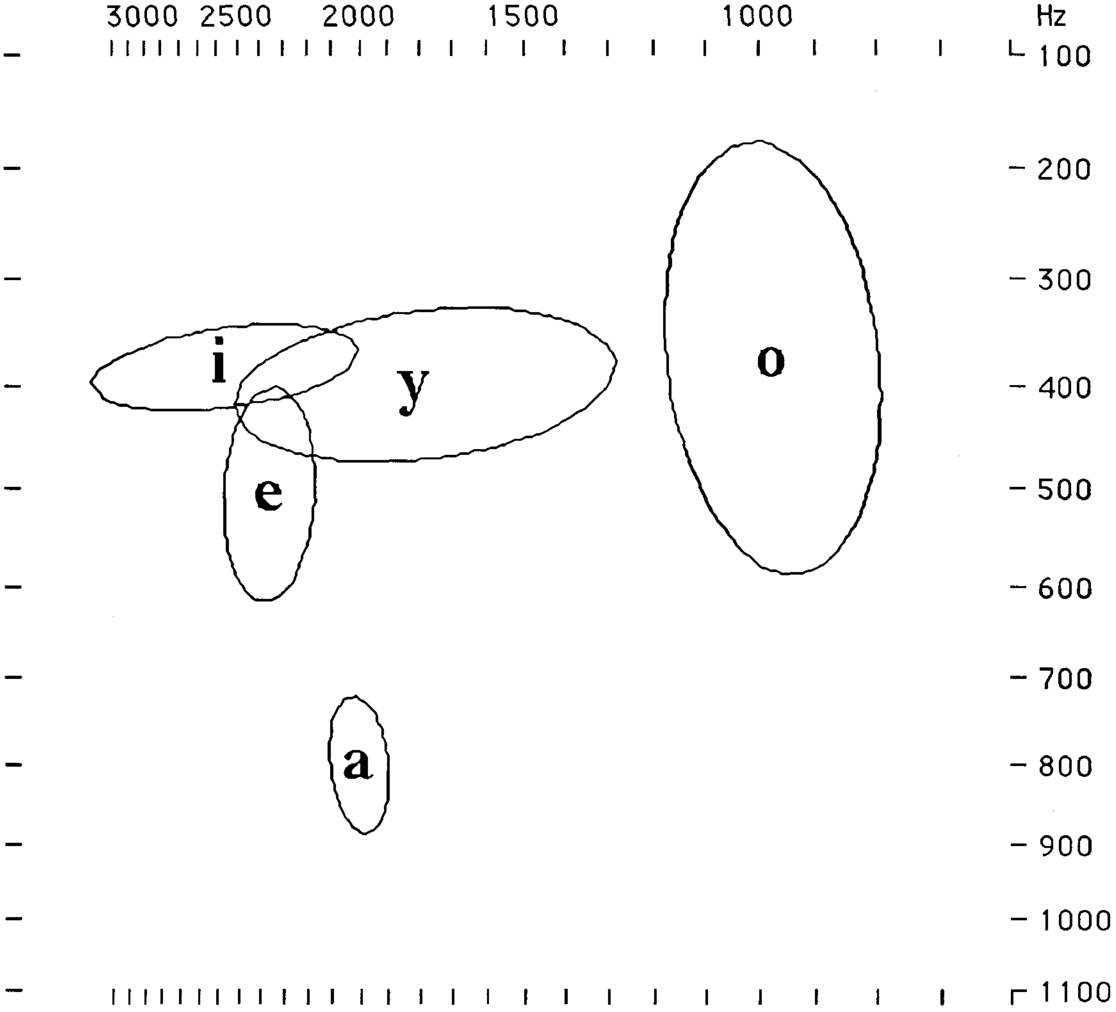


Figure 11. F1 vs. F2 of stressless vowels as produced by 6 female speakers.

Comparison of Figures 10 and 11 to Figures 2 and 3 indicates, as expected, that unstressed vowels are subject to greater variability in formant values, and are more centralized than stressed vowels.

The stressless vowels can also be compared to the stressed vowels in terms of rounding. Formant plots illustrating F1 vs. F3 are shown in Figures 12 and 13. Again, **o** has been omitted to make these figures easier to read.

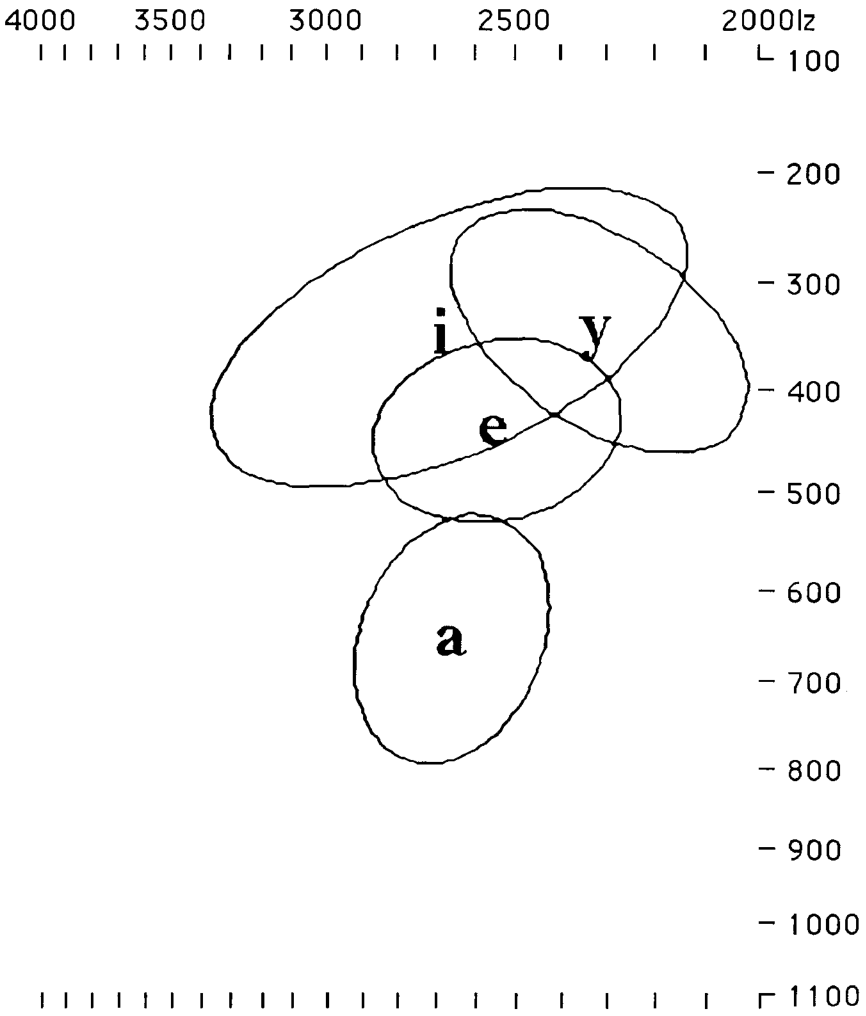


Figure 12. F1 vs. F3 of stressless vowels for 6 male speakers.

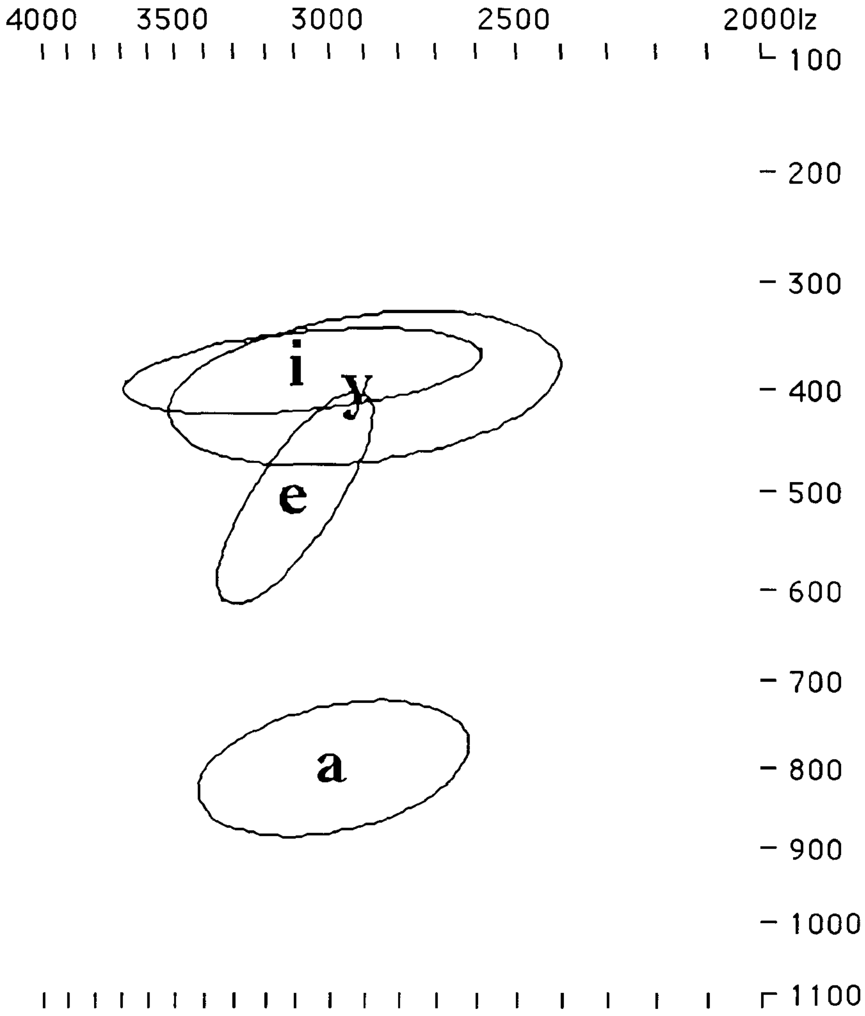


Figure 13. F1 vs. F3 of stressless vowels for 6 female speakers.

Comparison of Figures 10-13 to Figures 2-5 illustrates the reduced effect of F3 in distinguishing among the *stressless* versions of front vowels, when compared to the effect of this same formant in distinguishing among *stressed* versions of front vowels. For example, stressless **i** and **y** overlap for both male and female speakers on both F1 vs. F2 and F1 vs. F3 plots. The similarity in F3 values for **i** and **y**, evident in Figures 12 and 13, indicates that **y** unrounds somewhat under reduction.

As illustrated in Table 7, various sequences of the vowels **i**, **e**, **a**, **o** can occur in Wari', but like vowels do not occur in the same syllable. The vowel **y** can be followed only by **i**, and **ø** does not occur in any sequences, further demonstrating the unusual nature of these two vowels.

Table 7. Words illustrating sequences of vowels.

ao	**eo**	**io**	**ai**	**ei**	**oi**	**yi**
ʃao	**ʃeo**	**ʃio**	**ʃai**	**wijei**	**toi**	**ʃyinʔ**
to pound	to be agreeable	to be cold	to be hot	its blossom	to fly (pl.)	snail
kaoʔ	**jeoʔ**	**ʔioʔ**	**painʔ**	**wijeinʔ**	**noinʔ**	
to eat	my grandfather	louse	to toast	to be bitter	navel	

Consonants

The consonants of Wari' are shown in Table 8.

Table 8. Wari' consonants.

	Bilabial	Dental	Post-alveolar	Velar	Labialized velar	Glottal
Stop	**p**	**t**		**k**	**kʷ**	**ʔ**
Nasal	**m**	**n**				
Fricative			**ʃ**		**ʍ**	
Tap		**ɾ**				
Approximant			**j**		**w**	**h**

There is in addition a special consonant, transcribed **tᵖ** by Everett and Kern (in press), and **t͡ʙ̥** by Ladefoged and Everett ([in press]) which has been observed in only about 25 words in the language. We discuss this sound briefly at the end of this section.

There are no voicing contrasts in Wari'. As is common in such circumstances, the stops, **p**, **t**, **k**, and **kʷ**, are typically voiceless unaspirated and may become partially voiced when intervocalic. The voice onset time (VOT) was measured between the beginning of the release burst and the onset of periodic voicing. Mean values for each stop in intervocalic position are shown in Table 10. These values are based on ten speakers' productions of two tokens each of 10 words. Missing values for two speakers — one male and one female — reduced our subject pool to ten speakers for this investigation. The words used for VOT measurements are listed in Table 9.

Table 9. Words illustrating stop consonants in intervocalic position.

	a	**e**
p	**papa**	**ʔan pe**
	stingray	to put down
t	**pita**	**kote ka**
	fish (sp.)	his father
k	**koka**	**ʃikeʔ**
	fish (sp.)	shell corn
kʷ	**tokʷa**	**tokʷe**
	corn drink	Brazil nut

The measurements were submitted to a repeated-measures analysis of variance test with VOT as the dependent variable and consonant (**p**, **t**, **k** or **kʷ**), token (first or second), quality of following vowel (**a** or **e**), and speaker gender as the independent variables. The test revealed a highly significant difference in VOT by consonant (F=37.98; p<.001). Other effects investigated (gender, quality of following vowel [**a** or **e**], token number, quality of following vowel by token, and consonant by quality of following vowel by token) were not significant. Accordingly, the values shown in Table 10 are the mean VOTs pooled across all variables except place of articulation. These VOT values vary much like those in other languages (see Cho and Ladefoged, [forthcoming] for a survey of the effects of place of articulation on VOT in some similar languages).

Table 10. Mean VOT (ms) of Wari' stops for 5 male and 5 female speakers.

	p	**t**	**k**	**kʷ**
VOT	19	26	50	58

Only the nasals and non-labialized stops can occur in final position, as shown by the examples in Table 11. Final stops are usually unreleased.

Table 11. Words illustrating the complete set of final consonants.

p	**t**	**k**	**ʔ**	**m**	**n**
kap	**ʔat**	**wak**	**paʔ**	**ʍam**	**nan**
grub worm	bone	to cut	to kill	fish	wound

The status of glottal stop is somewhat different from that of the other consonants of Wari', principally in its anomalous distribution. Like the other stops, it occurs word-initially and word-finally, as shown in Tables 1 and 11. It is clearly phonemic, as shown by the phrases in Table 12. Spectrograms of the first pair of phrases are shown in Figure 14. The glottalization is quite subtle, but speakers and listeners can regularly distinguish these phrases.

Table 12. Phrases establishing the phonemic status of glottal stop.

weʔ na she is vomiting	**we nam** she calls her "my older sister"
miʔ na memem he gives fruit	**mi na memem** the fruit tree is producing

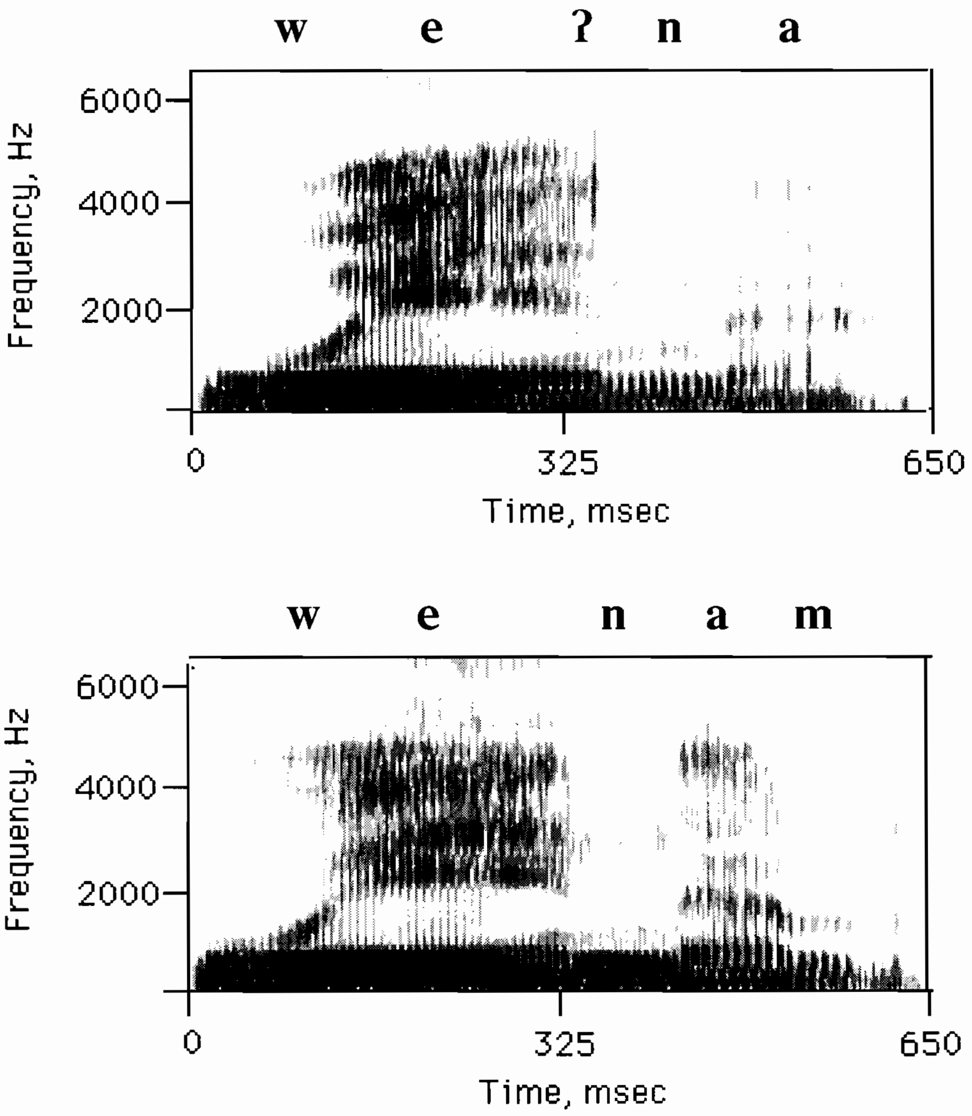


Figure 14. Spectrograms of the phrases **weʔ na** 'she is vomiting' and **we nam** 'she calls her "my older sister"' as produced by one male speaker.

The anomalous facet of the distribution of the glottal stop is that it can occur after a nasal and before a word boundary. In so doing it violates the normal Wari' syllable structure, which is CV(V)(C). On many (perhaps all) of the occasions when **ʔ** occurs after a consonant at the end of a word, it forms a separate morpheme. Everett and Kern (in press), noting the occurrence of **ʔ** after **m** and **n**, decided that **m**^{ʔ} and **n**^{ʔ} were single phonemes, and thus that words ending in these sounds did not violate Wari' syllable structure. The disadvantage of this analysis is that it posits two phonemes that differ from all other consonants in being restricted to appearance in word-final position. The contrasts involved are illustrated in Table 12. Spectrograms illustrating these contrasts are shown in Figure 15.

Table 12. The contrast between plain and glottalized nasals.

m	hwam	fish	**mʔ**	mamʔ	to find
n	nan	wound	**nʔ**	tanʔ	to arrive (pl.)

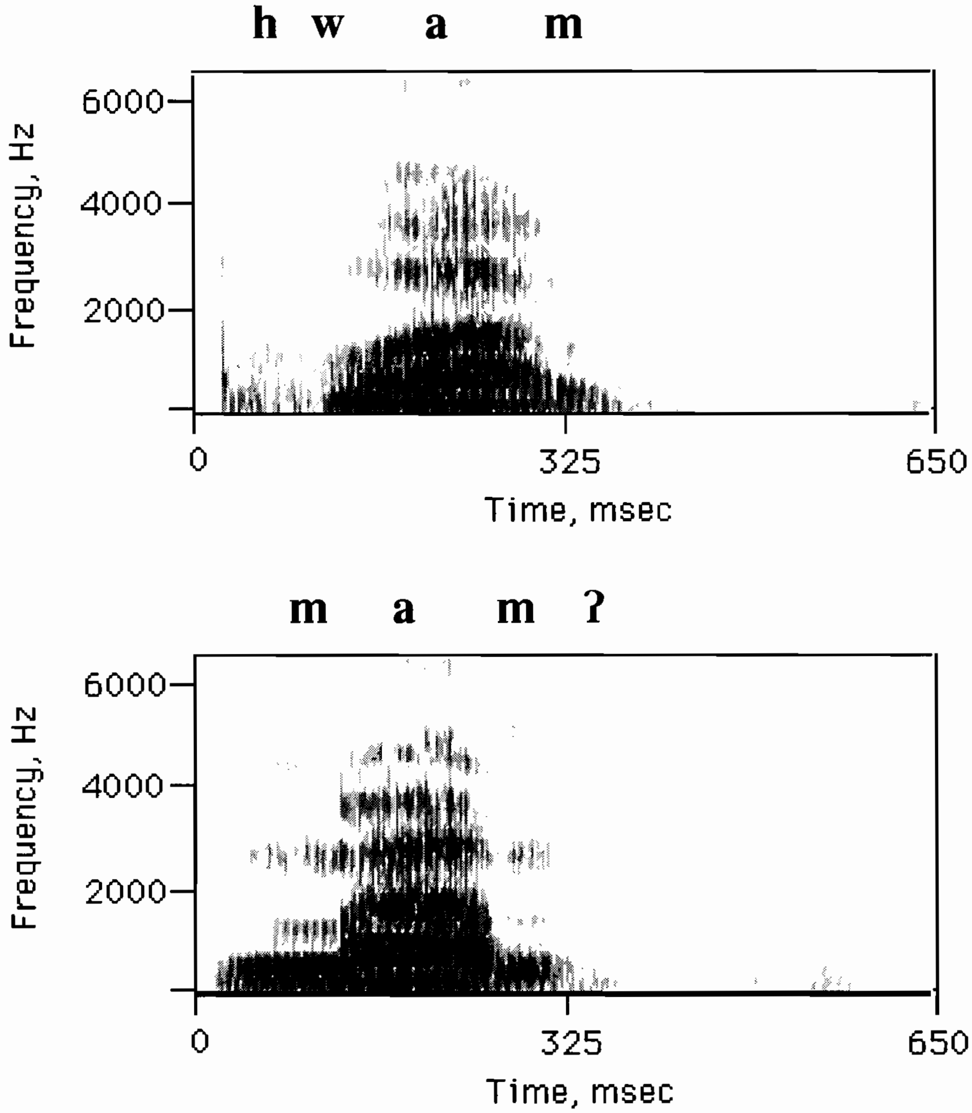
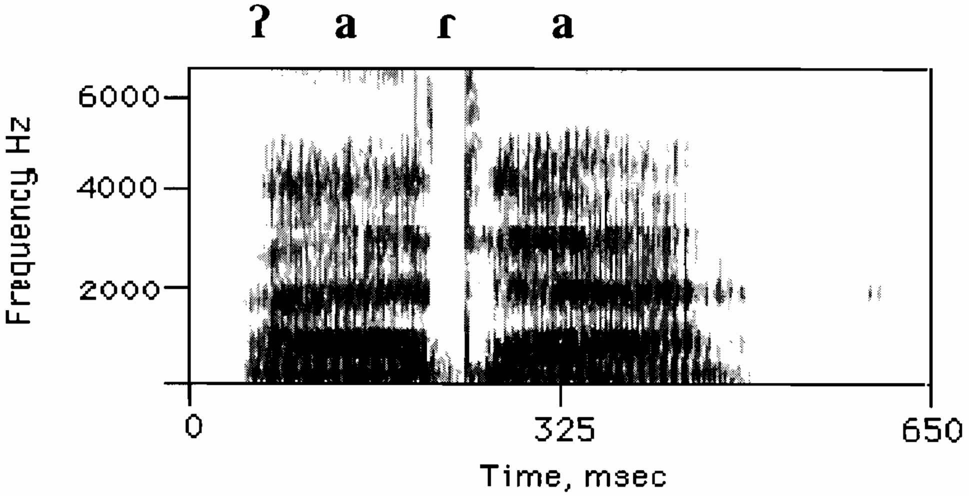
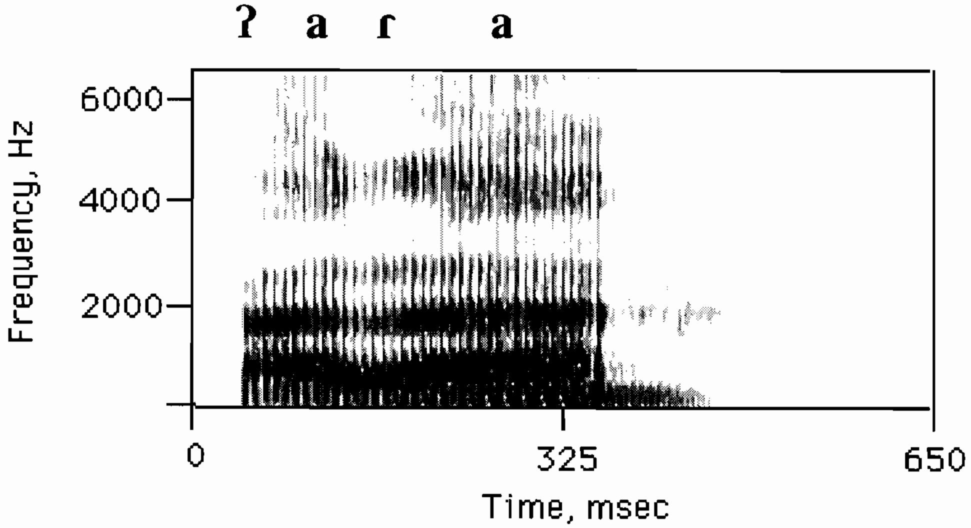


Figure 15. Spectrograms of the words **hwam** ‘fish’ and **mamʔ** ‘to find’ as produced by one male speaker.

We also looked at the acoustic characteristics of intervocalic glottal stop in the words **kaʔa** ‘bird (sp.)’ and **maʔe** ‘OK (fem.)’. In many cases, voicing was continuous, although of low amplitude and often heavily glottalized as well. In other cases, we observed rather long silent intervals (in one case as high as 130 ms). The typical intervocalic glottal stop, however, fell in between these two extremes, and consisted of short stretches of irregular voicing separated by brief periods of silence or very low-amplitude non-periodic noise.

The distribution of the tap **ɾ** is also anomalous, in that it can occur only intervocalically. We do not know the historical origin of this sound. It is similar to American English **ɾ** in words such as **pɪɾi**, ‘pity’, except that it has a dental articulation. This sound is subject to variation within and across speakers; we have noted three variants. It may be realized as a voiceless dental stop, a tap, or as a partially voiced dental stop. Each of these three variants is represented in the spectrograms in Figure 16.





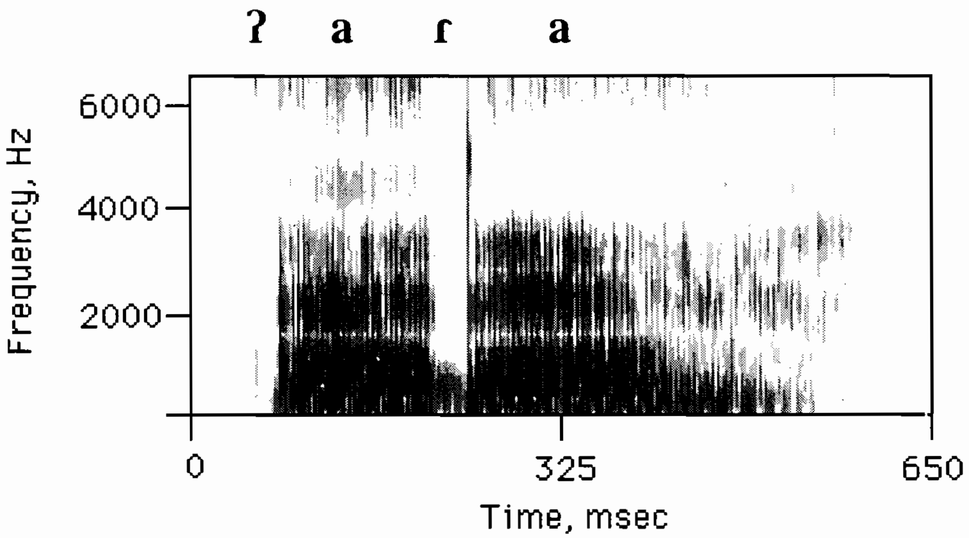
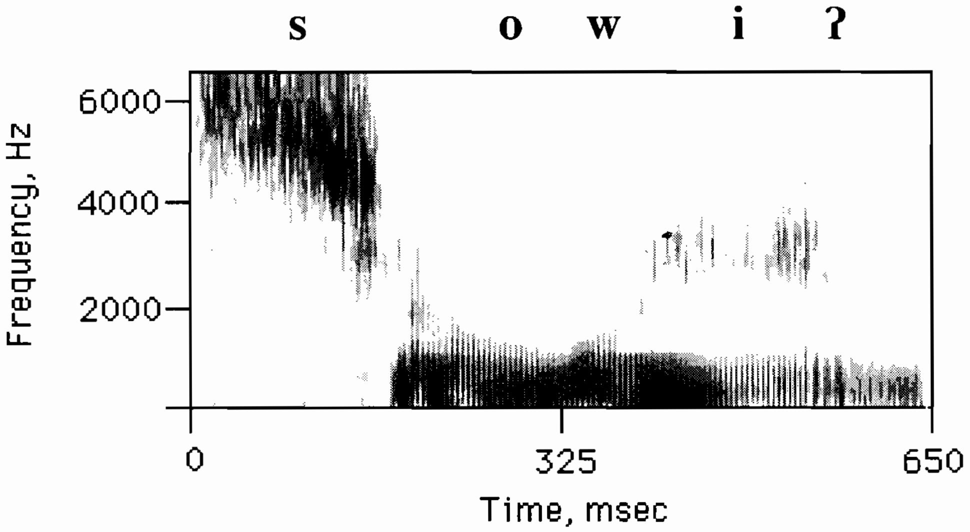


Figure 16. Spectrograms of the word **ʔaɾa** ‘to do’ as produced by three speakers. From top to bottom: male, male, female.

A post-alveolar fricative **ʃ** (represented by *x* in the orthography used by Barbara Kern) was listed in the consonant chart, but neither the symbol **ʃ** nor the post-alveolar fricative designation are entirely appropriate. This sound varies considerably from speaker to speaker, and from one phonetic context to another. We have noted **s**, **ts**, **ʃ**, **tʃ**, and (rarely) productions that we have characterized as **sts.** None of the **sts** variants of *x* that we observed exhibited a release burst following the silent interval. The only strong distributional tendency we observed for this sound is that it is overwhelmingly (but not exclusively) the fricative exponent which occurs in intervocalic position. Spectrograms of various realizations of this sound appear in Figures 17-20.



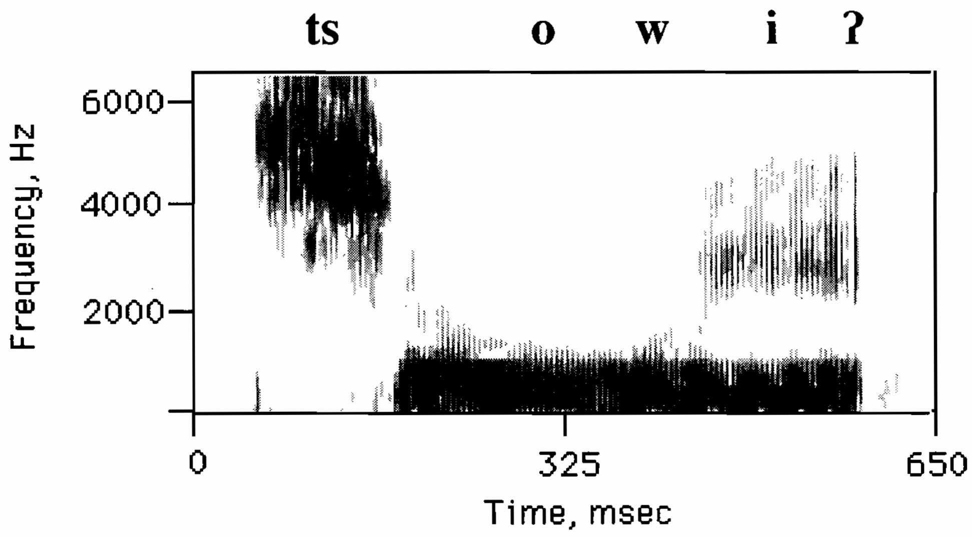


Figure 17. Two tokens of the orthographic word *xowi'* 'rain', produced by a single female speaker, illustrating **s** and **ts** variants of *x*.

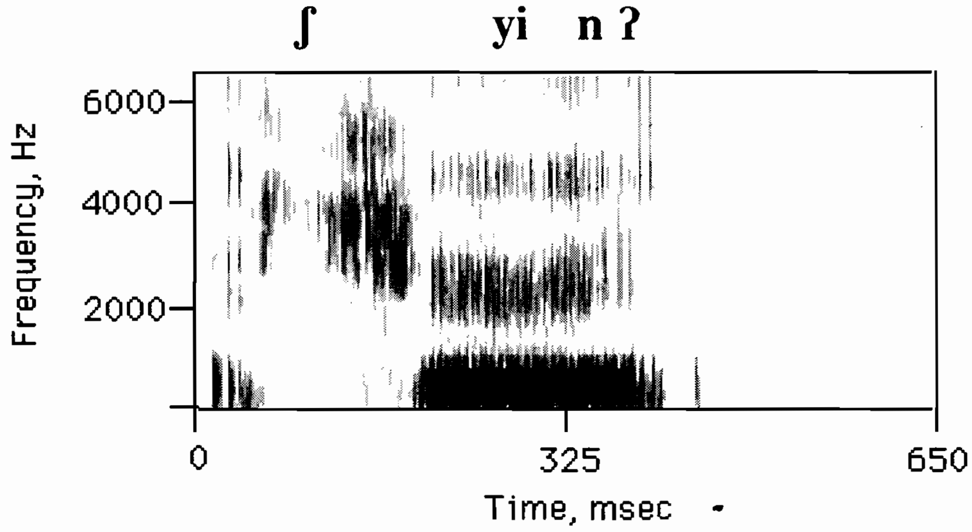


Figure 18. One token of the orthographic word *xuin'* 'snail', produced by a male speaker, illustrating **ʃ** variant of *x*.

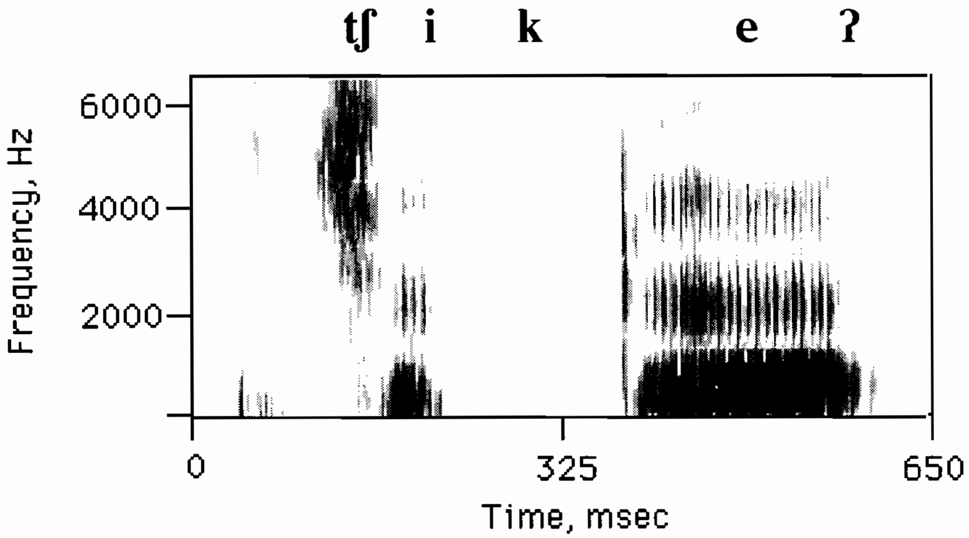


Figure 19. One token of the orthographic word *xique'* 'shell corn', produced by a male speaker, illustrating **tʃ** variant of *x*.

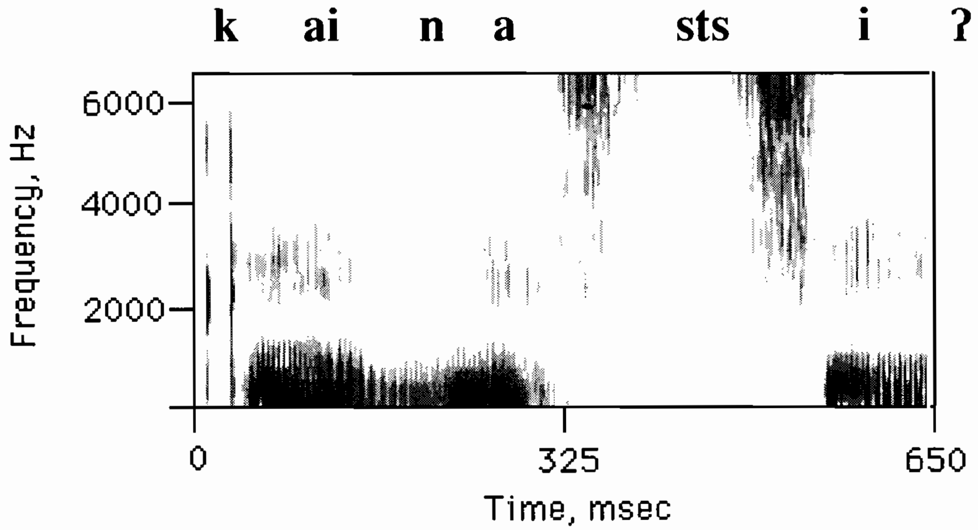


Figure 20. One token of the orthographic word *cainaxi'* 'our daughter (spoken by a woman)', produced by a female speaker, illustrating **sts** variant of *x*.

Finally in this account of the consonants we will consider the sound **t͡ʙ̥**, which occurs in a limited number of words, some of which are shown in Table 13. This sound, which is discussed in more detail by Ladefoged and Everett (in press), is a voiceless dental plosive which is released in such a way as to form a labial trill. It contrasts with **t** (and all other consonants), and has to be considered as a single sound. The articulation is not a combination of any two of the other consonants in Wari', and Wari' has no consonant clusters.

Table 13. Words contrasting **t͡ʙ̥** and **t**.

t͡ʙ̥ot͡ʙ̥o	to be pleasant	**toto**	to paint
t͡ʙ̥ot͡ʙ̥oweˀ	chicken	**towe**	to be fat
t͡ʙ̥owem t͡ʙ̥owem	dragonfly		
t͡ʙ̥ym	to be green	**tom**	to burn

The articulation of **t͡ʙ̥** may be seen from the palatogram and linguagram of **t͡ʙ̥ym** 'to be green' in Figure 21. The palatogram on the left shows that there was full lingual contact on the upper teeth and the front part of the alveolar ridge. The linguagram on the right shows the part of the blade of the tongue that makes this contact. There is only light contact in the midline near the tip of the tongue, which can be seen more clearly on the original photograph in the area between the dashed white lines. The part of the tongue in the midline further back on the blade did not contact the roof of the mouth. The shaping of the tongue blade and tip forms a narrow jet of air when the closure is released.

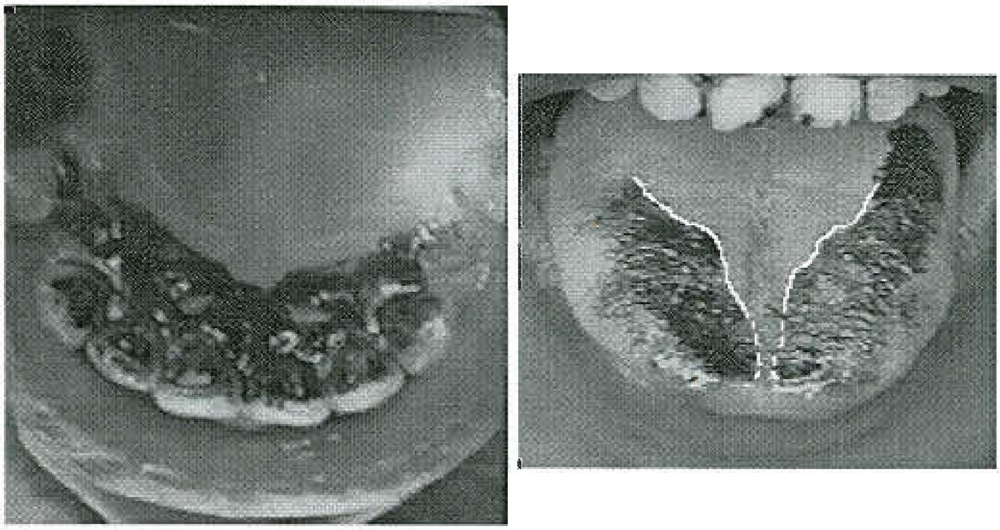


Figure 21. Palatogram and linguagram of Wari' [t͡ʙ̥ym] 'to be green'.

Spectrographic analyses of this sound are shown in Figure 22. In example (a) on the left of the figure, the lips were slightly apart when the **t** closure was released at what has been marked as time 0. They were sucked together at the time marked by the arrow at the top of the picture, and then blown apart, forming a bilabial fricative. By the time of the second arrow at the top of the picture they had been drawn together again so that a complete closure was formed. Almost instantly they were blown apart and sucked together again at the time of the third arrow, and released a little later into a very fricative vowel. There were thus three taps in this trill. Example (b) on the right of the figure illustrates the more common form of this sound as produced by our speakers (all of whom were thoroughly familiar with this sound). The lips vibrate but not with sufficient magnitude to cause a series of closures followed by releases; instead the lips are moving only slightly while in a bilabial fricative position.

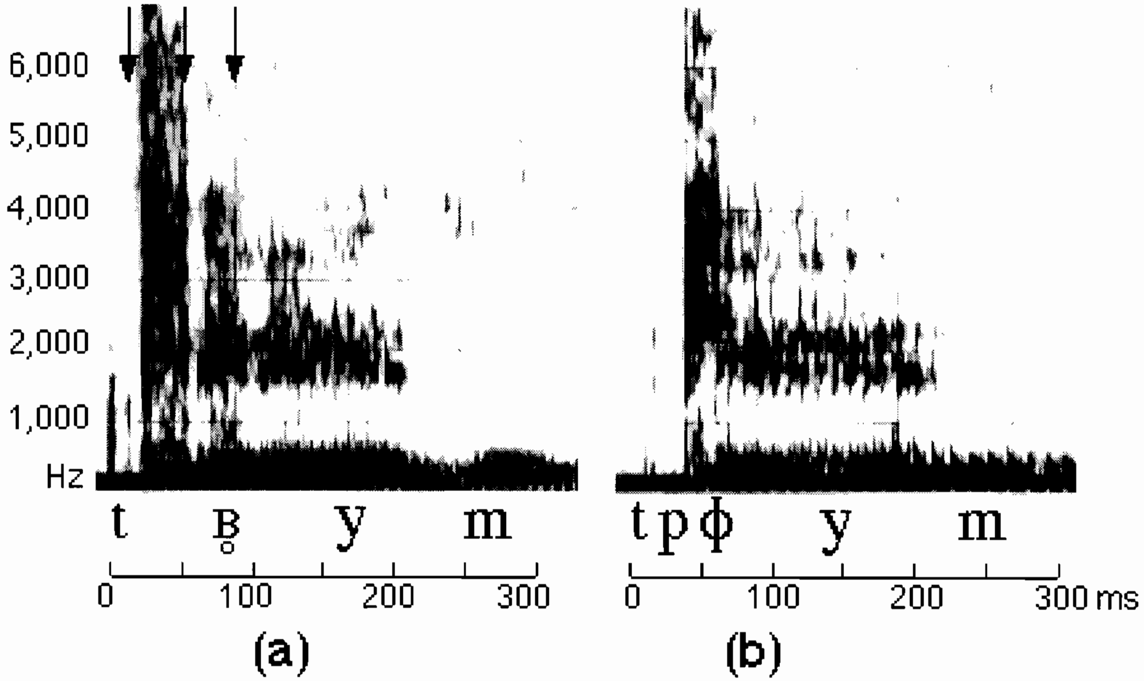


Figure 22. Spectrograms of the Wari' word [t͡ʙ̥ym], 'to be green' (a) pronounced with an observable trill; (b) the more usual form in which the tremors of the lips form a bilabial fricative rather than a trill.

Stress

The major stress in a Wari' phrase is on the last syllable of the verb or verbal compound; secondary stresses appear on other syllables. The following examples show the location of main stress, indicated by underlining, moving steadily away from the beginning of the sentence as the verbal compound is lengthened in various constructions. In the final example, **kep kep ʔaɾa ʃiɾapaʔ** is an embedded sentence; **na** 'consent' is the verb, so it is **na** that bears main stress.

kep ʃiɾapaʔ kam k^{w}aʔ
'touch' – 3 sg. irr.,1 sg. – f. – 'this ' m. or f.
She should touch me.

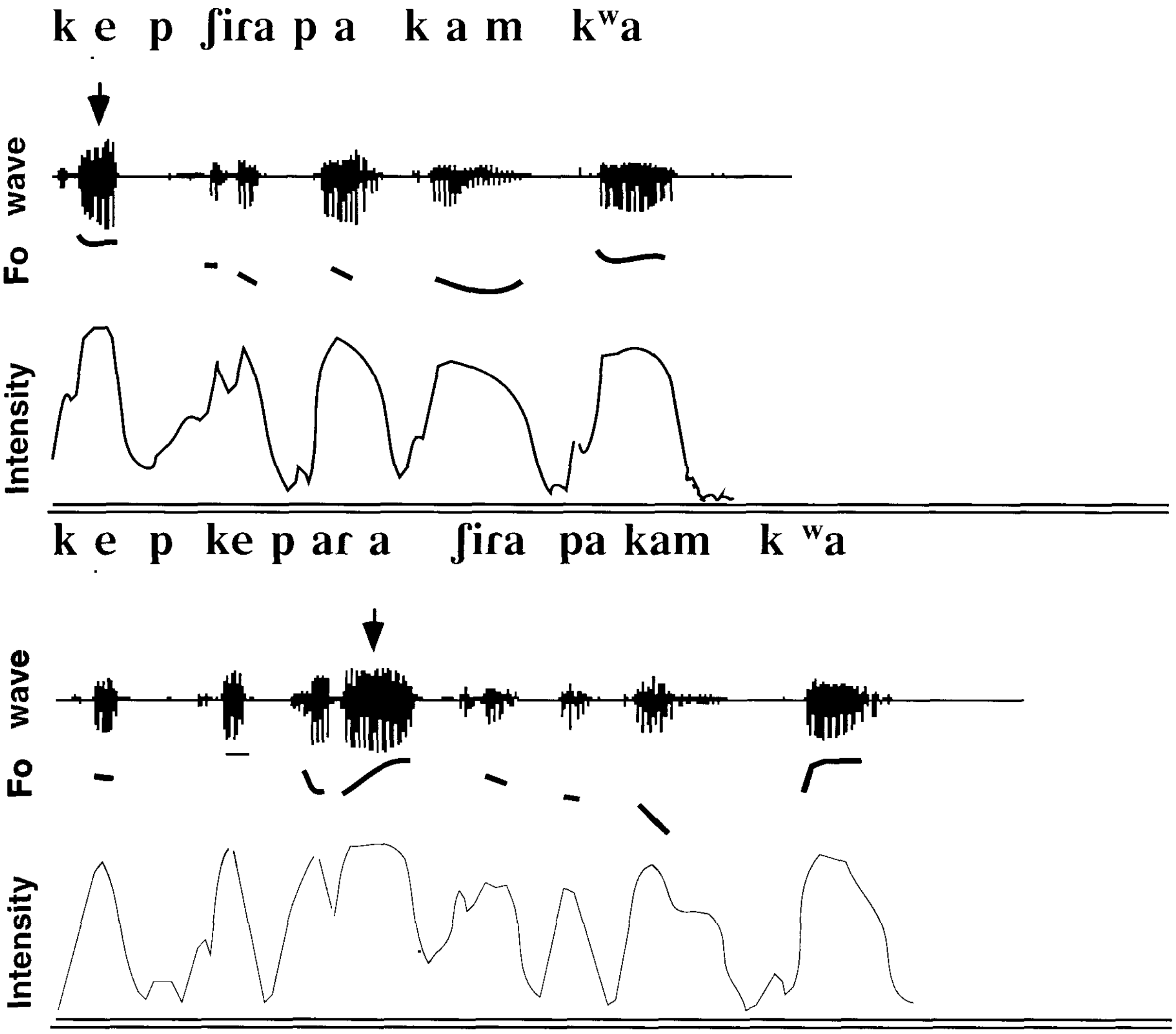
kep kep ʃiɾapaʔ kam k^{w}aʔ
'touch'-'touch' – 3 sg. irr.,1 sg. – f. – 'this ' m. or f.
She should continually touch me.

kep kep ʔaɾa ʃiɾapaʔ kam k^{w}aʔ
'touch'-'touch' – 'not' - 3 sg. irr.,1 sg. – f. – 'this ' m. or f.
She should not continually touch me.

kep kep ʔaɾa ʃiɾapaʔ ʔinam kam k^{w}aʔ
'touch'-'touch' – 'not' - 3 sg. irr.,1 sg. – verbal infl. - f. – 'this ' m. or f.
"She should not continually touch me" I (say) of her.

kep kep ʔaɾa ʃiɾapaʔ na ʔinam kam k^{w}aʔ me
'touch'-'touch' – 'not' - 3 sg. irr.,1 sg. – 'consent' - verbal infl. - f. – 'this ' m. or f. - emph.
"She should not continually touch me" consenting, I (say) of her.

The first three speakers did not record this set of examples, and some of the remaining 9 speakers did not produce the complete set of sentences in a fluent way, with just the stresses as shown here. It was, however, clear that the general pattern was as indicated. Figure 23 shows the analysis of one speaker producing the first, third and fourth of these sentences (the fourth is reproduced in truncated form). It may be seen that the largest peak in the intensity record and a peak in the pitch record occur on the stressed syllables; these points are marked with arrows in the figure.



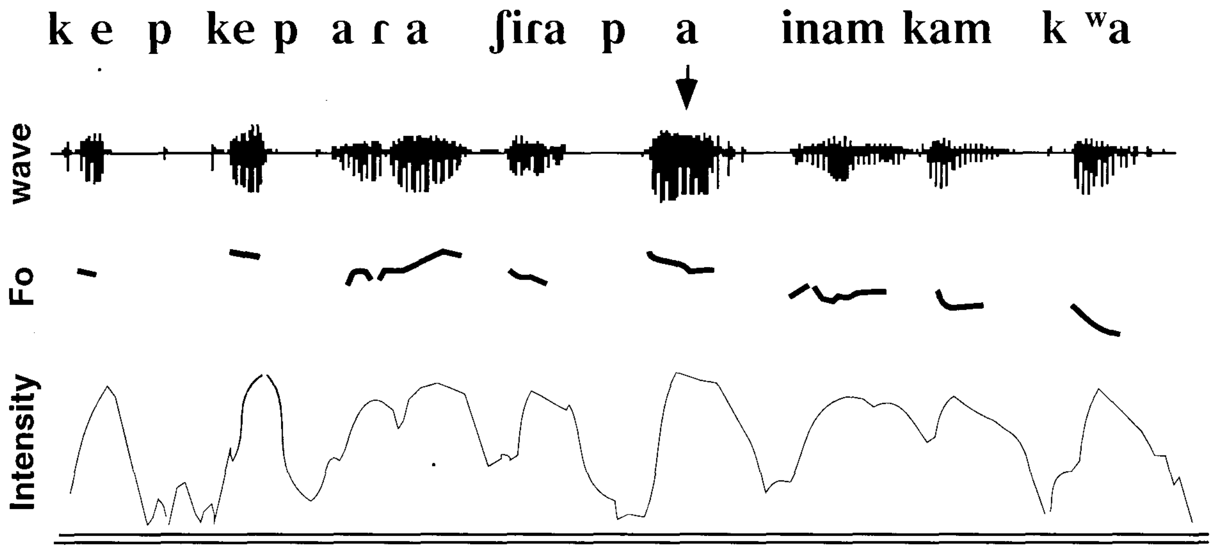


Figure 23. The waveforms, fundamental frequency and intensity in three sentences of Wari' (see text), indicating the major sentence stress.

References

Cho, Taehong and Peter Ladefoged, forthcoming. The effect of place of articulation on VOT: a cross-linguistic study.

Crothers, John, 1978. Typology and universals of vowel systems. Universals of human language, vol. 2, ed. by Joseph H. Greenberg, 93-152. Stanford, California: Stanford University Press.

Derbyshire, Desmond C., and Geoffrey K. Pullum, eds., 1986. Handbook of Amazonian languages, vol. 1. New York: Mouton de Gruyter.

Everett, Daniel L., and Barbara Kern, in press. Wari' (The Pacaas Novos language of Western Brazil). London: Routledge Publishers, Descriptive Grammar Series.

Grimes, Barbara F., ed., 1988. Ethnologue languages of the world. Dallas, Texas: Summer Institute of Linguistics, Inc.

Ladefoged, Peter, 1996. Elements of acoustic phonetics, 2nd edn. Chicago: University of Chicago Press.

Ladefoged, Peter, and Daniel Everett, in press. The status of phonetic rarities. Language 72.

Lindblom, Bjorn, 1986. Phonetic universals in vowel systems. Experimental phonology, ed. by John J. Ohala and Jeri J. Jaeger, 13-44. Orlando, Florida: Academic Press, Inc.

Lindblom, Bjorn, 1990. Phonetic content in phonology. PERILUS 1990:101-118.

Maddieson, Ian, 1984. Patterns of sounds. New York: Cambridge University Press.

Ruhlen, Merritt, 1975. A guide to the languages of the world. Language Universals Project, Stanford University.

Voegelin, C.F., and F.M. Voegelin, 1978 (1977). Classification and index of the world's languages. New York: Elsevier.

Acknowledgements

This paper is for the Wari' people, to whom we owe many thanks. It was supported by an NSF grant: "Recording the phonetic structures of endangered languages." We would also like to acknowledge the help of Dr. Sumiko Takayanagi in analysing this data.

The main features of Tsez phonetics

Ian Maddieson, Ramazan Rajabov*, and Aaron Sonnenschein*

I. Introduction

Tsez, formerly also known as Dido, is a language spoken by a population of around 8000 people in the highlands of the western part of Dagestan near the upper reaches of the Sulaq river and adjacent to the border with Georgia. Dagestan lies on the western shore of the Caspian Sea and is a constituent republic of the Russian federation. A sketch map of Dagestan is shown in Figure 1 with the area in which Tsez is spoken shown by cross-hatching. Kidiro is the administrative capital of the local regional government. Tsez has been spoken in this area since the beginning of the recorded history of the area. An additional 6-7000 Tsez speakers live in other parts of Russia or elsewhere in the world.

Figure 1. Map of Dagestan, showing the area where Tsez is spoken.

---

* Department of Linguistics, University of Southern California.

The dialect of Tsez spoken by the speaker studied for this project (the second author) is the Tsebari subdialect of the Asax dialect. This subdialect is spoken by 400 speakers, all of whom are fluent in the language. In addition, 60 per cent of the population is fluent in Russian and 70 per cent of the population is fluent in Avar, which serves a regional lingua franca. The older generation is also fluent in Georgian. There are no written texts in the Tsez language before the 20th century.

The Tsez language belongs to the Tsezic (or Didoic) language group. The other languages in this group are Hunzib, Xwarshi, Hinuq, and Bezht'a. Tsezic forms part of the Dagestanian branch of the Nakh-Dagestanian language family. It has usually been considered to constitute one branch of an Avar-Andi-Tsez grouping within Dagestanian, with Avar-Andi forming the other branch. However, this grouping is questioned by Nicolayev and Starostin (1994), who propose rather that Tsezic is an independent branch of Dagestanian. The Nakh-Dagestanian family is sometimes also called the North-East Caucasian family. This term should not be taken to indicate that there is an established larger family 'Caucasian' of which this is one branch (see Comrie 1981, Catford 1991), although probably North-East Caucasian and the North-West Caucasian languages, such as Abkhaz and Kabardian, are related. The third 'Caucasian' group, South Caucasian or Kartvelian (Georgian and its close relatives), is not obviously connected to the other two.

Like most of the languages of the Caucasian region, Tsez has an interesting consonantal system. There are stops at a wide range of places and every pulmonic stop or affricate has an ejective counterpart as well. Uvular consonants display a contrast between plain and pharyngealized categories. Pharyngealization also can be found with other consonants but only under special circumstances.

In contrast to the consonantal system, the vocalic system is a relatively simple one of five vowels. The main point of interest related to the vowels concerns the occurrence and distribution of the vowels which are phonetically pharyngealized.

There has been a long-standing debate in Caucasian phonology concerning whether or not the source of pharyngealization is vocalic, consonantal, or suprasegmental. The principal goal of this paper is therefore to present acoustic phonetic evidence which will assist in determining the nature of pharyngealization in Tsez. Before proceeding farther, we will first give a brief description of the segmental phonological system of Tsez.

2. Consonant inventory

There are 41 consonants in Tsez, as shown in the chart below. All the consonants in the chart have the pronunciation implied by the choice of IPA symbol used to represent them, except as noted below. Words illustrating many of these consonant contrasts are given in Table 1.

The Tsez dentals are apical. The uvular stops have some frication in their releases, which is sufficient to have caused some linguists (e.g. Bokarev 1959, 1967) to consider them to be affricates. The nature of /ʡ/, viewed here as basically a pharyngeal stop, will be discussed in more detail below. /ʜ/ is a voiceless lower pharyngeal fricative.

	Bilabial	Dental	Palato-alveolar	Palatal	Velar	Uvular	Pharyngealized uvular	Pharyngeal	Glottal
Plosive	p b	t d			k g	q	qˤ	ʡ	(ʔ)
Ejective stop	p’	t’			k’	q’	qˤ’		
Affricate		ts	tʃ						
Ejective affricate		ts’	tʃ’						
Fricative		s z	ʃ ʒ			χ ʁ	ʁˤ	ʜ	h
Nasal	m	n							
Trill		r							
Lateral affricate		tɬ							
Lateral ejective affricate		tɬ’							
Lateral fricative		ɬ							
Lateral approximant		l							
Approximant	w			j					

χˤ also occurs in Mokok dialect.

All consonants occur in both coda and onset positions, except the pharyngeal stop which is restricted to word-initial position in native vocabulary. In loan words from both Avar and Arabic sources, one finds pharyngeal stops in medial and final position as well. Examples are as follows: ʡiyada ‘to cry’ (native), maʡdin, ‘beautiful one’ (< Arabic), baʡarizi ‘to heat’ (< Avar), waraʡ (< Arabic) ‘honor’. The phonetic realization of this segment in intervocalic position is as an approximant, as is illustrated in Figure 9 below. The voiceless pharyngeal fricative can occur in any position. A glottal stop occurs in utterance-initial position if no other consonant precedes the first vowel. We do not at this time know the rest of the distributional pattern of the glottal stop or whether it should really be considered a part of the consonant inventory, or treated as a feature of the prosody of the language. There are also a few loanwords from Arabic in which a glottal stop occurs word-medially in coda position, for example /muʔminzabi/ ‘believers’.

In a restricted number of words apparent geminate consonants occur. However, it seems best to consider these as simply the result of two identical consonants occurring next to each other at morpheme boundaries. At present, we also take no position on whether consonants followed by [w] should be considered to form a set of labialized consonants.

Table 1. Some Tsez minimal sets illustrating consonant contrasts.

baru	‘wife’	bilu	‘penis’
daru	‘medicine’	gilu	‘middle of the river’
bero	‘ice’	gero	‘bell’
pero	‘warmth’	kero	‘canyon’
p’ero	‘chatter’ (N)	keron	‘canyon’
t’ili	‘pole’	k’eron	‘bagel (bread with hole)’
dili	‘lullaby’	maq’an	‘melody’
t’iri	‘attention, care’	maqan	‘barley’
tiri	‘log, beam’	ʁira	‘sexual desire’
tsetsa	‘seep’ (N)	xira	‘suck’
ts’ets’a	‘chop’	quqa	‘become dry’
sesru	‘lean meat’	q’uq’a	‘rub’
zezru	‘the one sensitive to cold’	neq’aj	‘sour’
ʒe	‘he’	meq’aw	‘badly behaving’
ʃe	‘letter’	ʡo	‘axe’
betʃnoda	‘vomit’	ʜo	‘oath’
betʃ’noda	‘cut’	retɬ’u	‘yoke’
biɬa	‘become similar’	ʁetɬu	‘pants’
bitɬ’a	‘go, walk’	hit’o	‘plunge’
bitɬa	‘plough’	ʜit’u	‘frying pan’
eni	‘mother’ (vocative)		
eli	‘we’		

3. Vowel inventory

There are 5 distinctive vowel qualities in Tsez, which may be represented with the symbols /i, e, a, o, u/. There are no restrictions on the position of a given vowel except for /e/, which can only rarely be found in word-final position. Examples of contrasts between these five vowels are given in Table 2.

Table 2. Some minimal sets illustrating Tsez vowels.

ik’a	‘to go’	bitsa	‘to bind’ (class iii)
ek’a	‘to spin’	betsa	‘to melt’
ak’a	‘official punishment’	batsa	‘to become bored’
ok’a	‘to beat’	botsa	‘to become clean’
uk’a	‘to bow’	butsa	‘to dry’

There is also a long vowel [aː] which occurs only in certain morphological forms, such as the future forms and the past participles of verbs, and the ergative case of nouns — for instance, /ik’a/ ‘to go’ but [aːki] ‘will go’; /uʒi/ ‘boy’ but [uʒaː] ergative case.

4. Ejectives

At every oral place of articulation where a voiceless pulmonic stop (plosive) occurs there is a corresponding ejective stop. There are also dental, palato-alveolar and lateral ejective affricates contrasting with pulmonic counterparts. Ejectives involve the co-production of a glottal closure with the oral articulation, and — in the canonical case — raising of the larynx so that the air between the glottal and oral closures is compressed, giving the release of the oral closure, which occurs first, a distinctive quality. The relative timing of the oral and laryngeal movements involved can be characteristically different in different languages (Ladefoged and Maddieson 1996).

An illustrative pair of spectrograms comparing a voiceless velar plosive and a velar ejective in intervocalic position is shown in Figure 2. Figure 3 illustrates a voiceless pharyngealized uvular plosive and the corresponding pharyngealized uvular ejective stop. These pairs demonstrate several interesting properties of the contrast as it is produced in Tsez. The onset to the closure for the pulmonic stops usually has a noisy voiceless transition, indicating that the vocal folds open before the oral closure is formed, producing a short amount of preaspiration. The stop closure is quite short and is followed by a relatively lengthy period of voiceless noise prior to the beginning of voicing for the following vowel. In the uvular case we do not consider the amplitude of this noise to be sufficient to justify calling this an affricate as suggested by Bokarev and others, but just regard all the voiceless plosives as having a certain amount of post-aspiration. Thus the spectrograms indicate that there is a relatively long opening gesture of the vocal folds during which a brief oral closure occurs. This pattern is most salient with uvulars in the acoustic data we have examined but it is typical in its general outline of other plosives as well.

In comparison with the plosives, the ejective stops in Figures 2 and 3 show a more abrupt cut-off of voicing at the end of the first vowel, particularly so in Figure 3. This is almost certainly due to the vocal folds closing firmly rather than to the completion of an oral closure. The oral closure that follows is released with stronger energy in the burst than is observed in the corresponding plosive. The interval between this release and the following vowel is not filled with continuous noise, as it is in the case of the plosive. Instead there is a period of silence or near-silence. This 'silent interval' indicates that the glottal closure is still maintained at this time, rather than being released at about the same time as the oral closure. The vowel onset often has several periods of glottalized phonation before modal phonation begins, as the glottal closure is relaxed. Again, the acoustic pattern indicates a relatively long laryngeal gesture overlapping a rather brief oral closure. Note, however, that the entire oral articulatory gesture for both the velar and uvular consonants in these words is long. In Figure 2, the second formant begins to rise and the third formant to fall almost from the beginning of the initial /a/ as the articulators move into position for the velar closure. A slow movement away from the velar position is apparent in the long rising transition of the third formant in the second vowel. In Figure 4, the first formant starts rising and the second, third and fourth start lowering at the beginning of the initial /i/ as the articulators move toward the position for the pharyngealized uvular. The slow movement away from this configuration is apparent in the long rising transition of the third formant in the final /a/.

Overall the total duration of the consonantal part of these tokens is similar, and the interval from stop release to the onset of the modal vowel (the Voice Onset Time or VOT) is very similar between the pulmonic and ejective stops. The difference between them lies not so much in durational properties as in the way that the oral closure is flanked by noise in one case, but by (near-)silence in the other.

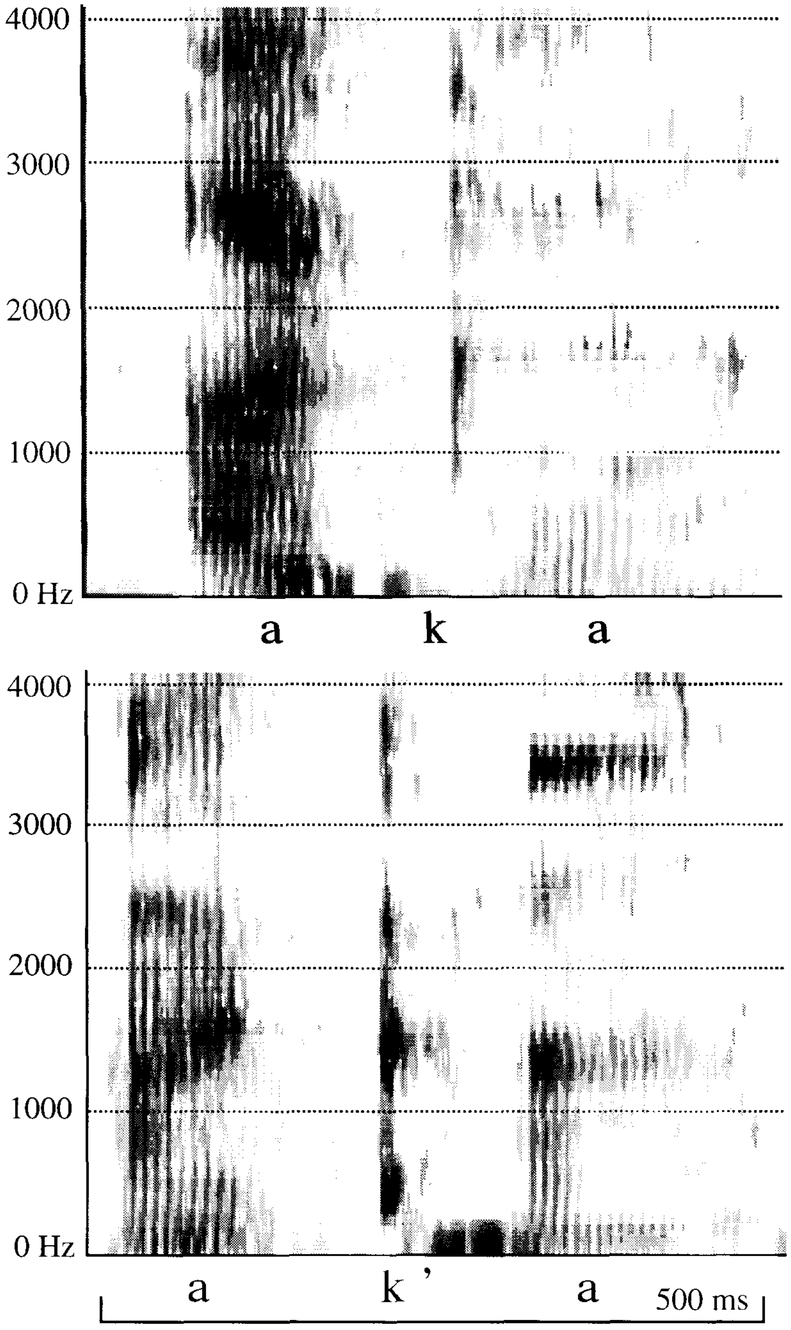


Figure 2. Spectrograms of the words /aka/ 'bracelet' and /ak'a/ 'to sharpen' illustrating pulmonic and ejective velar stops.

In order to provide some quantitative information on the difference between pulmonic and ejective stops and affricates three repetitions of a series of minimal or near-minimal pairs was recorded directly onto the Kay CSL speech analysis system using a head-mounted close-talking microphone. The words chosen are listed in Table 3. The speech was digitized at 20 kHz and simultaneous displays of waveforms and spectrograms were examined. Measurements were made of the time at which the initial vowel ended, the time of the oral release and the time at which the second vowel began. Onset and offset of the vowel were identified as the times at which modal voicing ended or began.

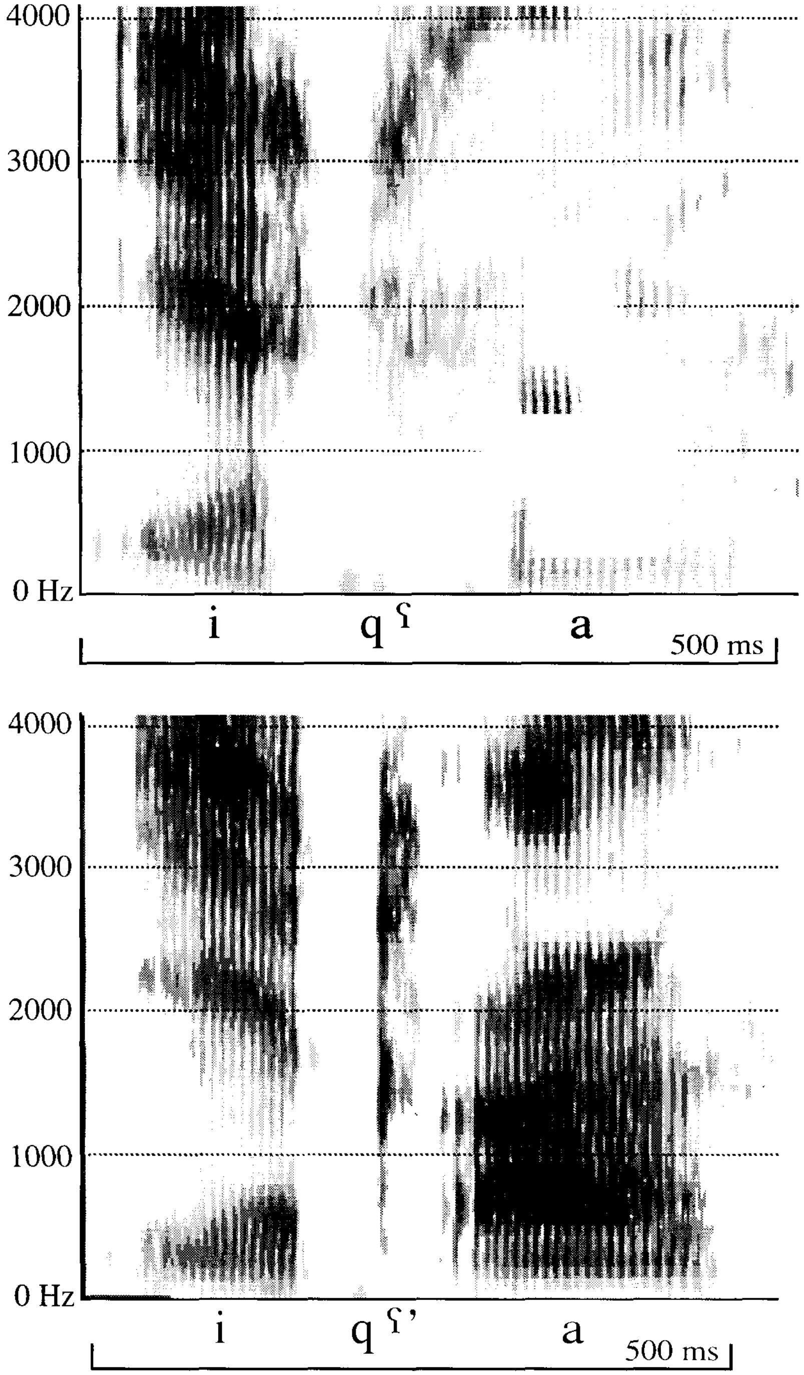


Figure 3. Spectrograms of the words /iqˤa/ 'to drain' and /iqˤ'a/ 'to be caught' illustrating pulmonic and ejective pharyngealized uvular stops.

Table 3. Minimal or near-minimal pairs contrasting pulmonic and ejective stops and affricates.

p	tɬopo	*low*	p'	tɬˤop'o	*dung*
t	uta	*to forage*	t'	ut'a	*to turn back*
ts	tsetsa	*to wring*	ts'	ts'ets'a	*to chip*
tʃ	bitʃa	*to stay*	tʃ'	bitʃ'a	*to dig*
k	aka	*bracelet*	k'	ak'a	*to sharpen*
q	iqˤa	*to drain*	q'	iqˤ'a	*to be caught*
tɬ	bitɬa	*to plow*	tɬ'	bitl'a	*to go*

From these time-points two durations were calculated, the duration of the interval from the offset of the vowel to the stop burst, called here rather loosely the 'closure duration', and the duration from the burst to the onset of the vowel, called here the VOT. Note that VOT as defined in this case includes the frication portion of affricates, and the 'closure duration' includes any preaspiration. The means for the three repetitions are shown in Table 4, rounded to the nearest millisecond.

Table 4. Closure duration and VOT of stops and affricates.

		Pulmonic			**Ejective**	
		Closure	VOT		Closure	VOT
bilabial stop	p	110	84	p'	113	78
dental stop	t	123	52	t'	113	83
velar stop	k	109	59	k'	109	83
pharyngealized uvular stop	qˤ	81	113	qˤ'	67	73
dental affricate	ts	90	95	ts'	71	87
palato-alveolar affricate	tʃ	97	100	tʃ'	92	99
lateral affricate	tɬ	105	99	tɬ'	100	83

Analysis of variance of the stop durations with both laryngeal type and place of articulation as main effects showed that laryngeal class (pulmonic vs. ejective) does not significantly affect either the closure duration (F (1, 16) = 1.58, p = .2268), or the VOT (F (3, 16) = 0.101, p = .7549). The overall mean closure duration for pulmonic stops is 106 ms, and for ejective stops is 101 ms. Mean VOT is 77 ms for pulmonic stops and 79 for ejectives. There are significant differences between different places of articulation, with pharyngealized uvular stops having significantly shorter closure duration than the other three places, and a significantly longer VOT than the bilabial or velar stops in this data set. This last effect is due to the very long mean VOT of the pulmonic pharyngealized uvular stops. Since we do not have a minimal contrast among plain uvulars, we do not know if these timing patterns are typical of uvulars in general or are only characteristic of the pharyngealized uvular stops.

The affricates show a marginally longer mean closure duration and VOT for the pulmonic class than for the ejectives, but significance is not better than the .01 level for either comparison.

5. Phonetic analysis of pharyngealization

Of particular interest to the phonetician is the analysis of pharyngealization in Tsez. There has been a great deal of debate in the literature of Caucasian linguistics concerning the nature and role of pharyngealization and there have been, in particular, conflicting claims concerning the nature of pharyngealization in Tsez. Kibrik and Kodzasov (1990: 315) suggest that pharyngealization is a prosodic feature (in a sense similar to that used by J.R. Firth and his disciples) which operates at the level of the entire word. They mention specifically the words *ʔitɬ'u* 'cover' and *bitɬoqˤu* 'insect' where, they claim, the effect of pharyngealization is spread considerably beyond the syllable in which a pharyngeal or pharyngealized consonant is written. Others believe that pharyngealization is a phonemic feature related to vowels (Bokarev 1959; Nikolaev and Starostin 1994), or to uvular consonants and vowels (Alekseev and Rajabov, in press).

In the interpretation of the Tsez phonological system adopted here we agree that pharyngealization is a segmental feature, specifically a consonantal one. In our view there are two pharyngeal consonants /ʜ ʡ/, and a series of pharyngealized uvular consonants. As noted above there is a very audible difference between vowels with and without pharyngealization. When a word appears to begin with a pharyngealized vowel we will argue that this is because there is actually an initial occurrence of /ʡ/. Pharyngealization of other consonants, including labials and dentals, is noted as a possibility in word onset position. These other pharyngealized consonants in word-initial position are sometimes, but not always, the result of prefixing a consonant to a root which begins with an apparent pharyngealized vowel. The similarity of the positional restrictions between pharyngealized vowels and these additional, non-uvular, pharyngealized consonants suggests that these consonants should be regarded as consisting of the coalescence of another consonant and /ʡ/. Other than in initial position pharyngealization of vowels is only heard after pharyngealized uvulars.

Pharyngealized vowels.

A selection of minimal or near-minimal pairs differing in whether the initial vowel was plain or pharyngealized was recorded on digital audio tape and digitally transferred onto the Kay CSL speech analysis system. The digitized files were downsampled to 10 kHz. Formants were measured at the onset and at the midpoint of the vowels using a 12th order LPC to determine formant location. Simultaneous displays of a wideband spectrogram and an FFT analysis taken at the same time-point were consulted to verify the formant analysis. (A few supplemental tokens were recorded directly onto the CSL.) The mean values of the first three formants are given in Table 5 for the vowel midpoint and in Table 6 for the vowel onset. Four pairs of sample spectrograms are given in Figures 4-7. (No pair illustrating the contrast of /e/ and /ʡe/ is shown.)

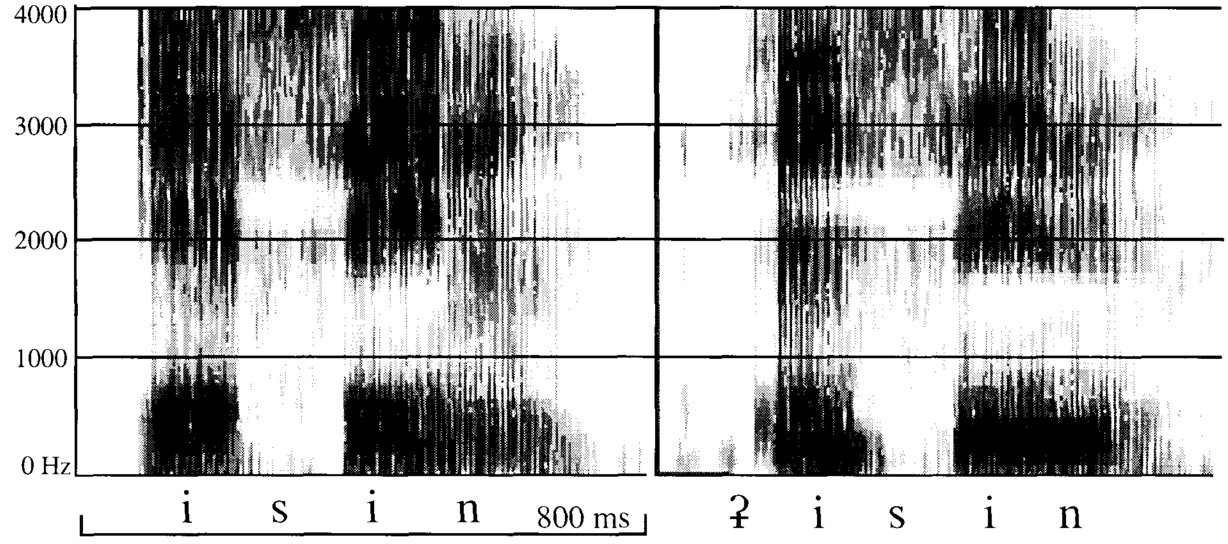


Figure 4. Spectrograms of the minimal pair /isin/ 'and snow' and /ʡisin/ 'small'.

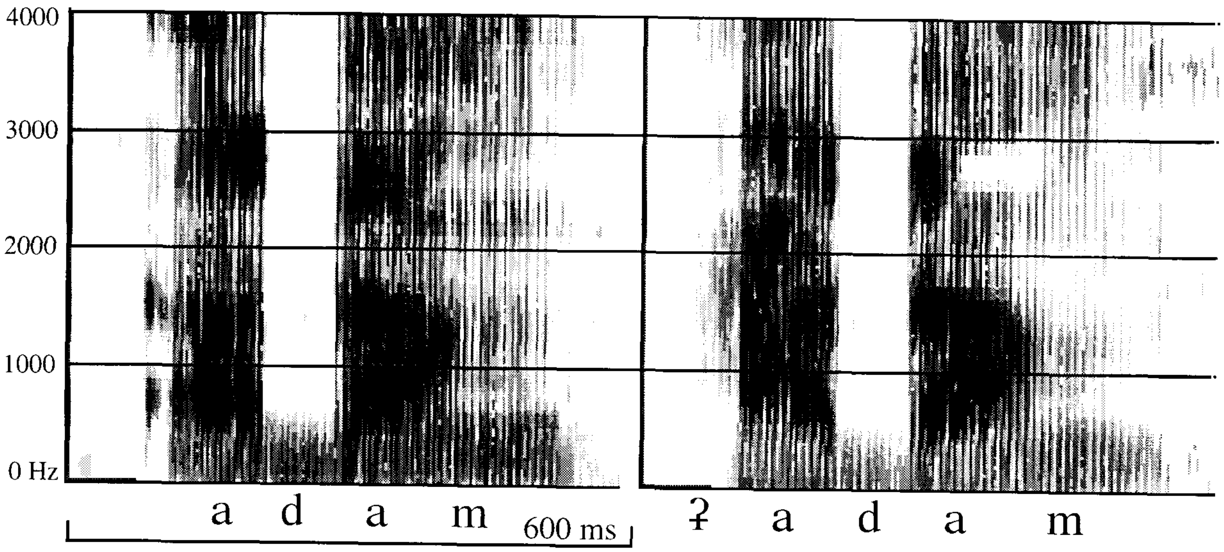


Figure 5. Spectrograms of the minimal pair /adam/ 'Adam (personal name)' and /ʔadam/ 'person' (< Arabic).

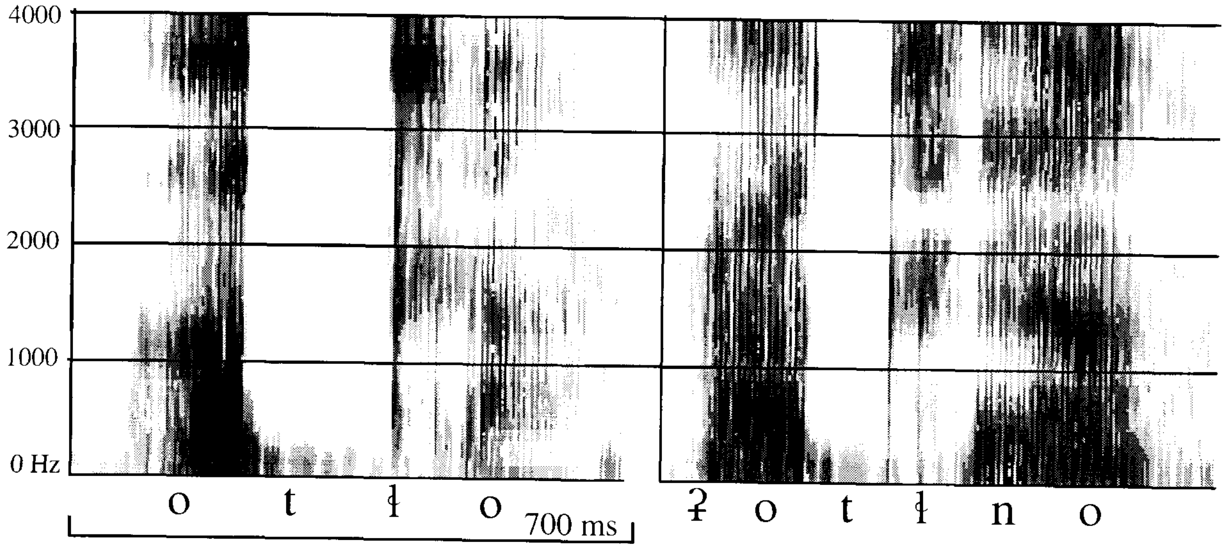


Figure 6. Spectrograms of the near-minimal pair /otɬo/ 'in the middle' and /ʔotɬno/ 'seven'.

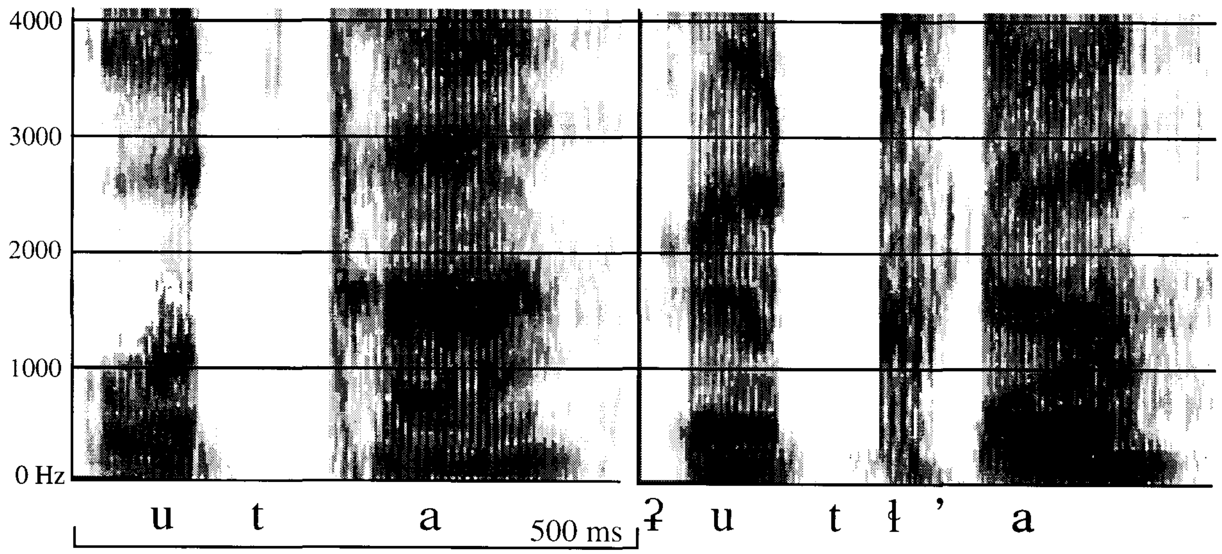


Figure 7. Spectrograms of the near-minimal pair /uta/ ‘gloss’ and /ʔutɬ’a/ ‘gloss’.

There are several points we wish to draw particular attention to in these spectrograms. In Figure 4 note the fairly low F1 and high F2 (above 2000 Hz) in /isin/, compared with higher F1 and lowered F2 (below 2000 Hz) in /ʔisin/. In Figure 5 there is a major difference between the F3 values of the first vowels in /adam/ and /ʔadam/ but the formants in the second vowels are similar. In Figure 6 there is a low F2 and a high F3 in the first vowel of /otɬo/ compared with an extremely raised F2 and lowered F3 in the first vowel of /ʔotɬno/. In Figure 7 there is a low F2 and a high F3 in the first vowel of /uta/ compared with an extremely raised F2 and a lowered F3 in the first vowel of /ʔutɬ’a/.

As these spectrograms demonstrate, the vowels affected by pharyngealization generally differ throughout their duration from their non-pharyngealized counterparts. Table 5 shows the mean values for vowels measured at the mid point of the first vowel in pairs like those in Figures 4-7. Each value is the mean of six tokens, from three repetitions each of two minimal or near-minimal pairs.

Table 5. Mean values for the first three formants of plain and pharyngealized vowels at vowel mid-point.

	plain			**pharyngealized**		
	F1	*F2*	*F3*	*F1*	*F2*	*F3*
i	320	2183	3065	411	1919	3060
e	636	1701	2741	669	1631	2751
a	766	1308	2730	804	1343	2639
o	533	1012	2397	553	1315	2241
u	348	937	3311	398	1491	2216

Measured at the mid-point of the vowels, the formant values of pharyngealized vowels are markedly different from those of plain vowels. The presence of pharyngealization significantly raises the first formant of the front vowels /i, e/, but has a smaller raising effect on the F1 of central or back vowels. It lowers the second formant of /i, e/ but raises that of /a/ and especially of /o/ and /u/. The particularly raised F2 of /o/ and /u/ in pharyngeal environments causes these vowels to make an auditory impression of being front rounded vowels, as noted for other Caucasian languages by Catford (1977, 1994) and for the Taiwanese language Amis by Maddieson and Wright (1991). The situation with the third formant is complex, since in pharyngealized cases a resonance is often apparent that really has no equivalent in the plain vowel. This is particularly the case with the /u/ and /a/ vowel qualities. The spectrograms in Figure 5 illustrate this problem rather clearly. In some instances the procedure we adopted results in measuring rather different resonances in different tokens, or at different time points in the same token. This should be borne in mind when considering the F3 results. The measured F3 is significantly lower in pharyngealized /a/ and /u/ than in their plain counterparts.

The spectrograms in Figures 4-7 also illustrate another important pattern. The formant differences between plain and pharyngealized vowels are greater at the vowel onset than at the mid-point (or the end of the vowel). The mean values measured at the onset are given in Table 6.

Table 6. Mean values for the first three formants of plain and pharyngealized vowels at vowel onset.

	plain			**pharyngealized**		
	F1	*F2*	*F3*	*F1*	*F2*	*F3*
i	298	2183	3099	466	1840	3010
e	533	1849	2776	714	1631	2917
a	778	1293	2876	862	1411	2192
o	510	934	2321	583	1341	2123
u	332	794	3229	395	1396	2043

The onset to mid-point differences are displayed graphically in Figure 8. Each panel of the figure displays the mean of the first three formants for one of the five vowel qualities in plain and pharyngeal environments. Formants for plain vowels are shown by open symbols, and for their pharyngealized counterparts by solid symbols. The general pattern for the pharyngealized cases is suggestive of an initial consonant imposing its transitions on the onset of the vowel, over and above any effects that may persevere through to the middle of the vowel as a syllable-level assimilatory effect. Either F1 and F2 or F2 and F3 are markedly closer together at the vowel onset than they are at the vowel mid-point. The spectrograms above also show evidence of a sharp onset to these words. As we noted in section 2, /ʔ/ occurs in initial position in the absence of any other consonant. However, /ʔ/ has minimal effect on the formant frequencies of adjoining vowels. It is present in each of the 'plain' vowel spectrograms on the left hand sides of Figures 4-7. A release that is often accompanied by some noise is seen in most of those in the 'pharyngeal' contexts on the right hand side of Figures 4-7. Formant structure can be followed in the noise portion of these releases, and particularly in the cases of /ʡadam/ in Figure 4 and /ʡutɬ'a/ in Figure 7 the transitional movements of the formants that were measured in the voiced portion of the vowels (and plotted in Figure 8) can be seen to originate from even more extreme positions. The most likely hypotheses to account for this pattern is that there is a specific oral consonantal constriction present at the beginning of these vowels, and that this is pharyngeal in place. This is why we suggest that the consonant inventory of Tsez includes the stop /ʡ/.

As Figure 9 illustrates, there is also a weak pharyngeal approximant which occurs in intervocalic position in a number of loanwords. This can be considered an allophone of the segment /ʡ/, which only occurs in the phonetic form of a stop in initial position. It is likely that the word-initial /ʡ/ varies between stop and continuant pronunciations but we have not examined this segment within utterances.

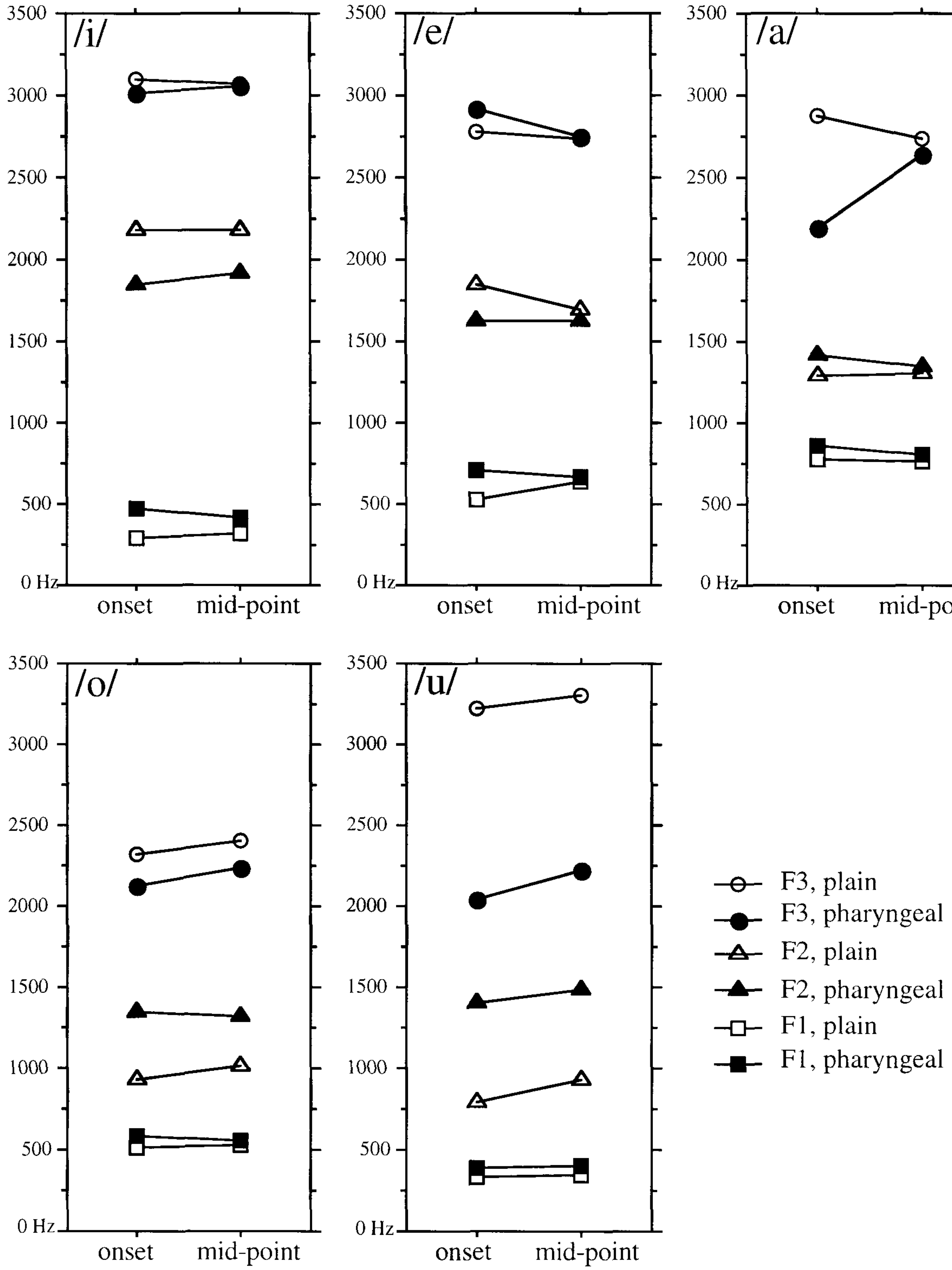


Figure 8. Mean values of the first three formants of plain and pharyngealized vowels at vowel onset and at the mid-point of the vowel.

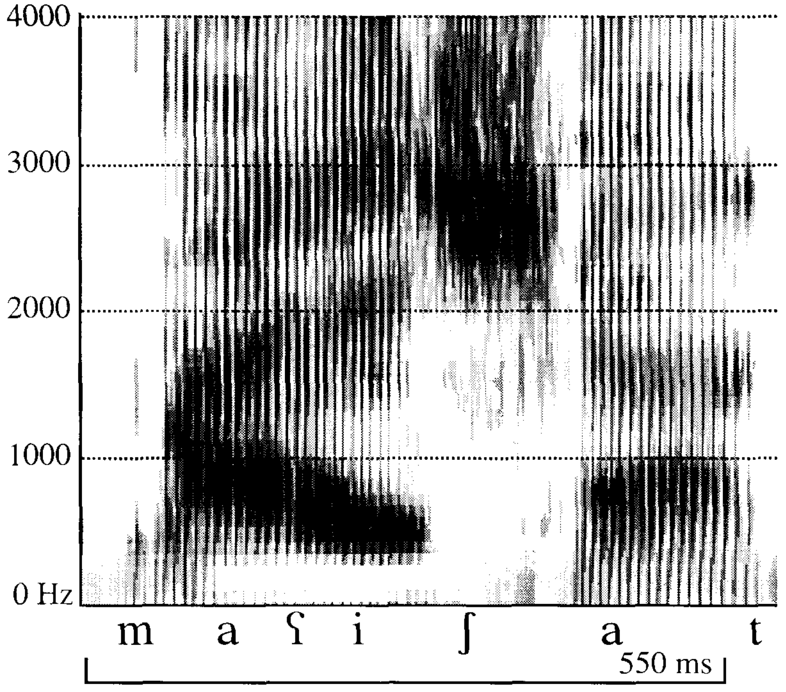


Figure 9. Spectrogram of the word /maʡiʃat/ 'belongings' illustrating the intervocalic approximant allophone of /ʡ/.

6. Pharyngealized uvulars

Illustration of the difference between plain and pharyngealized uvular plosives in final position is provided by the spectrograms in Figures 10 and 11. In the minimal pair /buq/ 'Sun' and /buqˁ/ 'hide!' in Figure 10, the pharyngealized uvular can be seen to have a lower 'locus' for F3 than the plain uvular. F3 lowers prior to the consonant /qˁ/ and can be seen to be rising in the aspiration noise of the stop's release. In contrast to this, the F3 rises at the end of the initial vowel in /buq/ and is flat after the stop release. However, apart from that portion of the vowel close to the uvular consonant closure, the formants of the two /u/ vowels in these words are very similar. The first and second formants are low, both lying below 1000 Hz, and a third formant is found at about 2300 Hz, with a stronger resonance occurring at about 3400 Hz. In other words, these vowels are unlike the pharyngealized /u/ vowels illustrated in Figures 7 and 8, which have a high F2. We found a similar absence of any marked raising of F2 in the vowel /o/ before pharyngealized uvulars.

Figure 11 shows spectrograms of the minimal pair /ʀaq/ 'side' and /ʀaqˁ/ 'wound' (N). There are some differences in the formant transitions going into the intervocalic consonants but very much stronger differences in the noise structure of the final releases. Pharyngealized uvulars after /i/ have already been illustrated in Figure 3. The differences between plain and pharyngealized vowels with the qualities /i, e, a/ are less marked than those for the back rounded vowels /o, u/ (see Figure 8), but the overall impression from the data we have examined is that the pharyngealization of a postvocalic pharyngealized uvular is not extended to the vowel that precedes it. There is, of course, some degree of coarticulation, but the evidence is against the idea that pharyngealization is a prosodic feature that spreads to all segments in a word.

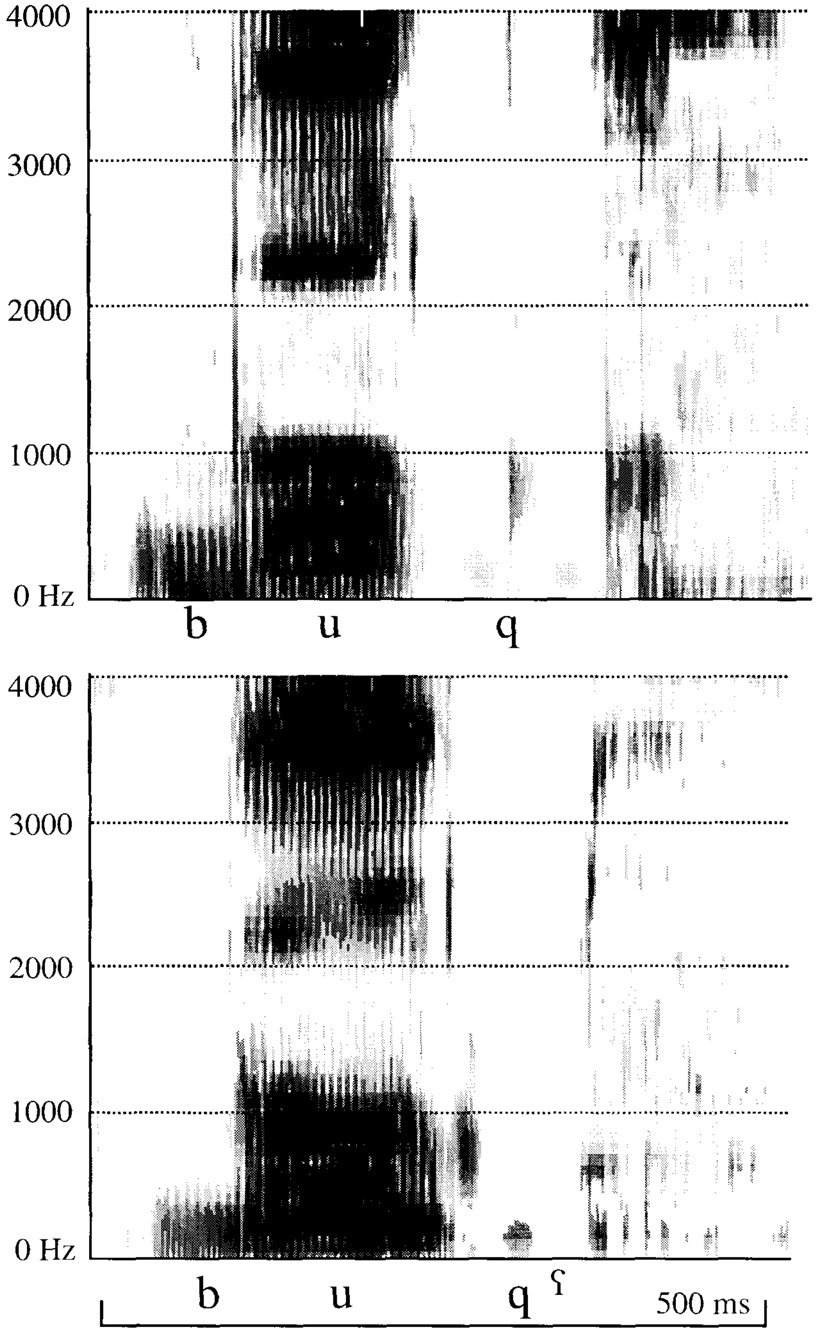


Figure 11. Spectrograms of the words /buq/ 'Sun' and /buqˤ/ 'hide' (imperative) illustrating the difference between final plain and pharyngealized uvular stops.

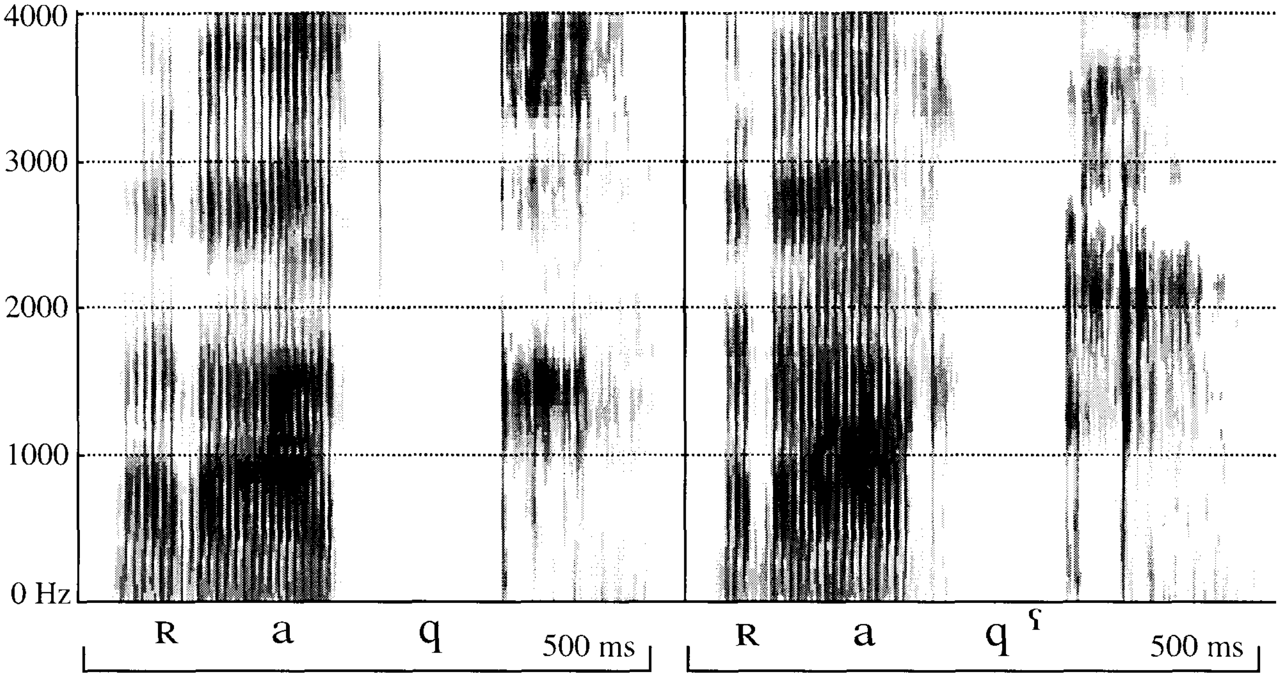


Figure 11. Spectrograms of the words /ʀaq/ 'side' and /ʀaqˤ/ 'wound' (N) illustrating the difference between final plain and pharyngealized uvular stops.

7. Directions for future research

The report on the phonetics of Tsez presented in this paper represents the speech of only one speaker. While we have no reason to believe that this speaker is atypical, it is important to extend these studies to multiple speakers in order to be more confident of the conclusions.

This paper is also confined to analysis of acoustic data. It would be desirable to be able to add articulatory 'flesh' to these acoustically-based observations. One major conclusion reached is that there are pharyngeal stops in Tsez, and another is that pharyngealization spreads only to the vowel that follows a pharyngeal or pharyngealized consonant. Articulatory studies using techniques such as high-speed MRI or fiberoptic laryngoscopy might be able to confirm these deductions. Many other facets of the articulation of Tsez sounds also deserve study.

We are also aware that there are many other topics of interest which should be dealt with in a more complete report, such as the nature of geminates, the production of labialization, and the systems of stress and intonation. It is our hope that we will be able to do at least some of these things in the future.

References

Alekseev, Mixail E. and Ramazan Rajabov (in press). Tsez language. In Michael Job and Rieks Smeets, eds. *Indigenous Languages of the Caucasus, Volume 3.* Caravan Books, Delmar, NY.

Bokarev, Evgenij A. (1959). *Tsezskie (Didojskie) Jazyki Dagestana.* Akademija Nauk SSSR, Moscow.

Bokarev, Evgenij A. (1967). Tsezkij jazyk. In *Jazyki Narodov SSSR, Volume 4*. Edited by V. V. Vinogradov. Nauka, Moscow: 404-420.

Catford, J. C. (1977). Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology* 6: 283-314.

Catford, J. C. (1991). The classification of Caucasian languages. In *Sprung from Some Common Source: Investigations into the Prehistory of Languages*, edited by S. M Lamb and E.D. Mitchell, Stanford University Press, Palo Alto CA: 232-268.

Catford, J. C. (1994). Vowel systems of Caucasian Languages. In *Papers from the Fourth Conference on Non-Slavic Languages of the USSR (1985)*, edited by H. I. Aronson, Slavica Publishers, Ann Arbor: 44-60.

Comrie, Bernard. (1981). *The Languages of the Soviet Union.* Cambridge University Press, Cambridge.

Kibrik, Aleksandr. E. and Sandro V. Kodzasov (1990). *Sopostavitel'noe izuchenie Dagestanskix Jazykov. Imja. Fonetika.* Moscow State University, Moscow.

Ladefoged, Peter and Ian Maddieson. (1996). *Sounds of the World's Languages.* Blackwells, Oxford.

Maddieson, Ian and Richard Wright. (1991). The vowels and consonants of Amis: a preliminary phonetic report. *UCLA Working Papers in Phonetics* 91: 45-65.

Nikolaev, Sergej L. and Sergej A. Starostin. (1994). *A North Caucasian Etymological Dictionary.* Moscow: Asterisk

The Phonetics of Paicî Vowels

Matthew Gordon and Ian Maddieson

1. Introduction

Paicî is one of the twenty-five or so Austronesian languages indigenous to the French "overseas territory" of New Caledonia. It is spoken in a narrow band stretching across the center of the main island, Grande Terre, from Poindimié and Ponerihouen on the east coast to Pouembout and Koné on the west. The location is shown in the map in Figure 1. The number of speakers was estimated at approximately 5000 in the 1970's. It is now likely to be on the order of 8000 as the indigenous component of New Caledonia's population has shown vigorous growth in the last few decades. Paicî has the largest number of speakers of any of the languages of Grande Terre, and because of its importance it has been chosen by the territorial education authorities as the first local language to be (re-)introduced into use in primary education.

Figure 1. Map of New Caledonia showing areas in which the various indigenous languages are distributed (based on a map prepared by Ozanne-Riviere and Rivierre). Language names are italicized.

Ironically, recent administrative changes which have given the indigenous people greater influence in the political life of New Caledonia might pose a threat to the future of even the stronger languages, such as Paicî. This is because French is the main language for inter-ethnic communication, is needed for participation in the modern political and economic life of the territory, and most school-based education is only available in French. In view of these factors the number of speakers might well be expected to decline in future years even as the ethnic population grows.

From a phonetic point of view, Paicî is principally of interest for its extensive vowel system. It has ten oral vowels and seven nasalized vowels, which makes Paicî something of a rarity cross-linguistically. Only 26, or 8.1% of the languages in UPSID sample of languages of the world (Maddieson 1984) distinguish ten or more contrastive vowel qualities, and most languages with nasalized vowels contrast no more than five qualities in this category (Maddieson 1991). Paicî is also one of a relatively small number of Austronesian languages which is tonal (Rivierre 1974, 1978).

Yet, despite the exceptional nature of its vowel system and its important political status in New Caledonia, Paicî has not been the subject of any previous detailed phonetic examination. Besides the two articles by Rivierre cited immediately above, the only published general linguistic works on Paicî are a substantial dictionary published by Rivierre (1983), which we have relied on extensively, and a short grammar and word list which appeared in Leenhardt's (1946) comprehensive book on East Melanesian languages under the name *Pati*. Rivierre's dictionary is founded in part on materials collected earlier by Leenhardt, Grace, Haudricourt and others but was entirely checked in the field and much expanded.

Following a brief discussion of the overall phonological inventory the acoustic structure of the Paicî vowel system will be examined from two angles below. First, the location of the oral vowels in the acoustic space will be described, in order to add to the relatively small cross-linguistic phonetic data base on large vowel inventories. By providing further data on how the vowels in large vowel systems are distributed we hope to contribute to understanding the principles governing this distribution. Second, the acoustic patterns of the nasalized vowels will be examined. Numerous studies have shown phonetic asymmetries between oral and nasalized vowels (e.g. Wright 1986, Maeda 1993). Given its large number of nasalized vowels, Paicî provides a particularly good opportunity to examine the acoustic consequences of nasalization. This paper is only a preliminary report, and it is hoped that more detailed studies will follow.

2. Consonants

Compared to a number of other New Caledonian languages, including Iaai (Maddieson and Anderson 1995) and Ndumbea (Gordon and Maddieson 1995), Paicî has a relatively simple consonant system, which is presented in Table 1. Nasals, voiceless plosives and prenasalized voiced stops occur at four places of articulation. The segments written in the 'alveolar' column are in fact slightly post-alveolar. The palatal stops are released with quite strong frication and could well be considered to be affricates. Plain and labialized bilabial stops and nasals contrast before non-rounded vowels (before /ɔ/ the bilabials are noticeably labialized in pronunciation, but do not contrast with non-labialized counterparts). There is also a labial-velar approximant /w/ which is similarly restricted in its distribution, occurring only before non-rounded vowels and, occasionally, /ɔ/. For convenience, this is placed in the same column as the labialized bilabials on the chart. The apical tap /ɾ/ also has a restricted distribution, as it only occurs word-initially in one grammatical element (/ɾʌ/ 3 pl subject prefix). In a very high proportion of its occurrences the same vowel precedes and follows the tap, suggesting that a process of epenthesis may be involved. When this segment precedes a nasalized vowel it is also nasalized or is pronounced as a very brief nasal segment. The alveolar lateral approximant /l/ occurs initially only in a few loanwords, such as /lāācí/ 'rice', but is relatively common at the beginning of the second syllable of a word. There are no consonant clusters or syllable-final consonants in Paicĩ.

Table 1. Consonants of Paicĩ (after Rivierre 1983: 21)

	Bilabial	Labialized Bilabial	Alveolar	Palatal	Velar
Voiceless plosive	**p**	**p^w**	**t**	**c**	**k**
Voiced pre-nasalized stop	**mb**	**mb^w**	**nd**	**ɲɟ**	**ŋg**
Nasal	**m**	**m^w**	**n**	**ɲ**	**ŋ**
Approximant		**w**	**l**		
Tap			**ɾ**		

3. Vowels and Tones

As noted earlier, Paicĩ distinguishes ten oral and seven nasalized vowel qualities. The ten oral qualities are given in Table 2, using the symbols suggested by Rivierre. As this table shows, Paicĩ reaches a total of ten by adding a series of three back (or at least non-front) unrounded vowels to the 'standard' 7-vowel set. This type of expansion of a vowel inventory is less common than the addition of front rounded vowels (Maddieson 1984). We will consider below the question of whether the acoustic analysis supports the suggested notation, particularly with respect to the degree of backness indicated for these vowels.

Table 2. Oral vowels of Paicĩ.

i **ɯ u**
e **ɤ o**
ɛ **ʌ ɔ**
a

The seven nasalized vowel qualities are given in Table 3, again using the symbols suggested by Rivierre. Vowels are always nasalized after nasals, and always oral after prenasalized voiced stops, but contrast elsewhere, as in the pair /cóo/ 'dress, cloth' vs /cṍõ/ 'look, watch'. In the nasalized vowel set the contrast between higher mid and lower mid qualities that is present among the oral vowels is lacking. Rivierre indicates that the remaining nasalized mid vowel is higher mid in the front unrounded and back rounded series, but lower mid in the back unrounded series. We will consider below the question of whether this reflects an articulatory difference between these vowels or could be accounted for by acoustic effects alone. Rivierre also suggests that the low nasalized vowel /ã/ varies in range between [ã] and [ɛ̃]. We will look for confirmation of the greater acoustic range of this vowel in our data.

Table 3. Nasalized vowels of Paicĩ.

ĩ **ɯ̃** **ũ**

ẽ **õ**

ʌ̃

ã

Each of the vowel qualities in Tables 2 and 3 can occur distinctively long or short. The orthography used in Rivierre's dictionary as well as in recent literacy and evangelical materials marks long vowels by doubling the vowel symbol. We have maintained this convention in our transcriptions. (The orthography otherwise uses only the five basic symbols of the Roman alphabet for vowels, requiring the resort to a variety of diacritics to mark further distinctions of height, rounding and nasalization.) Sequences of unlike vowels also occur quite frequently, although it is rare to find two long vowels adjacent.

Rivierre (1974) distinguishes three level tones, high [´], mid [¯] and low [`]. Some items are also 'toneless' in that their tone is predictable from the tone of the item they follow (usually they are at the same level). Immediately adjacent short vowels may bear different tones but a long vowel never bears more than one tone. A very high proportion of Paicĩ words with more than one syllable have the same tone on all their syllables, suggesting that the tonal patterns may be characteristics of the world. We have marked tones on items cited in this paper following Rivierre but have not made any independent study of their production.

4. Acoustic Analysis

4.1. Materials

A wordlist illustrating all the consonants and vowels of Paicĩ was prepared on the basis of the dictionary and the guide to the orthography prepared for schools. Words selected to illustrate consonants contained principally /a/ vowels, and words selected to illustrate vowels principally contained bilabial consonants (the high frequency of bilabials making this context the easiest in which to find minimal or near-minimal contrasts among the vowels). This list was recorded using a high-quality directional microphone with two groups of speakers in outdoor settings near Napoemi and Tibarama in the region of Poindimié. The participants in these recordings were eight native speakers of Paicĩ, five men and three women, ranging in age from their late forties to early twenties. The actual words recorded differed slightly in the two sessions, and some individual tokens were unusable because of wind noise, overlaid voices or other problems. These recordings, made in New Caledonia in February 1993 by the second author, serve as the basis for our analysis of the Paicĩ vowel system.

4.1. Oral Vowels

In order to describe the principal acoustic characteristics of the Paicĩ vowels the frequencies of the first three formants were measured in a selected subset of the words recorded. The words from which the measurements were made are listed in Appendix 1. The analyses were performed using the Kay Elemetrics Computerized Speech Laboratory (CSL) The audio recordings were digitized at 20 kHz and formants obtained algorithmically using LPC analysis. Either 12 or 14 coefficients were employed, depending on the speaker's sex and the difficulty resolving the formants. A window of about 25 ms (26.5 ms) centered around the middle of the vowel's duration was chosen for the LPC. An FFT power spectrum calculated over the same window was consulted for confirmation of the values obtained from the LPC analysis. An alternative placement of the analysis window was tried when LPC and FFT analyses appeared in conflict.

Values for long and short vowels are generally very similar, with the only consistent and sizable difference between long and short counterparts being that between /ɯ/ and /ɯː/. For this pair, the overall mean F2 value of the short vowel is higher than that of the long vowel (1596 Hz for /ɯ/ and 1419 Hz for /ɯː/) and a difference is found for every speaker for whom at least one token each of short /ɯ/ and long /ɯː/ was measured (all speakers except M5). However, the difference is not statistically significant, perhaps because of the small number of tokens involved, and in the plots and tables which follow long and short /ɯ/ are not separated.

The measured values of the first two formants of each vowel token for the five male speakers are plotted together in Figure 2. In this and subsequent figures of the same type, the origin for both scales is in the upper right of the figure. The distances along the formant scales are denominated in Hz but proportional to intervals in Barks. Each vowel symbol on the chart represents one or more tokens of the vowel indicated by the symbol. An ellipse is drawn around each cluster of points representing a single vowel type. The ellipses have a radius of two standard deviations along the axes of the first two principal components of the distribution.

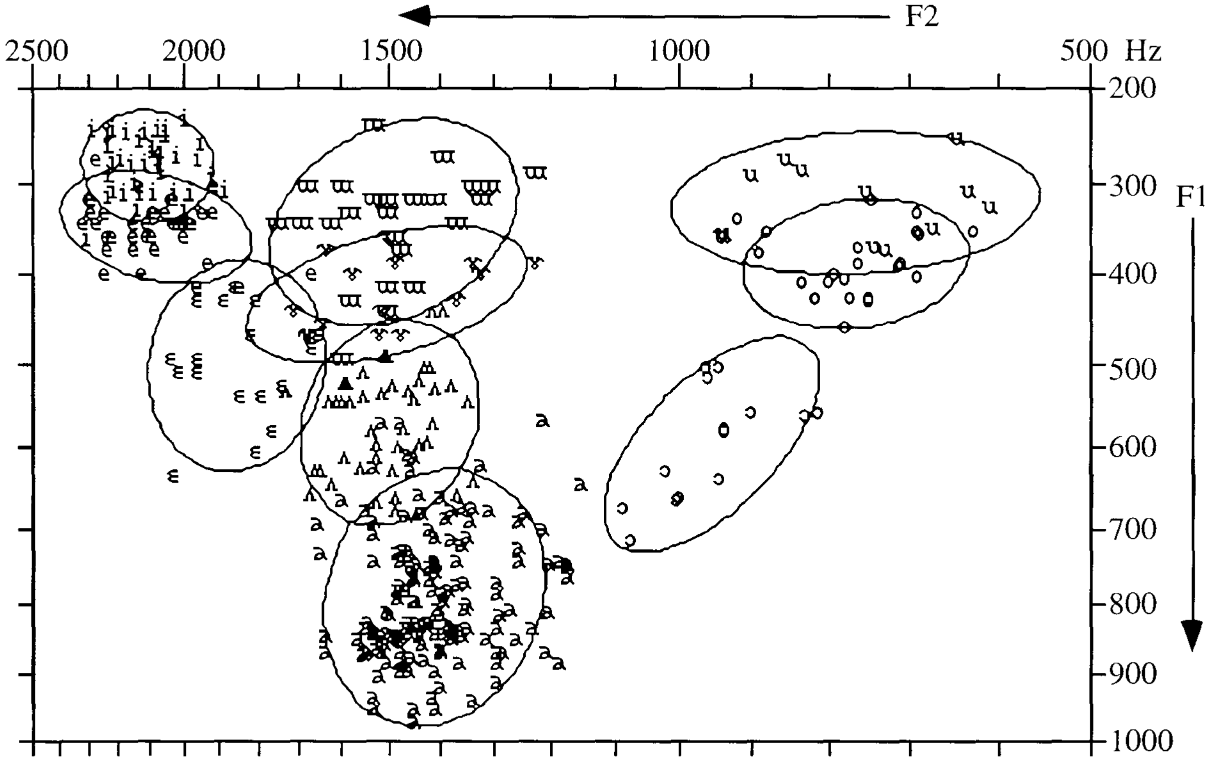


Figure 2. Values of the first two formants in oral vowels, male speakers, all tokens.

The scatter of points for each vowel in Figure 2 includes some variation attributable to differences in consonantal context, especially for the /a/ vowels where a wider range of contexts was available, and some attributable to cross-speaker differences. To show a clearer picture of the typical vowel positions, the mean formant values for each vowel for each individual speaker were calculated and each mean plotted as a single point. These data for the male speakers are shown in Figure 3. The residual scatter in this figure is mainly attributable to inter-speaker differences. The large scatter seen for /a/ in Figure 2 is much reduced in Figure 3 and is thus shown to be due largely to the contextual effects.

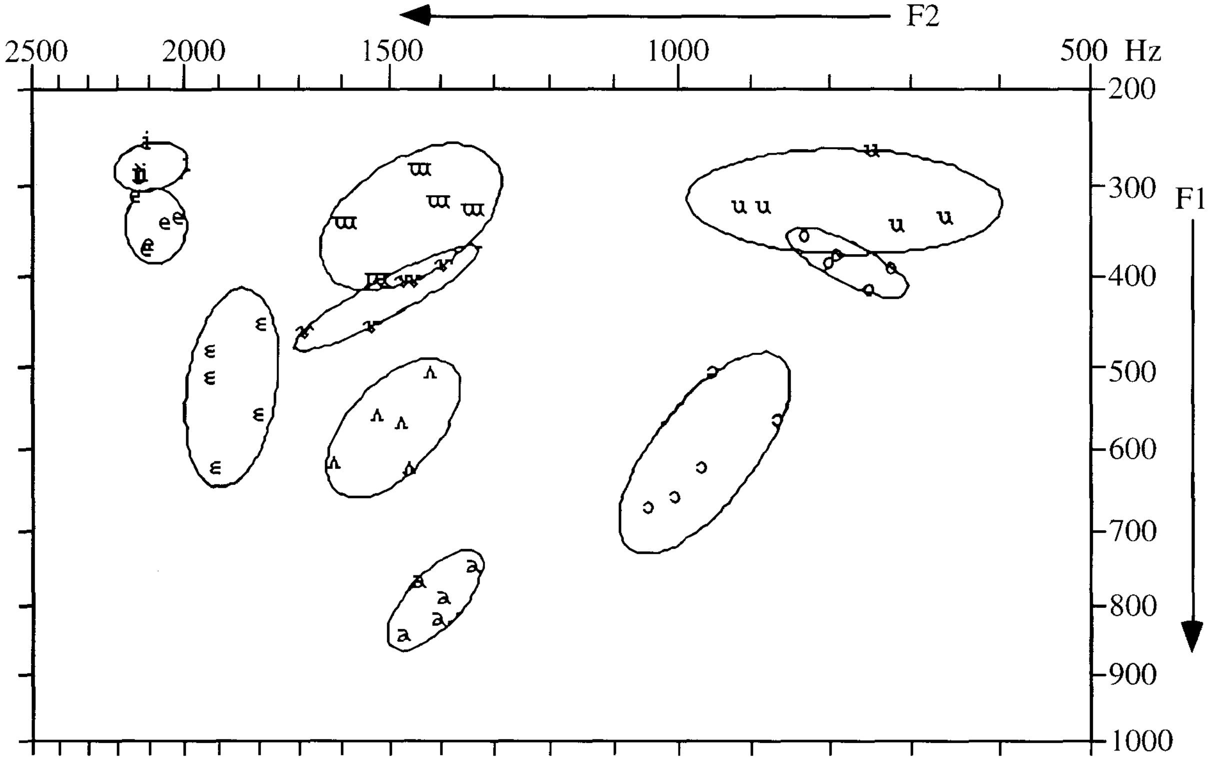


Figure 3. Values of the first two formants in oral vowels, male speakers, mean value for each vowel.

Similar displays of the distribution of the vowels of the female speakers in a two-formant space are shown in Figures 4 and 5. Figure 4 shows the scatter of points from all the tokens measured, and Figure 5 plots just the means for each vowel for each speaker.

Figures 2-5 show substantial differences in the degree to which vowels are separated in the two-formant space. The three high vowels and the three higher mid vowels form three pairs whose members are not far apart, especially in the case of the pair /ɯ/ and /ɤ/ for the female speakers. The three lower mid vowels on the other hand are more clearly separated from their higher mid counterparts. The relative crowding of the upper part of the acoustic vowel space may be what permits the quite large inter-speaker differences in the first formants of these lower mid vowels seen in figures 3 and 5; since the higher-mid vowels are relatively high, speakers are free to choose a realization of the lower mid vowels anywhere within a rather broad region.

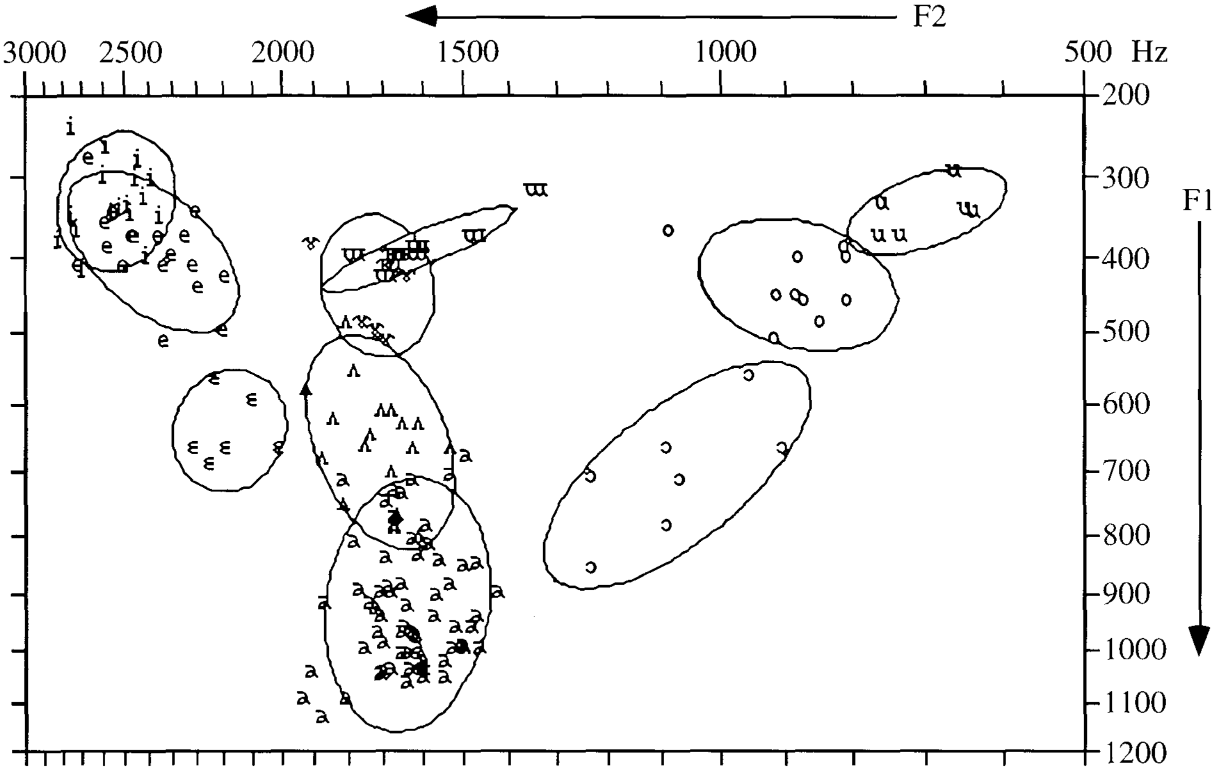


Figure 4. Values of the first two formants in oral vowels, female speakers, all tokens.

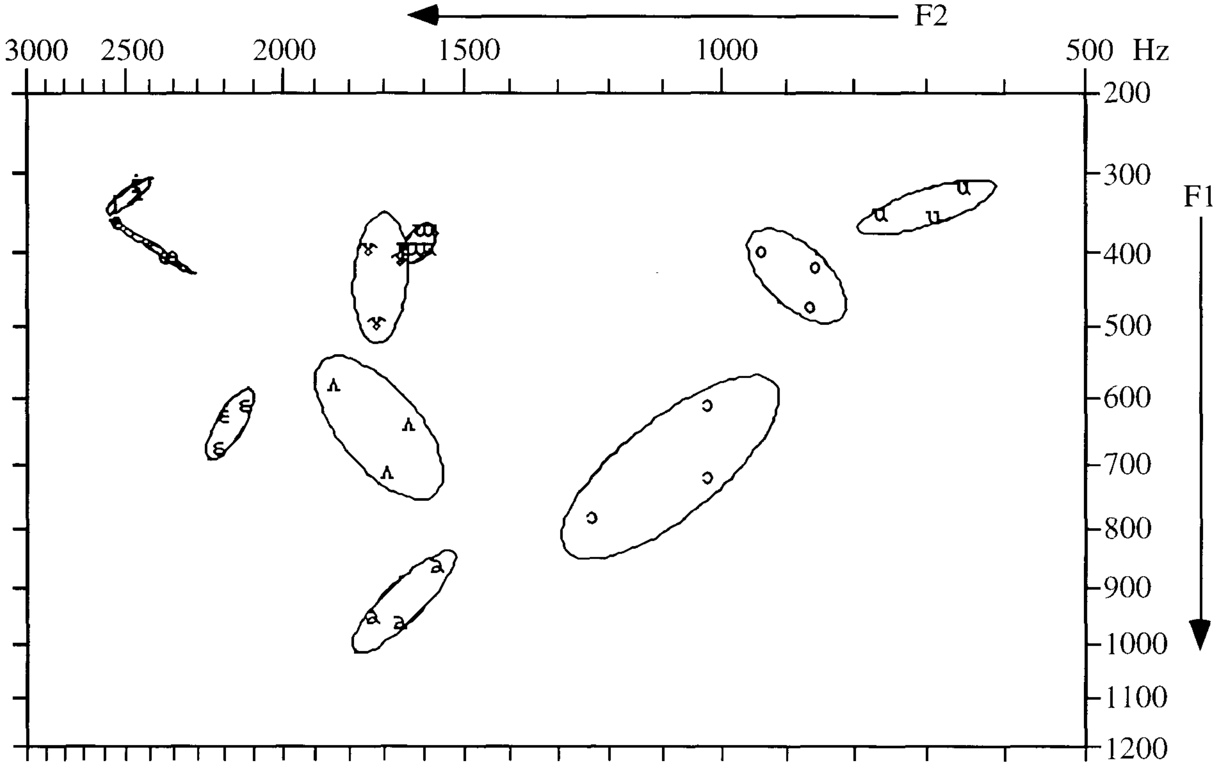


Figure 5: Oral vowels, female speakers, mean value for each vowel.

All the speaker means for the first three formants of each vowel, as well as the number of tokens measured, are shown in Appendix 2.

As noted in connection with Figure 2, the high and higher mid vowels are not greatly separated in the F1/F2 space, but the front pair i/e have quite distinct third formant values. The F3 for /i/ is about 300 Hz higher than that for /e/, the mean of the speaker means for male speakers being 2948 for /i/ and 2646 for /e/ (significant at the .0015 level). This formant presumably assists in differentiating these two vowels, but for the other two vowel pairs concerned there is no reliable difference in F3 in our data.

One of the main features of interest in the Paicĩ vowel system are the acoustically 'interior' vowels transcribed /ɯ, ɤ, ʌ/. For the majority of speakers studied the F2 values of these vowels suggest that these may not be truly back in their articulation — lower second formants would be expected — but might better be considered as central vowels. For the sake of consistency, however, we will continue to use Rivierre's transcription.

4.2. Nasalized Vowels

The values of the first three formants of the nasalized vowels were also estimated, using the same procedure outlined above for the oral vowels. We are aware of the problems involved in modeling nasalized vowels with an LPC analysis, but note that our method involves cross-checking with FFT spectra. We feel that a reasonable characterization of some of the most important acoustic properties of these vowels can be obtained in this way. The speaker-means of the resulting measurements are included in Appendix 2. Due to the absence of long vs. short pairs for certain nasalized vowels and an overall paucity of tokens containing nasalized vowels , values for long and short nasalized vowels are collapsed together. Differences in the nasalized vowels dependent on length are discussed later in this section.

Means of the first two formants for each of the nasalized vowels for the male speakers are plotted in figure 6, and for each of the female speakers in Figure 7. The scales are the same as in Figures 4 and 5.

The nasalized high vowels /ĩ, ũ/ have similar values of the first two formants to their oral counterparts, suggesting that they do not differ much from them with respect to their oral articulations. On the other hand, /ɯ̃/ has a substantially lower second formant than /ɯ/ among the female speakers. It is possible that this vowels is produced by some speakers with a more retracted position than its oral counterpart. But as only a single token of this vowel, and that following a labialized consonant, was available, any interpretation should be cautious.

Among the mid nasalized vowels the front vowel /ẽ/ has a substantially higher F1 than the oral vowel /e/, and compares more closely with the formants of /ɛ/. For the male speakers at least the first formant of /õ/ is substantially higher than that of /ẽ/. Compared to its oral counterpart /o/, this vowel has a higher F1 and F2, placing it more in the perceptual center of the vowel space. Formant values for /ʌ/ and /ʌ̃/ do not differ in a consistent direction, though there is a tendency for F1to be slightly higher than for the oral vowel /ʌ/. The nasalized low vowel /ã/ has a considerably lower F1 than its oral counterpart /a/, reducing the distance between this vowel and /ʌ̃/.

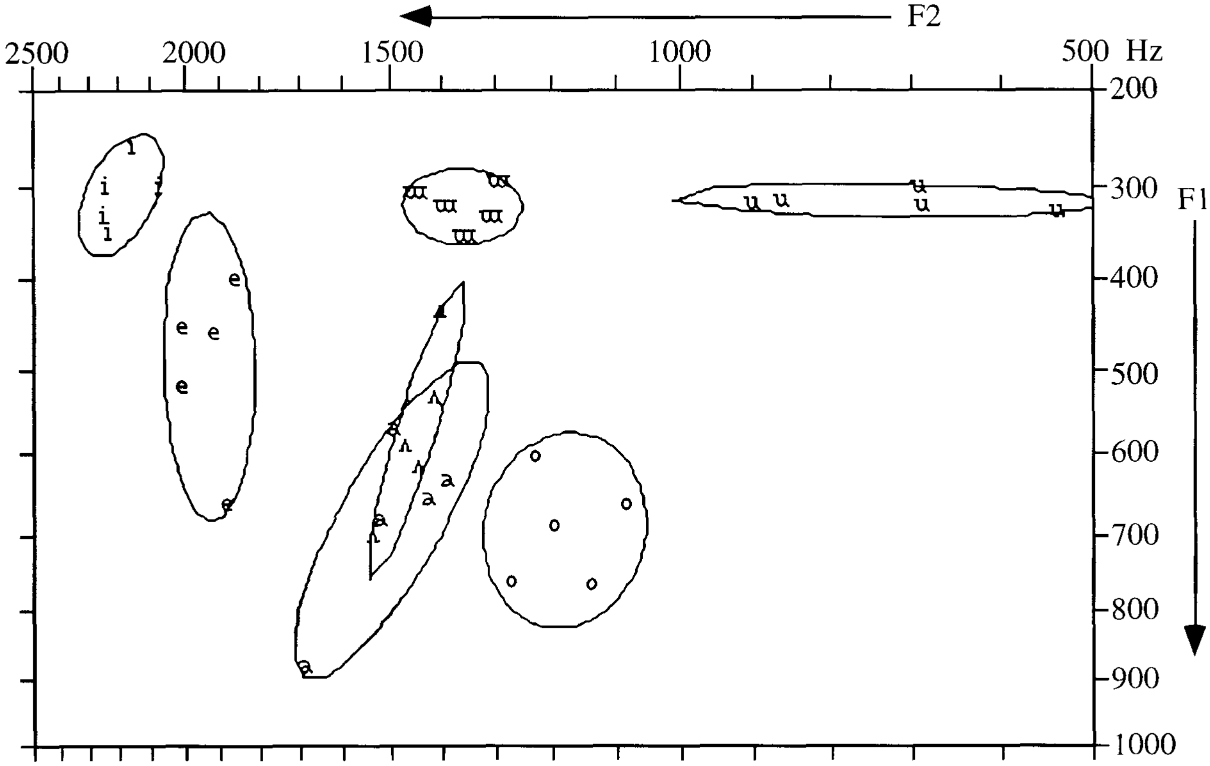


Figure 6. Mean values of first two formants of nasalized vowels for male speakers.

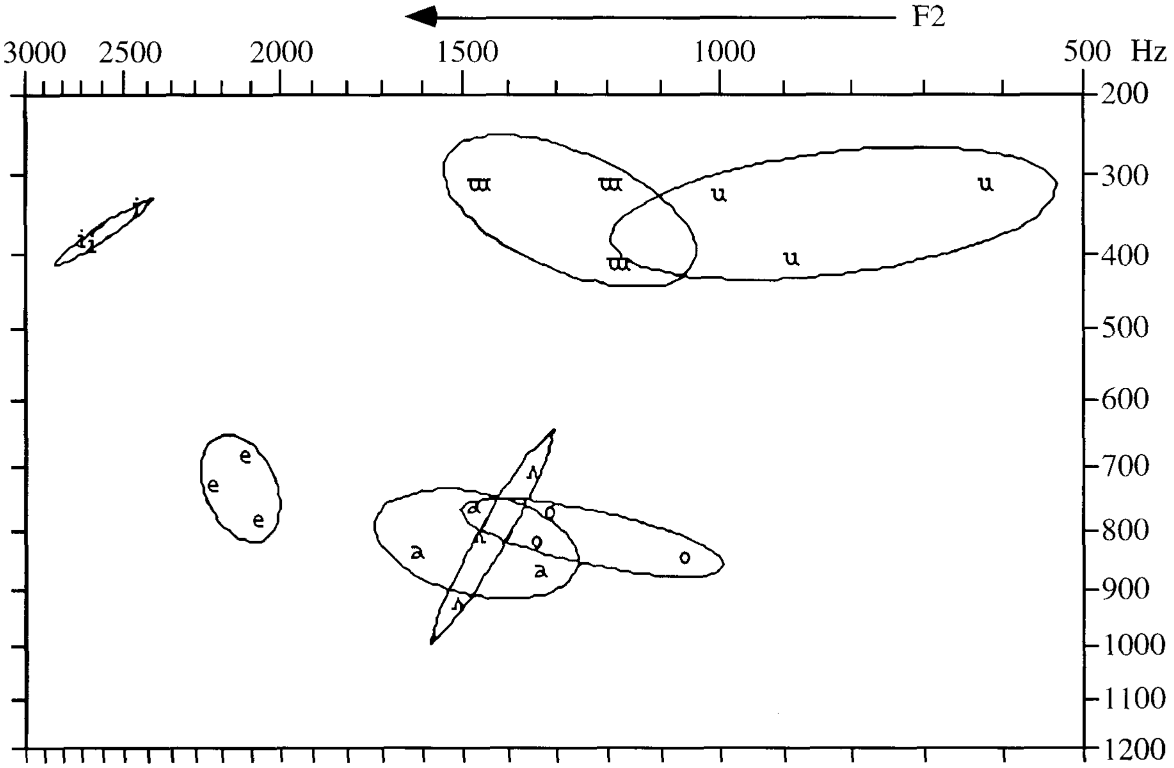


Figure 7. Mean values of first two formants of nasalized vowels for female speakers.

In summary, the high nasalized vowels remain high in the acoustic space, the mid nasalized vowels appear to be acoustically closer to the lower mid vowels in the oral set rather than the higher mid vowels, and the low nasalized vowel is acoustically raised.

5. Discussion

It is a common cross-linguistic observation that nasalized vowels tend to be more centralized than their oral counterparts. Wright (1980) suggests that the addition of a nasal formant in the vicinity of F1 triggers a perceptual shift in the height of vowels, as the 'center of gravity' of low-frequency energy in the vowel is shifted. Because the first nasal formant has very similar values across all vowel qualities (Maeda 1993), its presence has different perceptual effects on vowels of different heights. For non-low vowels, the first nasal formant is typically higher than the first oral formant, resulting in an upward shift in the central focus of low frequency energy. The perceptual consequence of this shift is that non-low nasalized vowels may be perceived as lower than their oral counterparts; over time, this may lead to an articulatory adjustment, as in the history of French. The apparent low position of the nasalized mid vowels in Paicĩ may be a reflection of this tendency.

In low vowels, by contrast, the first nasal formant is lower than the first oral formant; the result is a downward shift in energy when the vowel is nasalized and, hence, creates the perception of a raised vowel. Our measurements of /ã/ in Paicĩ as having a lowered F1 probably reflect just the acoustic effect anticipated, and not any articulatory adjustment.

The effect of nasalization on pereceived vowel backness is less well understood with effects observed in natural language at odds with those predicted by perception models (Beddor 1993: 182). Beddor suggests that the nasalization of back vowels might lead to "perceptual retraction" due to a decrease in the separation of F1 and F2 resulting from the addition of a nasal formant between the first two oral formants. This effect of nasalization should not be as great in front vowels due to the relatively great distance between the first two oral formants. Wright, however, found that the greatest effect on perceived backness occurred in the front vowel pair /i/ and /ĩ/. He also found that nasalization created the percept of fronting in /õ/, which is in line with the Paicĩ findings.

The effect of nasalization on natural speech vowels other than peripheral vowels has not been examined in the literature, due to the paucity of languages with 'interior' nasalized vowels. Our data contain few tokens of these vowels, but we may make some interpretative suggestions. Recall that in Paicĩ, nasalization did not appear to have a consistent effect on /ʌ/, while F2 of /ɯ̃/ was found to be often lower than F2 of /ɯ/. It may be that /ʌ̃/ has an F2 close to that of a nasal formant and hence its central frequency is not modified, but it could also be that articulatory efforts are made to keep values for /ʌ̃/ similar to those of its oral counterpart because of the proximity of nasalized /ʌ̃/ to both /õ/ and /ã/. Were /ʌ̃/ shifted either downward or backward in the vowel space, it would perceptually merge with these neighboring vowels. /ɯ̃/ on the other hand, has more room for backing, since there are no particularly nearby vowels with which it would merge. The nearest back vowel, /u/ is substantially more backed than /ɯ̃/.

Acknowledgments

This work would not have been possible without the generous cooperation of the speakers who shared their knowledge of their language, and the hospitality of the community. Much essential assistance was provided by Stephen Schooling. The project was funded by the National Science Foundation.

References

Beddor, Patrice S. 1993. Perception of nasal vowels. In *Nasals, Nasalization and the Velum*, ed. by M. Huffman and R. Krakow. Academic Press, New York: 171-196.

Leenhardt, Maurice. 1946. *Langues et Dialectes de L'Austro-Mélanésie.* Institut d'Ethnologie, Musée de l'Homme, Paris.

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

Maddieson, Ian. 1991. Testing the universality of phonological generalizations with a phonetically-specified segment database: results and limitations. *Phonetica* 48: 193-206.

Maddieson, Ian and Victoria B. Anderson. 1995. The phonetics of Iaai. *UCLA Working Papers in Phonetics* 87: 163-182.

Maeda, Shinji. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, Nasalization and the Velum*, ed. by M. Huffman and R. Krakow. Academic Press, New York: 147-170.

Rivierre, Jean-Claude. 1974. Tons et segments du discours en langue paici (Nouvelle Calédonie). *Bulletin de la Société de Linguistique de Paris* 69: 325-340.

Rivierre, Jean-Claude. 1978. Accents, tons et inversion tonale en Nouvelle Calédonie. *Bulletin de la Société de Linguistique de Paris* 73: 415-443.

Rivierre, Jean-Claude. 1983. *Dictionnaire Paicî-Français.* SELAF, Paris.

Wright, James T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In *Experimental Phonology*, ed. by J. J. Ohala and J. J. Jaeger, Academic Press, Orlando: 45-67.

Appendix 1. Wordlist used for analysis of vowel formants

(´ = high tone,¯=mid tone, `=low tone)

i	pī	crab, fish eggs
	mbīmbīī	wet, soaked
	kīī	key
	ápí	yam stake
	ã́mbíi	(be in) mourning
	ã́ndí	traditional money
ĩ	mĩ̄	liana, tropical creeper
	mĩ̄ĩ̄	red, yellow
	mĩ́mĩ́	gnat
	kĩ́ɾĩ́	little
e	pé	ray
	mbēē	companion
	mbwēē	buttocks
	ŋgē	feast, ceremony
	ācēē	mosquito
	ãmbwé	flying fish
	mbéré	abandoned settlement

ã	mwã̄tō	share of food
	mwãɾã	again
	mwã̄ɾĩ̄	money
	wã	clan, house
	tã̄ã̄	family
	kã̄ã̄	cry
	ãwãã	The Amoa river
	ã̄mbʌ̄ī	shark
ʌ	pʌ̄ʌ̄	foundation (of house)
	mbʌ̄	death
	mbwʌ̄	remainder
	ɾʌ	they
	kʌ̄kʌ̄	crow
	ātʌ́	pandanus basket
ʌ̃	pwʌ̃̄	tortoise
	pwʌ̃̄ŋge	your (sg.) mouth

ẽ	mẽ̄ẽ̄	eyes
	pwẽ̄ẽ̄	his mouth
	kẽ̄ẽ̄/kēē	small shell
ɛ	mbɛ̄ɾɛ̄	brother
	kɛ̄	crab (species)
a	pāɾāwē	skin
	mbā	wall
	pwáa	white
	mbwátɔ̃́	short
	ndāaé	his spine
	ndā	spear
	lāací	rice
	cś	oven
	ɲɟā	year, season
	ɲjāā	mountain
	kākẽ̄	manatee
	ŋgāandāaé	his forehead
	ŋgāɾā	taro terrace
	ápí	yam stake
	āpwɯ̄	shoulder
	ārāo	my front
	ámbwá	limestone, chalk
	ācēē	mosquito
	āŋgā	parrot (species)
	āɲɟīʌ	Ajië (a New Caledonian language)
	ákāē	yes, ok
	alala	touché!
	ātʌ́	pandanus basket

ɤ	mbwɤ̄/mbwē	banyan tree
	kɤ̄ɤ/kũ̄ũ̄kũ̄/kẽ̄ẽ̄	deaf
	kɤ̄ɾɤ̄/kēɾē	spider
ɯ	mbɯ̄ɾɯ̄	war token
	mbwɯ́ɯ́	green
	ndɯ̄ɯ̄ɾɯ̄	mud
ɯ̃	mwɯ̃	drown, flood
ɔ	kɔ̄ɔ̄	egret
	kɔ̄rēe	grasshopper
	ŋgɔ̄ɔ	crab
o	kóó	cold, frost
	kōōnẽ	Koné (place)
	pópáa	rain
	mbōo	descend
õ	cṍṍ	look at, see
	mȭtɔ̄	grass, “bush”
	mȭȭ	cold
u	mbū	1 pl incl pn (“we two”)
	mbūɾɛ̄	seashell
	kúɾú	“bougna”
ũ	pũ̄ɾũ̄	head
	mũ̄	flower

Appendix 2. Mean values of the first three formants of each vowel for each speaker. Speakers are identified by a letter indicating gender (M or F) and a number.

Speaker & vowel		**F1**	**F2**	**F3**	**Speaker & vowel**		**F1**	**F2**	**F3**
M1	n				**M5**	n			
i	13	288	2129	3008	i	6	291	2139	2743
ĩ	5	365	2240	3038	ĩ	4	309	2248	3200
e	12	364	2111	2636	e	6	340	2064	2533
ẽ	4	477	2006	2693	ẽ	2	567	1927	2623
ɛ	4	513	1930	2534	ɛ	2	558	1807	2489
a	38	789	1397	2369	a	26	820	1409	2470
ã	8	609	1492	2122	ã	4	734	1523	2116
ʌ	11	556	1526	2388	ʌ	7	621	1459	2481
ʌ̃	3	564	1407	2660	ʌ̃	1	657	1442	2294
ɤ	3	454	1544	2162	ɤ	3	386	1397	2607
ɯ	7	403	1534	2139	ɯ	2	320	918	2503
ɯ̃	3	367	1356	2032	ɯ̃	1	344	1310	2275
ɔ	4	562	870	2429	ɔ	2	672	1048	2416
o	8	390	727	2456	o	3	354	833	2436
õ	3	717	1161	2185	õ	3	742	1198	2323
u	3	335	662	2315	u	2	324	918	2395
ũ	2	338	537	2052	ũ	3	327	862	2152
M2					**F1**				
i	8	288	2152	3101	i	8	344	2658	3160
ĩ	3	346	2251	3002	ĩ	4	375	2720	3214
e	8	368	2115	2796	e	7	409	2400	3026
ẽ	2	551	2010	2726	ẽ	3	644	2180	3301
ɛ	2	620	1917	2538	ɛ	2	628	2207	3027
a	32	841	1479	2454	a	27	958	1656	2720
ã	4	961	1686	2235	ã	4	713	1615	2407
ʌ	9	615	1621	2399	ʌ	8	713	1812	2966
ʌ̃	2	757	1535	2362	ʌ̃	2	664	1488	2458
ɤ	3	460	1692	2225	ɤ	3	497	1726	2647
ɯ	6	340	1600	2241	ɯ	4	399	1641	2410
ɯ̃	1	331	1393	2317	ɯ̃	1	399	1337	2551
ɔ	3	658	1011	2259	ɔ	3	721	1021	3000
o	5	413	754	2321	o	4	476	864	2808
õ	3	710	1082	2276	õ	2	721	1456	2386
u	3	343	719	2363	u	2	357	690	2099
ũ	1	331	689	2220	ũ	1	303	813	2068

M3					**F2**				
i	7	279	1998	3022	i	6	**325**	2544	3409
ĩ	7	326	2079	3065	ĩ	5	**371**	2761	3370
e	9	334	2018	2523	e	6	**363**	2643	3271
ẽ	3	419	1857	2669	ẽ	3	**735**	2141	3066
ɛ	7	453	1796	2421	ɛ	2	**676**	2221	3138
a	29	746	1342	2245	a	16	951	1732	3063
ã	4	678	1387	2288	ã	3	**781**	1738	2636
ʌ	10	507	1418	2206	ʌ	5	**583**	1838	3090
ʌ̃	1	458	1396	2347	ʌ̃	1	**757**	1597	3076
ɤ	3	404	1477	2093	ɤ	2	**782**	1662	2089
ɯ	6	319	1406	2149	ɯ	3	**372**	1595	2556
ɯ̃	1	303	1296	1834	ɯ̃	1	**303**	1599	2275
ɔ	3	508	955	2094	ɔ	2	**782**	1228	3179
o	3	376	792	2175	o	3	**402**	937	2510
õ	3	823	1271	1848	õ	3	**765**	1483	2437
u	2	322	888	2143	u	2	**320**	649	2347
ũ	4	334	897	1976	ũ	1	413	937	2298
M4					**F3**				
i	6	256	2114	2867	i	6	**316**	2561	3111
ĩ	5	263	2162	2960	ĩ	6	**331**	2544	3027
e	6	312	2149	2742	e	6	**408**	2432	2887
ẽ	3	483	1914	2694	ẽ	1	**684**	2285	2768
ɛ	2	482	1931	2710	ɛ	2	**614**	2124	2737
a	26	767	1447	2430	a	24	**860**	1567	2523
ã	4	704	1423	2476	ã	4	**807**	1478	2169
ʌ	7	566	1475	2460	ʌ	5	**641**	1630	2779
ʌ̃	1	629	1466	2551	ʌ̃	1	**855**	1640	2591
ɤ	3	404	1454	2194	ɤ	3	**398**	1748	2716
ɯ	4	283	1444	2251	ɯ	3	**399**	1595	2459
ɯ̃	1	317	1448	2303	ɯ̃	1	**303**	1351	2220
ɔ	2	621	970	2315	ɔ	2	**613**	1023	2669
o	4	383	801	2488	o	3	421	853	2921
õ	3	828	1138	2204	õ	3	789	1226	2502
u	2	264	750	2384	u	2	354	762	2364
ũ	2	309	692	2287	ũ	2	275	656	2382

A Preliminary Report on the Phonetics of Tlingit

Ian Maddieson, Nicola Bessell*, and Caroline Laws Smith**

Introduction

The language usually known as Tlingit is the indigenous language of the people who lived on most of the islands of the Alexander archipelago and in adjoining parts of the North American mainland before European incursions into the area. A map of the historically Tlingit area is shown in Figure 1. This part of North America became the center of operations of the Russian-American Company in the late 18th century and now forms the south-eastern part of the US state of Alaska, a portion of the nearby Canadian territory of the Yukon, and a part of the Canadian province of British Columbia.

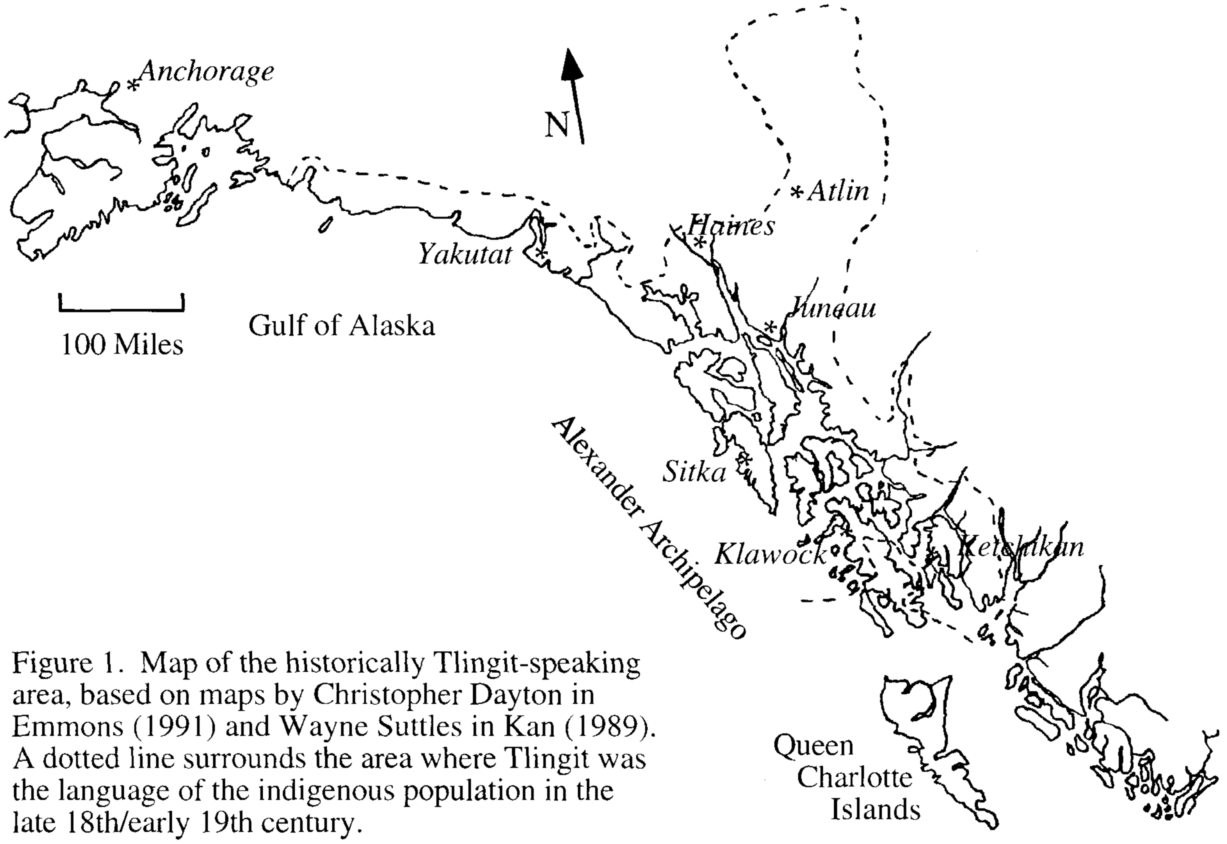


Figure 1. Map of the historically Tlingit-speaking area, based on maps by Christopher Dayton in Emmons (1991) and Wayne Suttles in Kan (1989). A dotted line surrounds the area where Tlingit was the language of the indigenous population in the late 18th/early 19th century.

The spelling Tlingit represents a semi-Anglicized version of the word /ɬinkít/ which means 'person' in this language. A more ruthlessly Anglicized pronunciation, [klɪŋkɪt], is commonly used in Alaska, including by many members of the ethnic group concerned. The consensus among linguists is that Tlingit is related to Eyak and the Athabaskan languages in a family usually called Na-Dene (following Sapir 1929, see Krauss 1973, 1979). Na-Dene may also include Tlingit's immediate southern neighbor, Haida.

---

*Nicola Bessell is at the Department of Linguistics, University of Texas, Austin.
**Caroline Smith is at the Department of Linguistics, University of Ottawa.

It is difficult to estimate the current number of speakers of Tlingit. Krauss's estimate in the mid-1970's was that there were "perhaps 2000 at most ... and most must be over forty" (Krauss 1979: 842). This estimate appears to reflect survey data collected even earlier by the Summer Institute of Linguistics. Time has taken its toll and there now appear to be few speakers under 60, and even many of the older speakers only use the language rather occasionally. Thus, few, if any, young children currently have the opportunity to learn the language in a home setting. Dauenhauer and Dauenhauer (1995: 101) comment that "close to 90% of the ethnically Tlingit population is now monolingual English-speaking" and they are pessimistic about the survival of the language.

There are, nonetheless, efforts under way to teach the language to younger members of the community through formal educational processes. Both a 'Beginning Tlingit' course-book for teachers and adult learners, with accompanying tapes, and a learner's guide to the spelling system are available (Dauenhauer and Dauenhauer 1991, 1984). A standardized orthography was developed in the early 1960's and used in Biblical translation as as well as in dictionaries of Tlingit nouns (Naish and Story 1963) and verbs (Naish and Story 1973). A slight modification of this orthography was agreed in the early 1970's and this has been used in a revised edition of the noun dictionary (Davis 1976, reprinted 1996) as well as in publications concerning Tlingit history and reminiscence, orations, and cultural traditions (Dauenhauer & Dauenhauer 1987, 1990, 1994) and in the teaching materials mentioned above. An orthography different from the standard and employing many diacritics has been used in Nyman and Leer (1993).

Technical linguistic study of Tlingit began with Swanton (1908, 1909) and took a significant step forward with Boas (1917). In more recent years, besides the works cited in the previous paragraph, important contributions have been made by Krauss (see Krauss 1973, 1979 for a summary), Leer (1978, 1990, 1991, and elsewhere), and Pinnow (1966). None of the previous studies have, however, paid a great deal of attention to the phonetics of the language. The most significant discussion so far of phonetic details of Tlingit is to be found in Bessell (1996). Bessell's observations prompted the present more intensive investigation.

This preliminary report is the result of a brief visit to Juneau in August 1996 to collect acoustic and articulatory data. The language reflected is that typical of the Central Tlingit area including Juneau, Sitka, Hoonah and Angoon. In all the descriptions below 'Tlingit' should be understood to refer to this variety of the language unless specifically noted. Some differences have been reported for the Inland or Yukon variety (Nyman and Leer 1993). Speakers we consulted also considered that the southern variety of the language, as spoken around Ketchikan and on Prince of Wales island (Klawock and Craig), is recognizably distinct, but we are not sure of the nature of these differences, apart from the possible survival of the /ɣ/ and /ɣʷ/ noted by Swanton and Boas and a tonal difference reported by Leer (1978). Preservation of /ɣ/ is also noted in de Laguna (1972) for the Northern Tlingit of Yakutat. The more radically different Tongass dialect (Leer 1978), in the extreme south, may now have no remaining fluent speakers.

Inventory

a) Consonants

Central Tlingit has a relatively rich inventory of consonants. The inventory of centrally articulated consonants is given in Table 1. (For convenience, the labial-velar approximant /w/ is listed in the column for labialized velars, not given a separate column.) Lateral segments are listed separately in Table 2. The transcriptions used follow the conventions of the IPA.

Table 1. Central consonants of Tlingit.

	Dental/ Alveolar	Post-alveolar	Palatal	Velar	Labialized Velar	Uvular	Labialized Uvular	Glottal
Unaspirated plosive	t			k	k^{w}	q	q^{w}	?
Aspirated plosive	t^{h}			k^{h}	k^{wh}	q^{h}	q^{wh}	
Ejective stop	t’			k’	k^{w}’	q’	q^{w}’	
Unaspirated affricate	ts	tʃ						
Aspirated affricate	ts^{h}	tʃ^{h}						
Ejective affricate	ts’	tʃ’						
Fricative	s	ʃ		x	x^{w}	χ	χ^{w}	h
Ejective fricative	s’			x’	x^{w}’	χ’	χ^{w}’	
Nasal	n							
Approximant			j		w			

Table 2. Lateral consonants in Tlingit.

Unaspirated lateral affricate	tɬ
Aspirated lateral affricate	tɬ^{h}
Ejective lateral affricate	tɬ’
Lateral fricative	ɬ
Ejective lateral fricative	ɬ’

This consonant inventory has several unusual properties. It has a large number of back stops and fricatives, and a larger than usual number of lateral consonants. It includes the very unusual category of ejective fricatives, which will be discussed at some length in a later section. It is also characterized by some striking absences. In common with many of the indigenous languages of the northwest quadrant of North America, Central Tlingit has no bilabial or labio-dental consonants in native vocabulary (a few of the more recent loanwords from English maintain bilabials). It has very few voiced (or sonorant) consonants; just a single nasal and the two approximants /j, w/. Particularly striking is the absence of a central lateral approximant, given the large number of obstruent lateral consonants. Although more recent loanwords from English may contain /l/, Tlingit even resisted incorporating any /l/’s in earlier loanwords. Hence English *gold* was borrowed as /kúun/ and *dollar* as /táanaa/ ‘money’. Nouns illustrating each of the consonants of the variety of Tlingit under discussion are given in Table 3 together with the established orthography. Initial and final occurrences are given in separate columns to draw attention to some important distributional restrictions. These will be discussed in a little more detail in a later section. The words in Table 3 are almost all monosyllables, but this should not be taken to indicate that the language is characteristically monosyllabic in its word structure. Many nouns have stems longer than one syllable, and in many contexts nouns occur with affixes added. Verbal forms are typically polysyllabic with several categories of affixes added to a stem.

Table 3. Words illustrating consonants of Tlingit.

	Initial			**Final**		
	IPA	**Orthog.**	**Gloss**	**IPA**	**Orthog.**	**Gloss**
t	táa	*dáa*	*weasel*	ʔaat	*aat*	*father's sister*
tʰ	tʰaan	*taan*	*sealion*	-	-	(does not occur)
t’	t’áa	*t’áa*	*board*	χw’áat’	*x̱’wáat’*	*Dolly Vardon (fish)*
ts	tsaas	*dzaas*	*thong*	kʰats	*kats*	*lime, pounded shell*
tsʰ	tsʰaa	*tsaa*	*harbor seal*	-	-	(does not occur)
ts’	ts’axʷéeɬ	*ts’axwéil*	*crow*	xaats’	*xaats’*	*distant blue sky*
s	saak	*saak*	*eulachon (fish)*	ʔaas	*aas*	*tree*
s’	s’aaw	*s’aaw*	*crab*	haas’	*haas’*	*vomit*
tʃ	tʃaɬ	*jal*	*buried storehouse*	qáatʃ	*g̱áach*	*rug*
tʃʰ	tʃʰaan	*chaan*	*mother-in-law*	-	-	(does not occur)
tʃ’	tʃ’áak’	*ch’áak’*	*bald eagle*	ʃáatʃ’	*sháach’*	*smelt, needlefish*
ʃ	ʃáa	*sháa*	*women*	tʃáaʃ	*jáash*	*branches used in underground store*
k	kaaw	*gaaw*	*drum, bell, hour*	ɬ’aak	*l’aak*	*dress*
kʰ	kʰaa	*kaa*	*yardstick*	-	-	(does not occur)
k’	k’aakán	*k’aagán*	*pinfish*	kʰaak’	*kaak’*	*forehead*
kʷ	kʷéeɬ	*gwéil*	*sack, bag*	naakʷ	*naakw*	*medicine*
kʷʰ	kʷʰáataa	*kwáadaa*	*quarter (coin)*	-	-	(does not occur)
kʷ’	kʷ’át’	*k’wát’*	*bird's egg*	ʔaakʷ’	*aak’w*	*small lake*
q	qákʷ	*g̱ákw*	*tree spine*	tɬ’áaq	*tl’áaḵ*	*arrowhead*
qʰ	qʰákʷ	*ḵákw*	*basket*	-	-	(does not occur)
q’	q’akʷ	*ḵ’akw*	*screech owl*	-	-	(does not occur)
qʷ	-	-	(does not occur)	náaqʷ	*náaḵw*	*octopus*
qʷʰ	qʷʰáan	*ḵwáan*	*people, tribe*	-	-	(does not occur)
qʷ’	qʷ’átɬ	*ḵ’wátl*	*cooking pot*	qeeqʷ’	*ḵeiḵ’w*	*kittiwake ('sea pigeon')*
x	xaak	*xaak*	*empty seashell*	s’áx	*s’áx*	*starfish*
x’	x’áa	*x’áa*	*point of land*	s’aax’	*s’aax’*	*grey cod*
xʷ	xʷátʃaa	*xwájaa*	*skin scraper*	káaxʷ	*gáaxw*	*duck*
xʷ’	xʷ’án	*x’wán*	*boots*	qáaxʷ’	*ḵáax’w*	*herring spawn*
χ	χaakʷ	*x̱aakw*	*fingernail*	s’áaχ	*s’áax̱*	*groundhog*
χ’	χ’áakʷ	*x̱’áakw*	*freshwater sockeye salmon*	tʃʰeeχ’	*cheix̱’*	*thimbleberry*
χʷ	χʷastáa	*x̱wastáa*	*canvas, denim*	húnχʷ	*húnx̱w*	*older brother*
χʷ’	χw’áaɬ’	*x̱’wáal’*	*down (feathers)*	ɬéeχʷ’	*léix̱’w*	*red face paint*
ʔ	ʔáa	*áa*	*lake*	-	-	(does not occur)
h	háat	*háat*	*riptide*	-	-	(does not occur)

Table 3, continued.

tɬ	tɬeet	*dleit*	*snow, white*	tʃaatɬ	*chaatl*	*halibut*
tɬʰ	tɬʰáa	*tláa*	*mother*	tɬʰ	-	(does not occur)
tɬ'	tɬ'eeq	*tl'eik̲*	*finger*	tɬ'áatɬ'	*tl'áatl'*	*yellow*
ɬ	ɬáx'	*láx'*	*heron, crane*	ʃáaɬ	*sháal*	*fish trap*
ɬ'	ɬ'áa	*l'áa*	*breast*	káaɬ'	*gáal'*	*clams*
n	naa	*naa*	*tribe*	ʔaan	*aan*	*town*
j	jaa	*yaa*	*sea trout*	taaj	*taay*	*hot springs*
w	waaq	*waak̲*	*eye*	jaaw	*yaaw*	*herring*

b) Vowels and Tones

Tlingit has a system of four distinctive vowel qualities, /i, e, a, u/. All four occur both long and short, with the short vowels tending to be pronounced more centralized than the long vowels. Examples of the vowel contrasts are given in Table 4. Long vowels, here as well as throughout this article, are written with a doubled vowel letter.

Table 4. Words illustrating Tlingit vowels.

	Short			**Long**		
i	hít	*hít*	*house*	kʰiit	*keet*	*killer whale*
e	tʰé	*té*	*stone*	tee	*dei*	*road*
a	t'á	*t'á*	*king salmon*	t'áa	*t'áa*	*board*
u	túʃ	*dúsh*	*tadpole*	túuʃ	*dóosh*	*cat*

Tlingit has a contrast between high and low tones. Examples of minimal or near-minimal lexical contrasts among nouns are given in Table 5. In this table, as well as throughout this article, high tone is marked with an acute accent and low tone is unmarked. These examples also demonstrate that the tonal distinction is independent of vowel length on the surface. It should be noted, however, that virtually all monosyllabic nouns with short vowels have high tone.

Table 5. Words illustrating Tlingit tonal contrasts.

High			**Low**		
ʃáa	*sháa*	*women*	ʃaa	*shaa*	*mountain*
kúun	*gúun*	*gold*	kuun	*guun*	*spring (water)*
íiq	*éek̲*	*beach*	iiq	*eek̲*	*copper*
náaqʷ	*náak̲w*	*octopus ('devilfish')*	naaqʷ	*naak̲w*	*dried wood used for tanning*
ɬáaχ	*láax̲*	*standing dead tree*	ɬàaχ	*làax̲*	*red cedar*
x'áan	*x'áan*	*anger*	χ'aan	*x̲'aan*	*fire, red*
kán	*gán*	*firewood*	kʰanéest	*kanéist*	*cross*
kʰeetʃín	*keijín*	*five*	tʃinkʰaat	*chinkaat*	*ten*

Apart from a small number of mostly grammatical function words (like deictics and question markers) no more than one syllable of a word with two or more syllables will normally have a high tone, and this will often be the last syllable of the word. In certain types of noun phrases and compounds, suppression of a non-final high tone occurs, which may serve as one signal of the linking of the elements involved. Thus /sán/ 'paternal uncle' gives the possessed form /aχ sánii/ 'my paternal uncle', but when the plural marker /hás/ is added only its high tone appears in the phrase: /aχ sanii hás/ 'my paternal uncles'. In this respect the tones of Tlingit are acting very much as an accentual system.

Tone is very important in the functioning of the complex verbal morphology but verbs do not have lexical tone contrasts in the same way as nouns. Depending on the morphological categories reflected, most verb stems in Tlingit may occur with long or short vowel and high or low tone. Quite long verbal forms often contain no high tone syllable at all, e.g. /ajaχtatʰiit/ 'would blow over it' (Dauenhauer and Dauenhauer 1990: 502). Tone in Tlingit thus functions in at least three ways, as the basis of lexical contrast among nouns, as a grammatical marker of categories within the verbal system, and as an indicator of phrasal constituency.

No general study has yet been conducted of tonal processes in Tlingit, but it is clear that there are several interesting features. As already noted by Boas (1917: 11) a number of affixes have a tone assigned to them based on a polarity principle; their tone is the opposite of the (final) tone of the stem. For example, the suffix /-ii/ which marks possession is high tone in /tu χaatíi/ 'his root', but low in /tu qáatii/ 'his salmon'. (This suffix is variable in other ways as well, having the vowel /u/ after a stem final /u/ or /w/ or labialized consonant, and in being variably pronounced long or short. It also has an approximant onset after stem-final vowels; /w/ after /u/ and /j/ elsewhere. Hence /tu t'áajii/ 'his hot springs' with stem /t'áaj/ and /tu t'áajii/ 'his board' with stem /t'áa/ are homophonous. Boas traces this latter behavior to the suffix having etymologically an initial /ɣ/.). Another suffix with polar tone is the directional /-tee/ 'to'.

Boas published the first instrumental phonetic data on Tlingit, several kymograms designed to study the tonal contrast. He reported that the F0 levels of low and high tones are quite close, having a ratio of about 14 to 15. Our informal observations support the suggestion that the tones are not widely separated in pitch.

Phonotactics

A number of the consonants have restricted distributions, as noted in Table 1 above. Labialized velars and uvulars are more common in word-final than in word-initial position and some do not occur at all in initial position. The absence of /q'/ from final position may be an accidental gap in our data, as other ejectives are relatively common in this position. Whereas there is a contrast between unaspirated and aspirated pulmonic stops in word- (and syllable-) initial position, only one series of pulmonic stops occurs in final position.

The word-final stops which occur in utterance-final position are usually released quite audibly with some sustained noise following the release burst. This fact has led to their being interpreted as aspirated, e.g. by Leer (1978) and implicitly in the Tlingit orthography. However, the amplitude of the noise following these utterance-final releases is typically markedly less than that which occurs in initial aspirated stops, as is illustrated by the spectrogram in Figure 2. The release burst of initial /tʰ/ is followed by an interval filled by noise with a very broad spectrum, whereas the release burst of final /t/ is followed by noise concentrated in formant-like bands. The final noise is often shorter and weaker than that observable in Figure 2, as for example in the final /k/ in Figure 6. When word-final stops are not in utterance-final position, there is no sustained noise after the release burst. In particular, when vowel-initial affixes are attached to words ending with /t, k, q/ there is clearly no aspiration whatsoever present. (Boas noted this but described it as a process of 'replacement of surds with sonants', anticipating the alternations in the modern orthography between ‘d’ and ‘t’, ‘k’ and ‘g’ etc., as in *hít ‘house’* but *hídi ‘his/her house’*). Finally, we note that utterance-final consonants of other classes are also frequently followed by a marked audible release and following noise, as for example the final /n/ of /χ’aan/ ‘fire’ in Figure 8. It seems clear that the noisy audible release for utterance-final stops is due to their position and not to their belonging to the aspirated category. On both phonological and phonetic grounds the word-final stops are instances of the unaspirated category. (The pattern for affricates is similar.)

The absence of aspirated stops and affricates in word-final position (and in any syllable-final position) is matched by the absence of /h/ from this same position . Absence of final aspiration is a quite common restriction among languages which have an aspiration contrast in their stops. Some quantification of the aspiration contrast between initial stops in Tlingit is given in the next major section below.

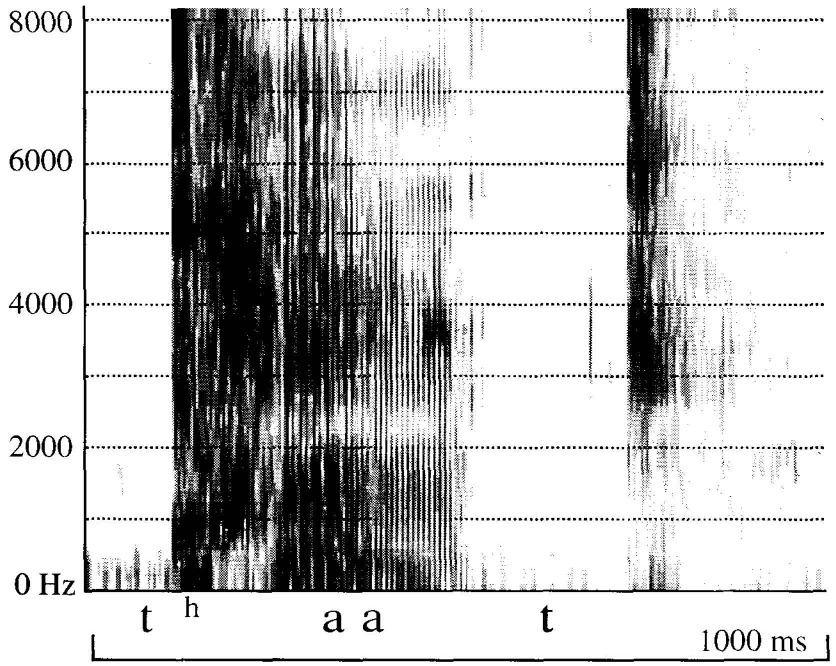


Figure 2. Spectrogram of the word /t^{h}aat/ ‘night’ (speaker 2, female), illustrating the difference between aspiration of word-initial /t^{h}/ and release of final /t/.

In utterance-final position the realization of the labialization of the labialized velar and uvular stops and fricatives includes a quite lengthy voiceless sonorant portion, sounding like a voiceless back rounded vowel. The percept can be of a separate syllable. We note also in this section that /j, w/ occur in both onset and coda position of syllables, but whereas /w/ can follow any of the four vowel qualities, /j/ does not follow /i/ and is only common after low vowels.

/ʔ/ occurs only in morpheme-initial position. It might be regarded as inserted before morpheme-initial vowels, but since there are also affixes which begin with a vowel not preceded by a glottal stop, its distribution cannot be predicted simply from knowledge of where morpheme boundaries lie. Compound words result in the occurrence of clusters such as /-tʔ-/, /-kʔ-/, /-sʔ-/. These are distinct from the ejectives /t’, k’, s’/, etc. Examples include /ɬitʔistúk/ ‘black bass’, /kukʔát/ ‘earring’, and /q^{h}usʔuuq’/ ‘plaything, toy’.

Consonant measurements

Unaspirated, Aspirated and Ejective stops.

As noted above, Tlingit has three categories of laryngeal action with stops (and also with affricates). The unaspirated stops are released with only a very short or zero delay before the vowel begins. The aspirated and the ejective stops are both released well before the vowel, but whereas the interval from release to vowel onset is filled with noise from relatively high-volume air flow in the aspirated stops, the ejective stops have a period of (near-)silence after the release. The glottal closure which accompanies the oral articulation is usually held until the vowel onset; an audible glottal onset to the vowel can be heard and no earlier acoustic transient indicating the glottal release can be seen.

In some languages, e.g. Hausa, the oral and glottal closures in an ejective stop are released very close together in time, and the interval before vowel onset is filled with low-amplitude noise, but in this respect Tlingit is similar to Navajo rather than to Hausa (Lindau 1984, Ladefoged and Maddieson 1996). Spectrograms of representative tokens of word-initial /kʰ/ and /k'/ are shown in Figures 3 and 4, and also in Figure 10. Figures 3 and 4 also exemplify the two-way word-final contrast between ejective and pulmonic stops. At their release the ejectives have a shorter and higher-amplitude noise than the pulmonic stops in utterance-final position, and frequently a separate release of the glottal closure can be heard and detected on spectrograms some 50-100 ms after the oral release. In Figure 3 this glottal release can be seen above the figure '6' in the legend.

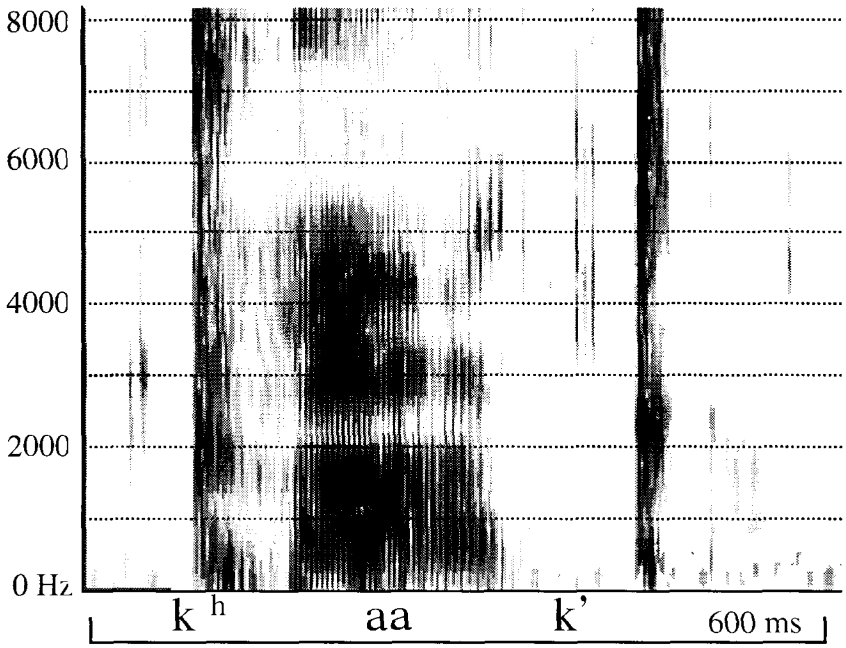


Figure 3. Spectrogram of the word /kʰáak'/ 'forehead' (speaker 5, female), illustrating initial /kʰ/ and final /k'/.

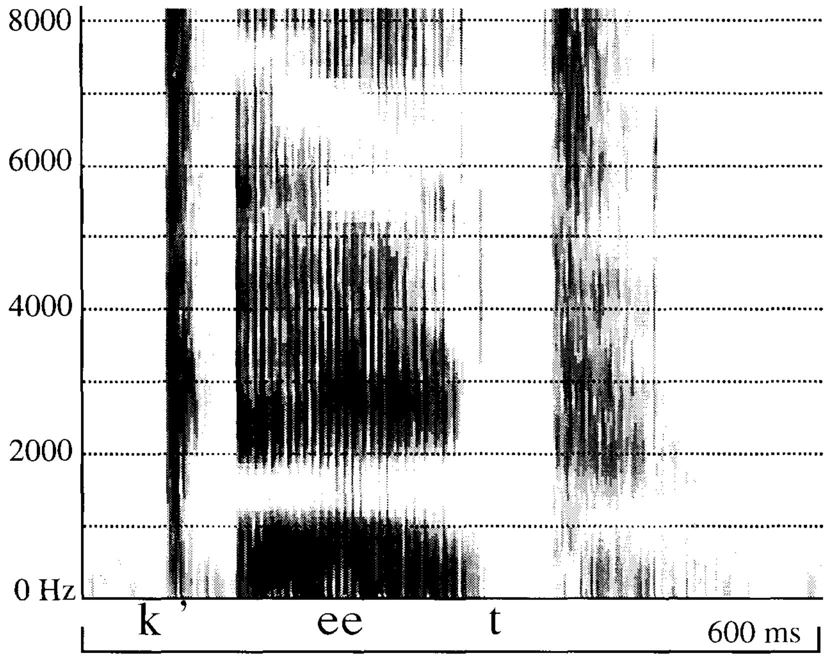


Figure 4. Spectrogram of the word /k'eet/ 'salmonberry bush sprouts' (speaker 5, female), illustrating initial /k'/ and release of final /t/.

In order to provide a basis for comparing Tlingit with other languages, measurements were made of the voice onset time for the stops /t, k, q; t^h, k^h, q^h; t', k', q'/. Monosyllabic words beginning with these sounds followed by a long /aa/ (with either high or low tone) were digitized from the audio recordings at 20 kHz and measurements were made from simultaneous displays of waveforms and wide-band spectrograms. (In a few cases a word of another pattern had to be measured; for example, the word /k'eet/ 'salmonberry bush sprouts' was consistently used for examples of /k'/.) The word-list used for these measurements is given in Table 6. Data from four speakers, 3 female and 1 male has been measured. Generally two tokens of each word were measured for each speaker, but a third token was measured when it was available.

Table 6. Word list for measurements of stop durations.

	First Choice			Alternate		
t	táa	*dáa*	*weasel*	táanaa	*dáanaa*	*money* (1 speaker)
t^h	t^haan	*taan*	*sealion*			
t'	t'aaw	*t'aaw*	*feather*			
k	káał	*gáal*	*clams*	káas'	*gáas'*	*house post* (1 speaker)
k^h	k^háax'	*káax'*	*chicken, spruce grouse*	k^háak'	*káak'*	*forehead* (1 speaker)
k'	k'éet	*k'éit*	*salmonberry bush sprouts*			
q	qáax^{w}'	*g̱áax'w*	*herring spawn*			
q^h	q^háa	*ḵáa*	*person*			
q'	q'áatʃ'	*ḵ'áach'*	*ribbon seaweed*			

The mean duration of voice onset time was 24.6 ms (s.d. 23.4) for unaspirated stops, 127.6 ms (s.d. 36.8) for aspirated stops, 102.7 (s.d. 40.6) for ejective stops. The means are for 25 tokens of each category. Analysis of variance showed a strong main effect of laryngeal setting on voice onset time, F (2, 72) = 61.139, p < .0001. Post-hoc comparison of means shows that this is contributed mainly by the shortness of the unaspirated category, the comparison of the unaspirated category with both of the other classes being significant at better than the .0001 level. The ejective stops have a shorter VOT than the aspirated ones, but the difference is only marginally significant.

Unlike in many other languages (Maddieson 1996), no significant difference in VOT is found between stops at different places of articulation in Tlingit (F (2, 72) = .185, p = .83). The overall mean durations for the three places represented here are 80.5 ms for alveolars, 84 for velars, and 90 for uvulars. The means for each combination of laryngeal setting and place are shown graphically in Figure 5. Each bar represents the mean of 8 or 9 measurements.

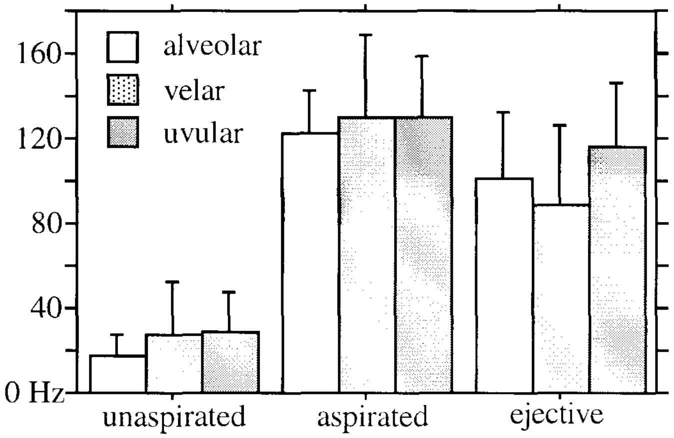


Figure 5. Means of VOT for stops separately by place and laryngeal setting.

Ejective fricatives.

Ejective fricatives occur in only a very small proportion of the world's languages (Maddieson 1984). Because of their relative rarity, these segments were a special focus of our investigations. Moreover, there has been some question as to whether the description of them as ejective was in fact correct. In an ejective stop the increased air pressure behind the articulatory constriction that is required to generate an explosive release is created by raising the larynx, using the closed glottis as a piston. A relatively small volume of air — that enclosed between the closed vocal folds and the oral articulatory constriction — is involved, but the volume change in the oral cavity is proportionally very large, so that sufficiently high pressures can easily be generated. However, a continuing flow of air is required to generate frication noise. Since the volume of air available is small and the mobility of the larynx is limited, the ejective mechanism (or glottalic egressive airstream) can only generate a short duration of continuing flow. It follows that ejective fricatives are expected to have a rather short noise duration. Bessell (1996) suggested that the durations observed in Tlingit were too long for these fricatives to be plausibly ejective. Instead, she proposed that the frication in these segments must be produced with pulmonic air flow with their distinctive sound created by producing them with a preceding or following glottal constriction. They would therefore be better described as 'glottalized' or 'laryngealized' fricatives.

We will describe some of the properties of these fricative sounds based on the acoustic and aerodynamic data we have collected, and will then return to consideration of how they are produced. We will continue to refer to the ejective fricatives by that term, as we feel that the evidence will show that this is appropriate for the speakers studied here. Some qualitative illustration of the acoustic patterns of the ejective fricatives and their pulmonic counterparts is provided by the series of spectrograms in Figures 6-10 below.

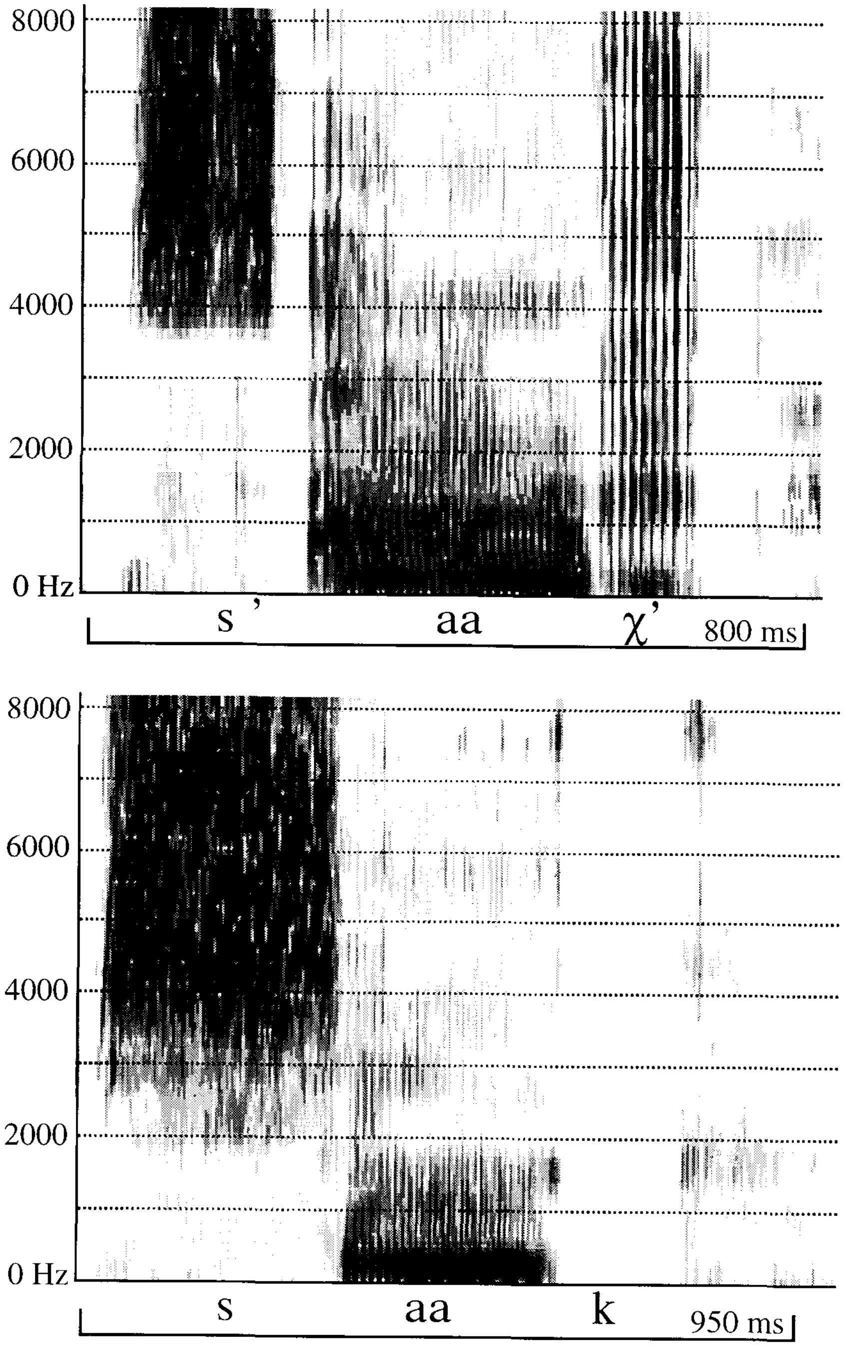


Figure 6. Spectrograms of the words /s'aaχ'/ 'gray cod' and /saak/ 'eulachon'. (Speaker 1, Female).

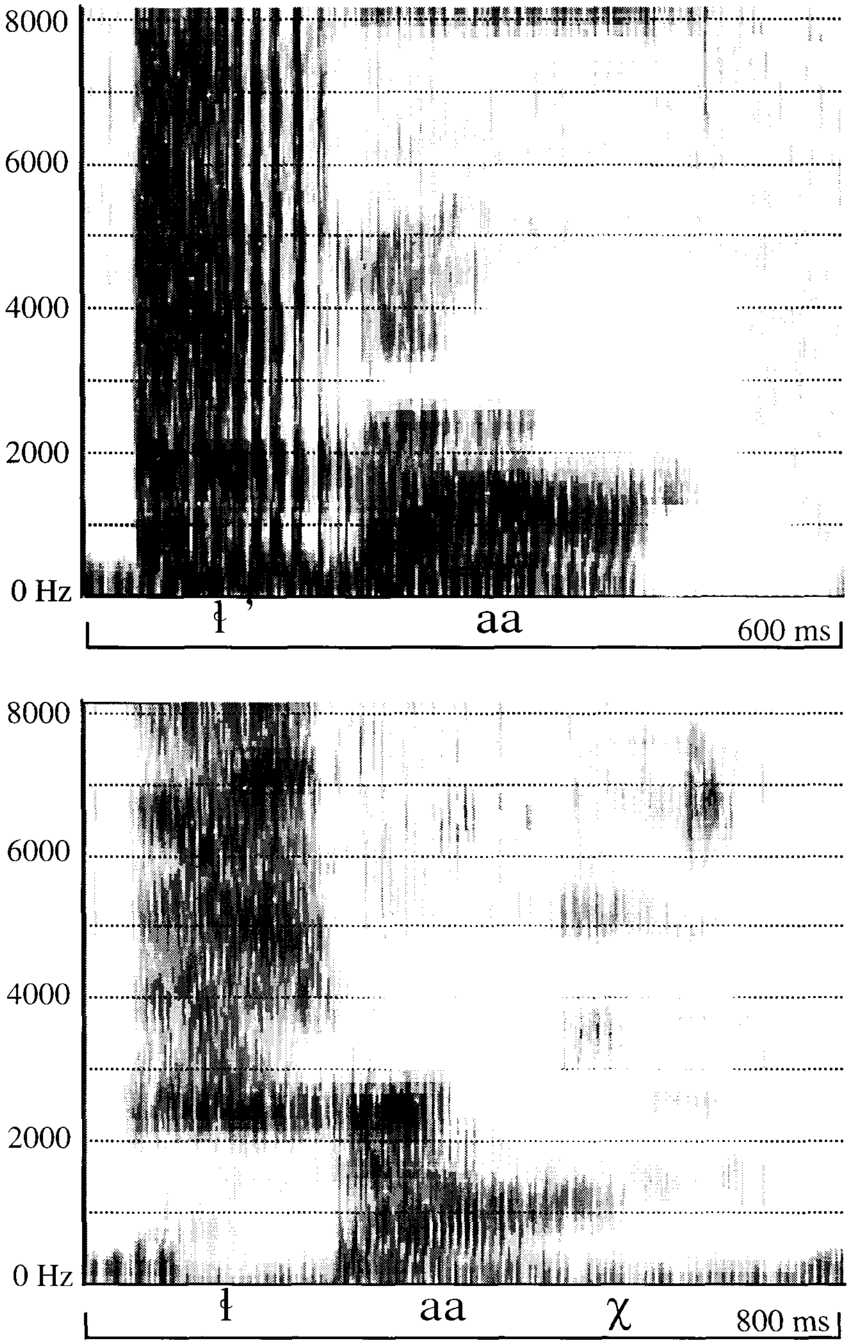


Figure 7. Spectrograms of the words /ɬ'aa/ 'breast' and /ɬaaχ/ 'red cedar'. (Speaker 6, Male).

In general, the noisy portion of an ejective fricative is separated from a preceding or following vowel by a period of (near-)silence, whereas the pulmonic fricatives directly abut the vowel. The separation of an initial ejective fricative's noise from the following vowel is well illustrated in Figures 6, 8 and 9. The onset to a following vowel clearly involves the release of a glottal closure; a very clear instance can be seen in Figure 6 where the first few periods of the vowel have particularly high amplitude as well as a low fundamental frequency. The glottal onset is also auditorily apparent when the frication is removed from the signal and the edited signal played. Figure 7 illustrates a less typical pattern in which there is no silent interval between the noisy portion of the fricative and the following vowel. The typical pattern with final ejective fricatives is shown in Figure 10. A less typical pattern, with virtually no silent interval before a final ejective fricative, occurs in the word /s'aaχ'/ 'gray cod' in Figure 6. In this token the release of a glottal closure can be seen on the spectrogram to occur about 80 ms after the end of the frication noise. Such a glottal release is often audible after the end of frication in final ejective fricatives, and detectable on spectrograms.

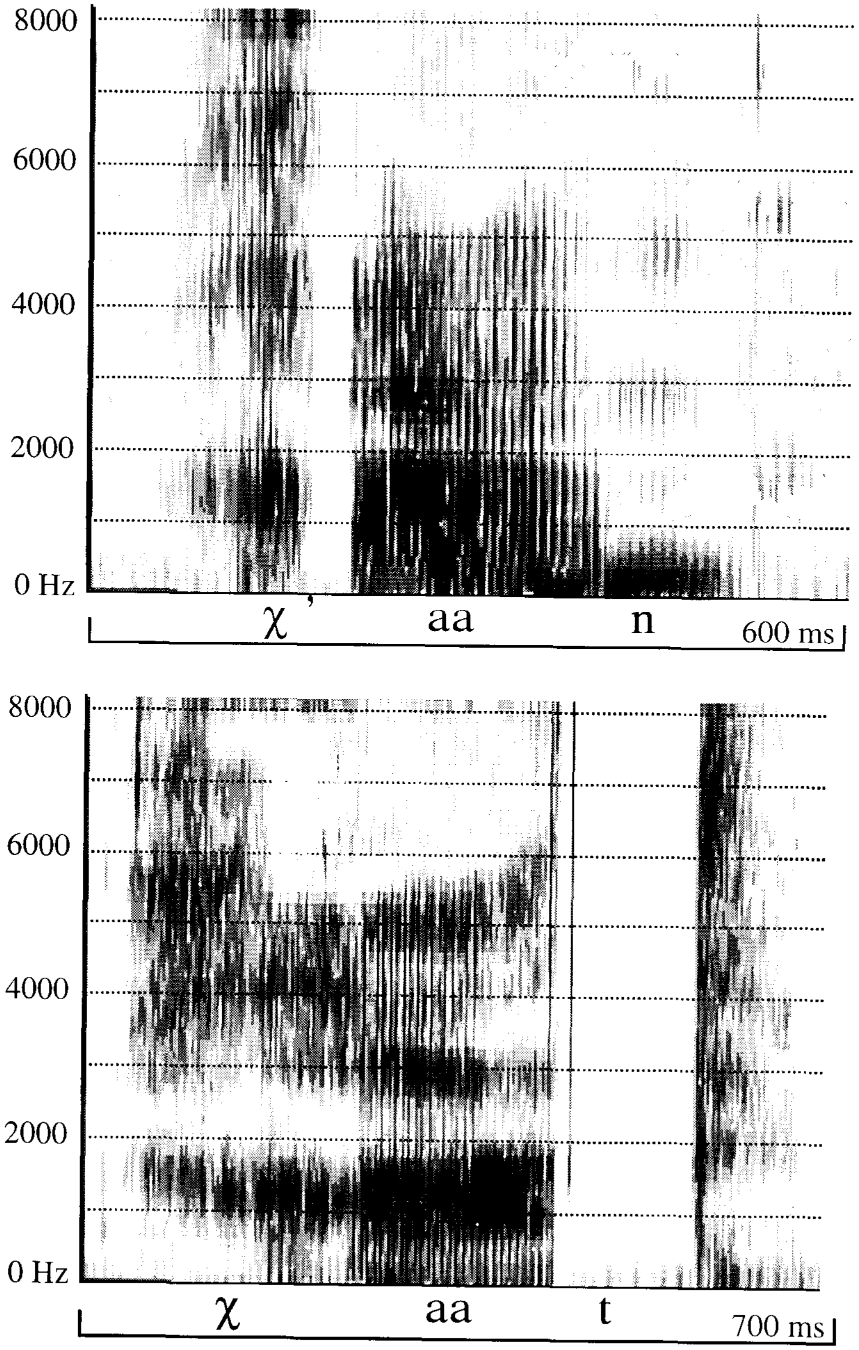


Figure 8. Spectrograms of the words /χ'aan/ 'fire' and /χaat/ 'spruce roots'. (Speaker 5, Female).

The noisy portion of an ejective fricative is also typically shorter than that of a pulmonic counterpart. This relationship is very evident in Figures 6, 8 and 9. A less dramatic difference is apparent in Figure 7. Not only is the duration of the noise of an ejective fricative different from that of a pulmonic fricative, but the noise is often qualitatively different. Ejective fricatives often have a more ‘scrapy’ quality, with the noise broken into a series of pulses. Such pulsing can be clearly seen in the lateral /ɬ’/ in Figure 7, and in the initial /χ’/ in Figure 9 as well as the final /χ’/ in Figure 6. Although pulsing is sometimes observed with pulmonic /χ/, as in Figure 9, it is more frequent and stronger with /χ’/. We have not observed pulsing during the frication in /s’/.

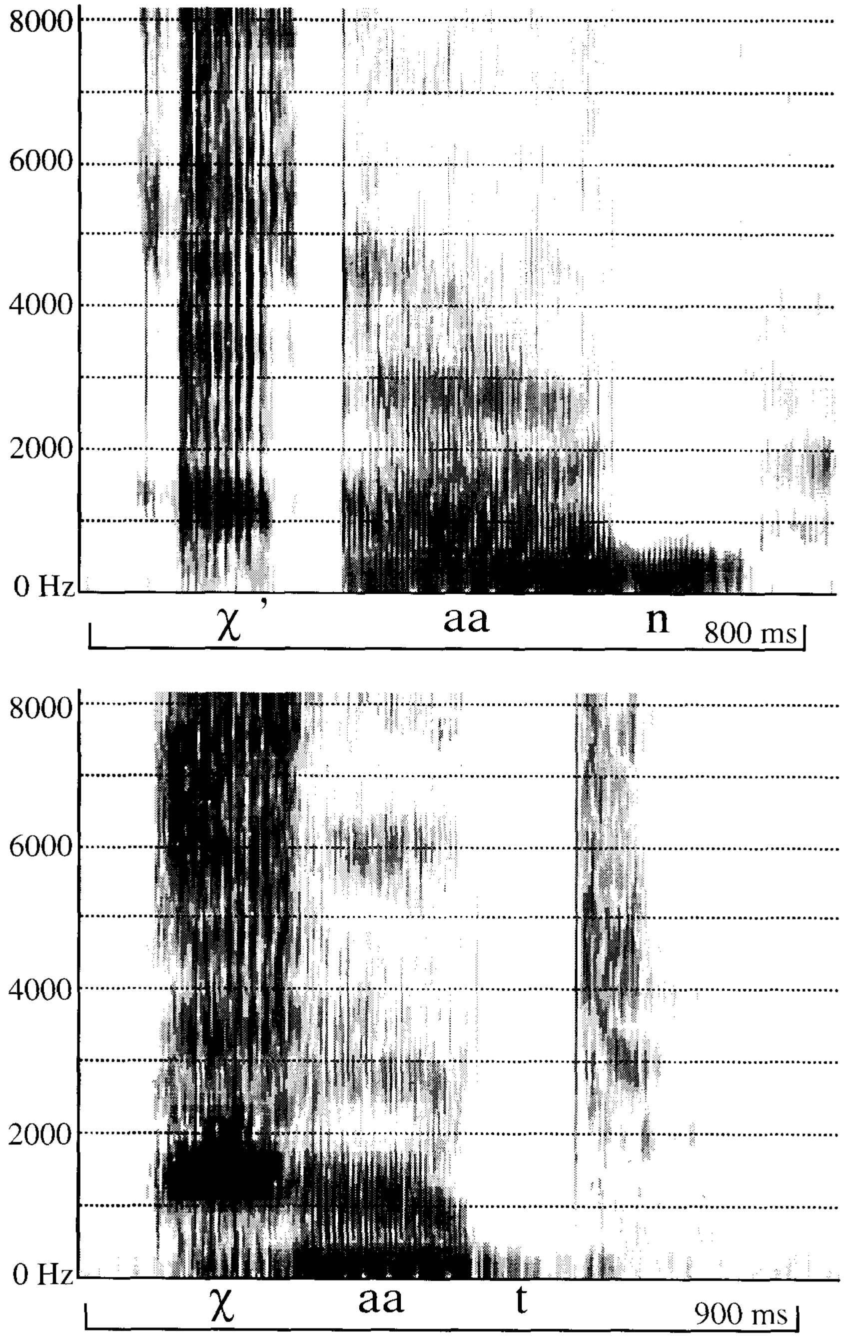


Figure 9. Spectrograms of the words /χ’aan/ ‘fire’ and /χaat/ ‘spruce roots’. (Speaker 1, Female).

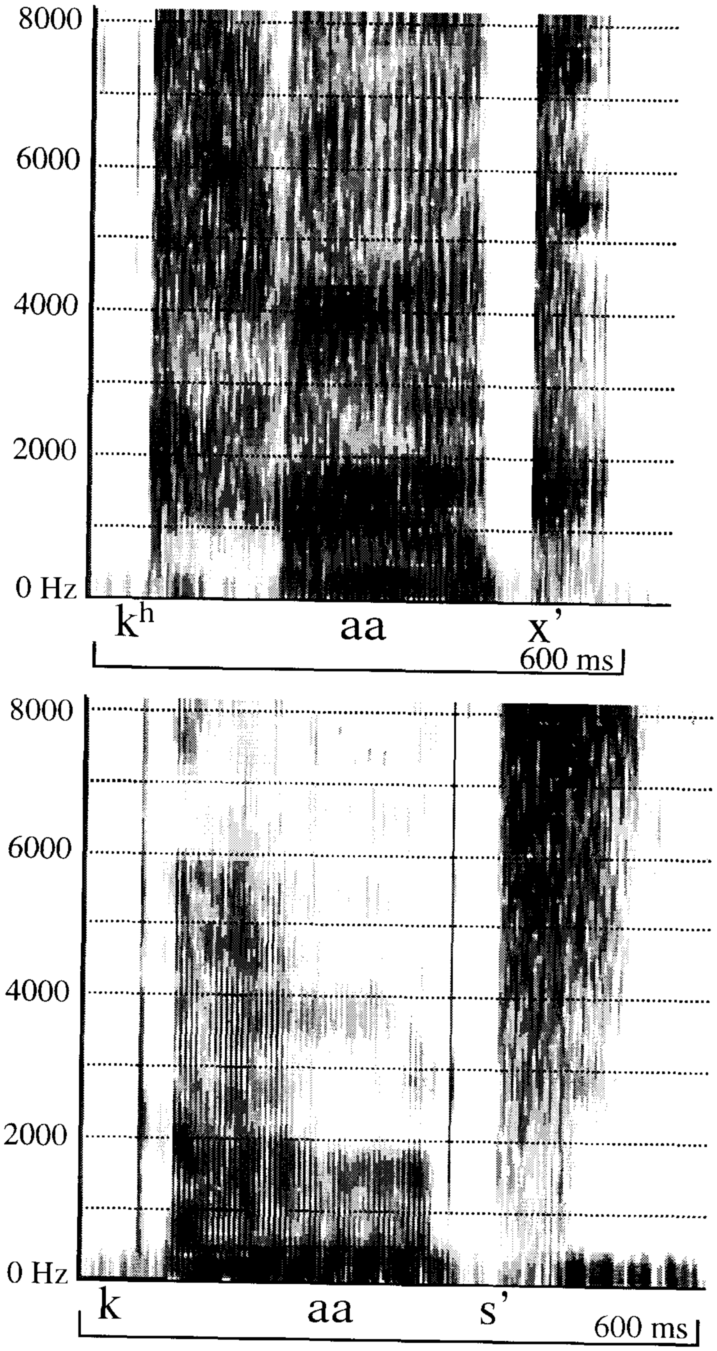


Figure 10. Spectrograms illustrating final velar and alveolar ejective fricatives in the words /k^{h}aax'/ 'chicken, grouse' and /kaas'/ 'house post'. (Speaker 1, Female).

A similar procedure to that described above for measuring duration in stops was followed in order to obtain some quantitative data on word-initial fricative durations. The word-list used is shown in Table 7. Normally two tokens of each word from each of four speakers was measured; a third token was sometimes measured when it was available.

Table 7. Word list for measurements of fricative durations.

	First Choice			Alternate		
ɬ	ɬaaχ	*laax̲*	*red cedar*	ɬáaχ	*láax̲*	*dead tree* (1 speaker)
ɬ'	ɬ'aak	*l'aak*	*dress*	ɬ'aa	*l'aa*	*breast* (1 speaker)
s	saak	*saak*	*eulachon (candlefish)*			
s'	s'áaχ'	*s'áax̲'*	*grey cod*	s'áaw	*s'áaw*	*crab* (1 speaker)
x	xaak	*xaak*	*skeleton, seashell*	xáats	*xáats*	*blue sky* (1 speaker)
				xát'aa	*xát'aa*	*sled* (1 speaker)
				xít'aa	*xít'aa*	*broom* (1 speaker)
x'	x'áax'	*x'áax'*	*(crab-)apple*			
χ	χaat	*x̲aat*	*spruce roots*			
χ'	χ'aan	*x̲'aan*	*fire*			

Our results show that in the set of words examined the interval from the onset of frication to the onset of the voiced vowel is quite similar — about 200 ms — for plain and ejective fricatives, the mean for this interval being 223 ms with plain fricatives (n = 32) and 194 ms for ejectives (n = 34). The difference is marginally significant in a one-way analysis of variance (F (1, 64) = 5.499, p = .0221). The means for different places of articulation, counting 'lateral' as one of the places, were not significantly different, though the sibilants tend to have a longer duration of this interval than the velar, uvular or lateral fricatives.

By contrast, the duration of the frication itself (i.e. subtracting the duration of any 'silent interval' before the vowel from the above measure) is significantly shorter in ejective fricatives than in pulmonic fricatives. The mean over all ejective fricatives is 148 ms and for pulmonic fricatives is 222 ms. This difference is highly significant (F (1, 64) = 41.84, p = .0001). There is no significant difference between the fricatives in their friction duration depending on their place of articulation, nor any significant interaction between laryngeal type and place. The mean friction durations for each of the individual fricatives is shown in Figure 11.

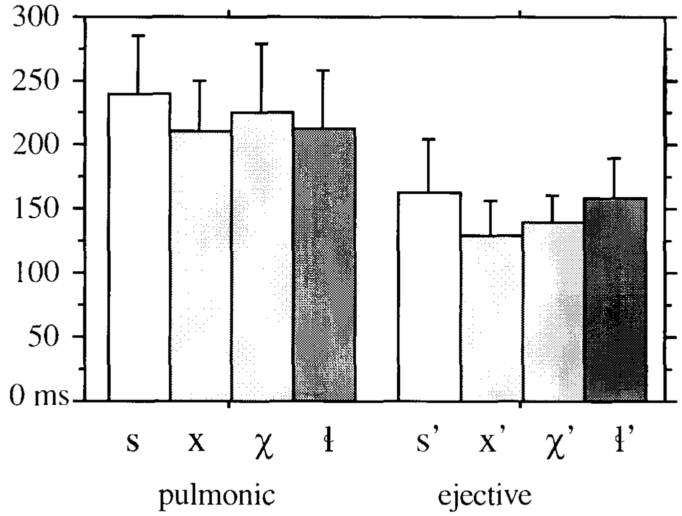


Figure 11. Mean friction durations in pulmonic and ejective fricatives.

The mean duration of the silent interval between the end of frication and the onset of the vowel is 46 for the ejective fricatives. The means for the different places do not vary by more than 3 ms from the overall mean. Because a few of the pulmonic fricatives have a brief silent interval this measure is not zero for this class of consonants, but the mean is only about 1 ms.

Aerodynamic data.

In addition to examining the acoustic records of the ejective fricatives and their plain counterparts, a simple technique was used to collect some aerodynamic data for those fricatives which have a constriction in the front of the mouth, namely /s, s', ɬ, ɬ'/. Speakers were asked to produce words or phrases containing these segments (as well as some other target segments) while holding a narrow plastic tube positioned so that it passed between their lips and its open end rested lightly against the center of the hard palate. The pressure in the oral cavity behind the constriction was detected through this tube and recorded using the pressure transduction capacity of the Macquirer hardware and software system. (The pressures recorded are deviations from the ambient atmospheric pressure.)

Data from five subjects, 4 female and one male, were recorded in this fashion. Three of these speakers are also among the group whose acoustic data was discussed above. Following each subject's session a series of calibration recordings of known pressures from 0 to 10 cm H_2O were made using a U-tube manometer connected to the same hardware. Each speaker's calibration data had a correlation coefficient (R^2) above .99 between the nominal calibration pressures and the output voltages recorded from the hardware. The response of the system can therefore be said to be linear over at least the range 0 to 10 cm H_2O.

Some problems were encountered with appropriate positioning of the tube, with the tube being blocked by the tongue and obstruction of the tube with saliva. Faulty records were usually quite easily recognizable by the failure of the pressure signal to rise appropriately at a time when the audio signal indicated that a constriction was present. Bad recordings that were recognized at the time of recording were repeated, others were discarded at a later stage of analysis. As a result of such factors, as well as the fact that some speakers preferred to give phrases rather than isolated words, somewhat different numbers and types of usable tokens were obtained from the five speakers examined. A further problem came from the very wide dynamic pressure range that was encountered. This issue will be discussed in more detail below.

Successful pressure recordings look like those shown in Figure 12 comparing the words /kʷééɬ/ 'sack' and /ʔééɬ'/ 'salt' as pronounced by Speaker 5, one of the female speakers. Two channels of data are shown, the intraoral pressure and an audio signal recorded simultaneously. The pressure signal is sampled at nearly 3 kHz; a three-point smoothing has been applied to the signals shown in the figure. The audio signal is sampled at over 11 kHz. Note that the gain of the audio channel was set at a level designed to facilitate judgements of the onset and offset of frication. As a result, the audio waveform is usually clipped during vowels, but not during the fricatives. A third signal, from an electroglottograph, was also recorded. This signal showed violent changes associated with the ejective fricatives, which is certainly consistent with their being genuinely ejective segments. However, the laryngograph signals are not amenable to analysis, perhaps because the electrodes tended to shift position, and are not shown in the figure.

Figure 12 shows that intraoral pressure reaches a considerably higher maximum during the ejective fricative and remains elevated for a shorter duration. The pressure curve in the ejective fricative is essentially parabolic in shape, whereas the pulmonic fricative shows a flattened peak.

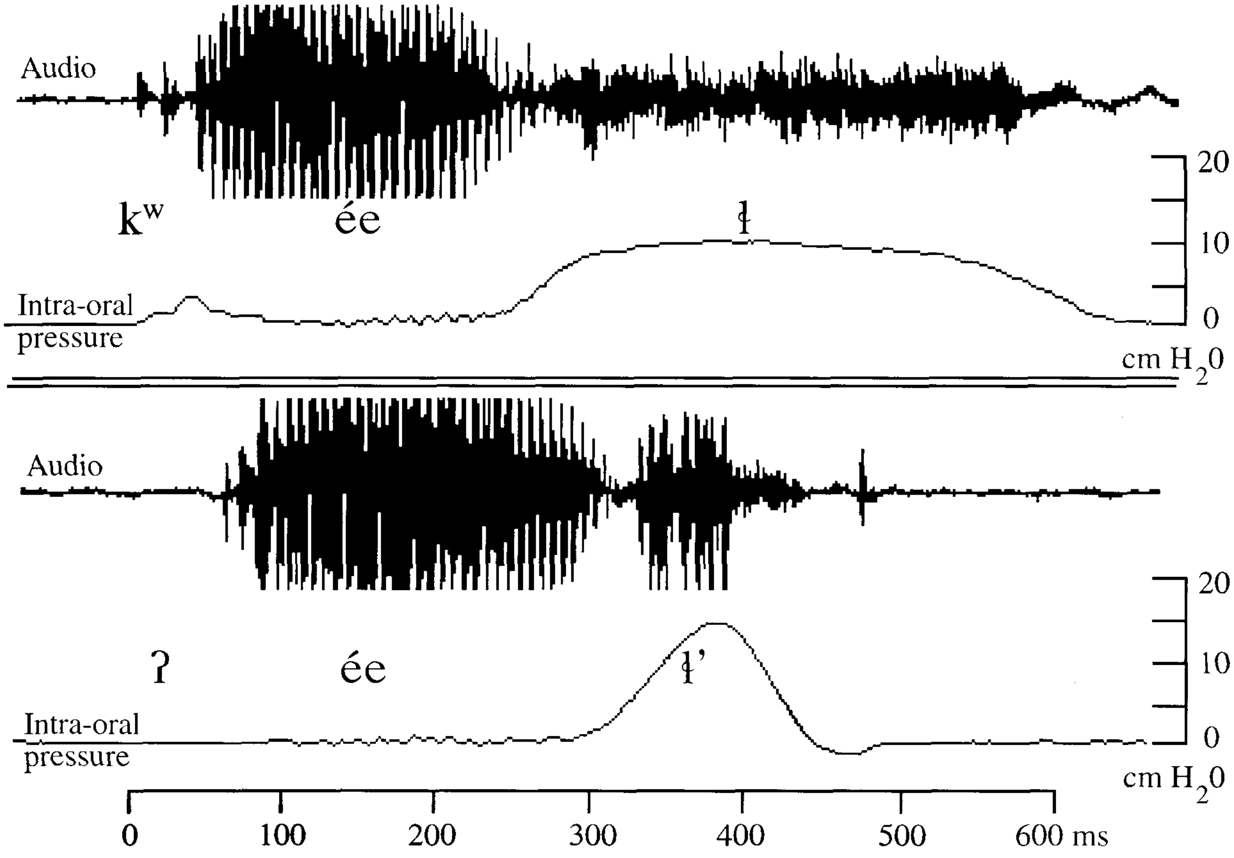


Figure 12. Audio and intra-oral pressure recordings of the words **kʷéeɬ** 'sack' and **ʔéeɬ'** 'salt' illustrating word-final plain and ejective lateral fricatives (Speaker 5, female).

The pressure data for the lateral fricatives /ɬ'/ and /ɬ/ and the sibilants /s'/ and /s/ from the five speakers was also subjected to statistical analysis. The number of tokens measured is given in Table 7. Most of these are of fricatives in word-initial or word-final position of words spoken in isolation. A certain number are of the same words in phrasal contexts. For one speaker the word-medial /s/ in /ɬaak'ásk/ 'black seaweed' was measured to make up for the small number of other tokens of /s/ available.

Table 7. Number of tokens of different fricatives measured for intra-oral pressure, by speaker.

	S1	**S2**	**S3**	**S4**	**S5**	**Total**
ɬ'	4	8	4	5	6	27
ɬ	4	6	3	4	6	23
s'	7	6	5	8	6	32
s	8	9	2	7	7	33

The maximum recorded pressure during the fricative constriction was measured for each token (with pressure being calculated from the individual regression equation derived from that speaker's calibration series). The mean results are plotted in Figure 13. A one-way analysis of variance showed that the peak pressure is significantly different for different fricatives ($F(3, 111) = 111.4$, $p < .0001$). Post-hoc tests using Fisher's PLSD test adjusted for unequal cell sizes show all pairwise comparisons to be highly significant ($p. < .0005$ in the lowest case). The sibilant alveolar fricatives have higher peak pressures than their lateral counterparts. Obviously, however, the main difference is between the two ejective fricatives, which have much higher peak pressures, and their plain voiceless counterparts, which have lower peak pressures.

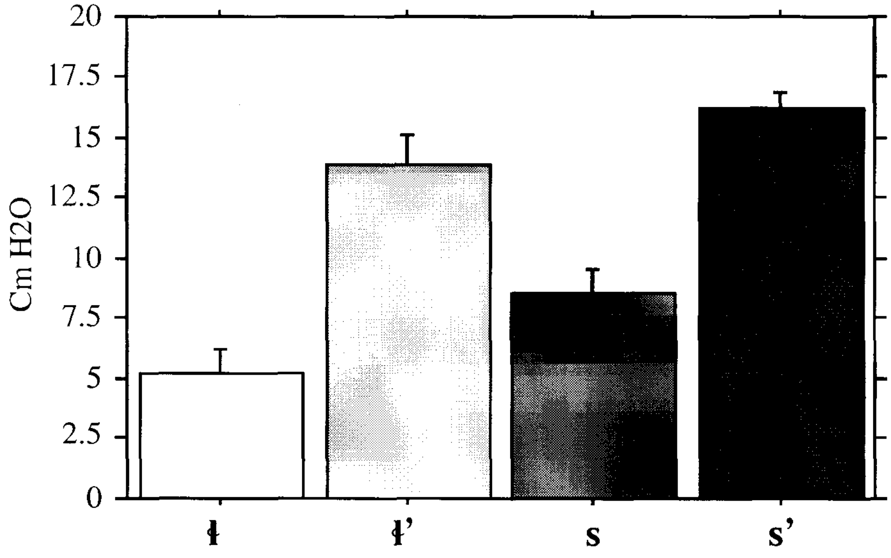


Figure 13. Mean measured values of peak pressure in plain and ejective fricatives. Data from five speakers, as reported in Table 7.

This comparison actually underestimates the magnitude of the differences between the ejective and plain fricatives. A considerable number of the ejective tokens had a peak value which exceeded the response range of the system settings being used, resulting in 'clipping' of the signal. Most of this clipping was not apparent until the signals were examined at a later time on an expanded time scale. Since the true peak value was higher, a second analysis was made using small adjustments to partly reflect the degree of clipping. When a peak was judged to be 'severely clipped' 1.0 cm H2O was added to the calculated pressure value, 0.5 was added to 'clipped' tokens and 0.1 to 'slightly clipped' tokens. The adjusted results are plotted in Figure 14. These cautious adjustments probably still understate the true differences between the plain and ejective fricatives, but emphasize that the sibilant ejective has decidedly higher pressure than its lateral counterpart. Post-hoc comparisons of the adjusted values show all pairwise comparisons to be significant at better than the .0001 level. Although there are some significant differences between the speakers, Figure 15 shows that the pattern of results is similar for all five speakers. Speaker 3 is the one male speaker in this data set.

Table 8. Unadjusted and adjusted peak pressure values in cm H_2O, rounded to one place of decimals, for the four fricatives /ɬ', ɬ, s', s/.

	unadjusted	**adjusted**
ɬ'	13.9	14.1
ɬ	5.2	5.2
s'	16.3	17.0
s	8.6	8.6

The values plotted in Figures 13 and 14 are given in Table 8. The ordinary driving pressure from the lungs during phonation is usually taken to be within the range 5-10 cm H_20 (Baken

1987). The peak pressures for the plain fricatives are therefore within the range considered as normal for pulmonic pressure during speech (at ordinary levels of loudness). That for /s/ is near the higher end of this range, while that for /ɬ/ is somewhat lower. Peak pressures in both /ɬ', s'/ are substantially higher than this pressure, and equate to levels observed in shouting.

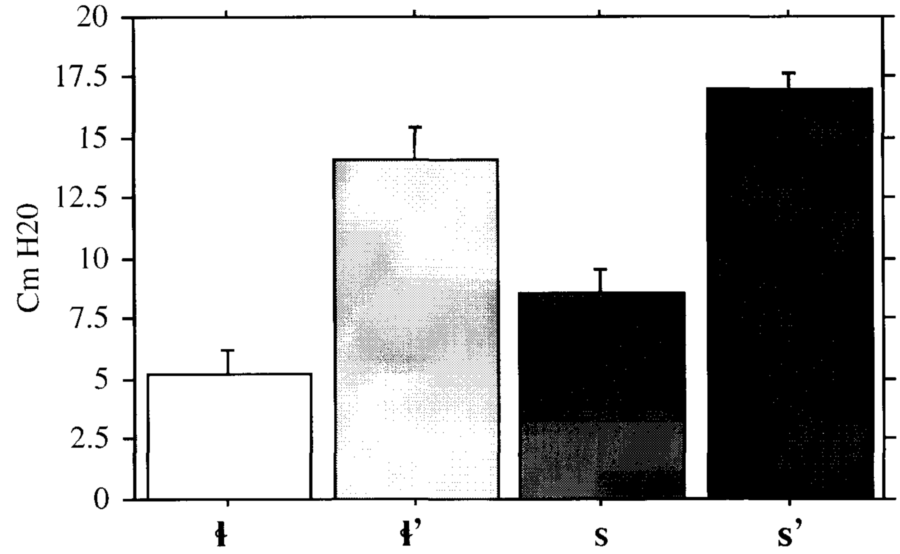


Figure 14. Mean adjusted values of peak pressure in plain and ejective fricatives. Data from five speakers, as reported in Table 7.

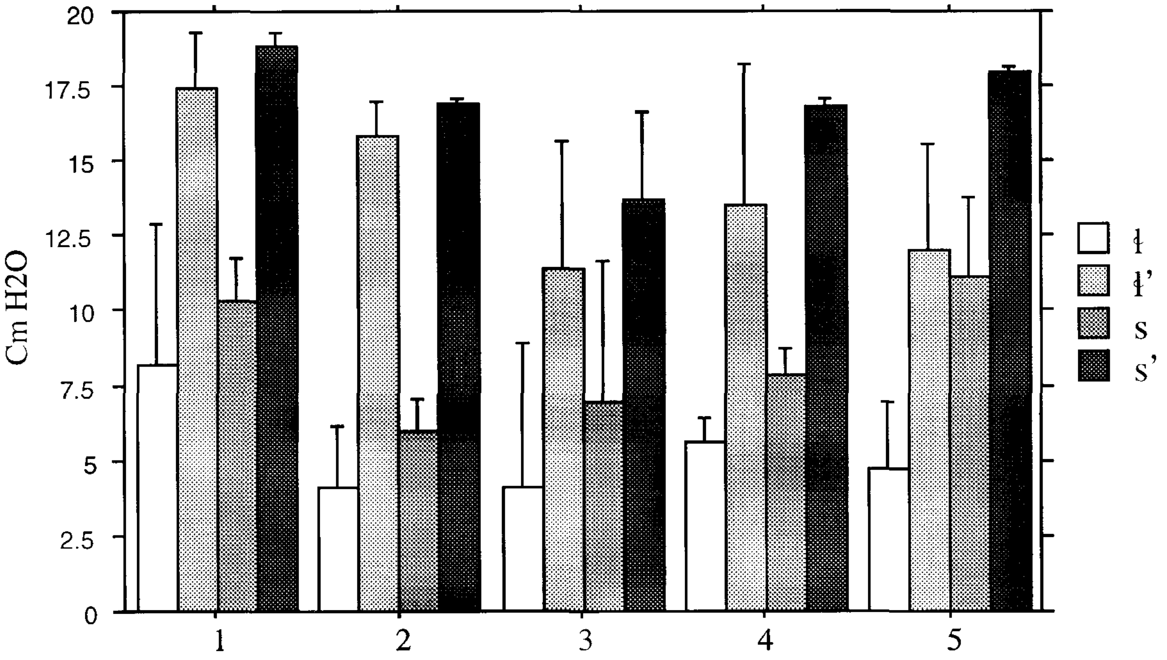


Figure 15. Adjusted peak intraoral pressure in plain and ejective fricatives, by speaker.

Figure 12 illustrated a shorter duration of elevated intra-oral pressure for the ejective fricatives, and a shorter duration of detectable frication noise in the audio waveform. The duration of the elevated pressure and of frication were measured on all the tokens in which the relevant onsets and offsets could be determined. The software used quantizes the pressure signal at 8 bits (i.e. as one of 256 unit values). In order to make consistent measurements, the beginning and end of the elevated pressure duration were determined by applying an algorithm which located the time-points furthest from the peak of pressure at which values of the pressure function were 10 units above the baseline pressure with no intervening values that were lower.

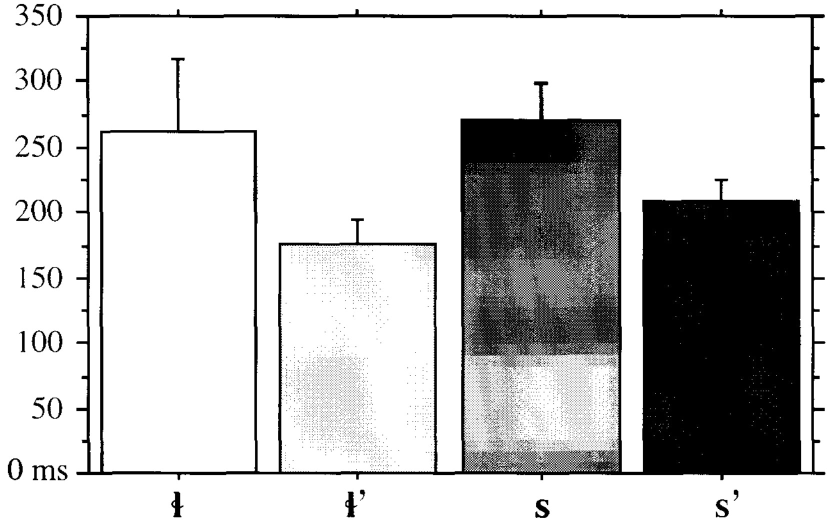


Figure 16. Mean elevated pressure duration (see Table 9).

The means of the resulting measurements are shown in Figure 16. Analysis of variance showed a significant main effect for the four different fricatives (F (3, 104) = 9.553, p <.0001). Post-hoc comparison of means showed that there is no significant difference between /ɬ'/ and /s'/, nor between /ɬ/ and /s/, but that /ɬ'/ has a significantly shorter duration of elevated pressure than /ɬ/ (p = .0002), and /s'/ has a significantly shorter duration than /s/ (p = .0015).

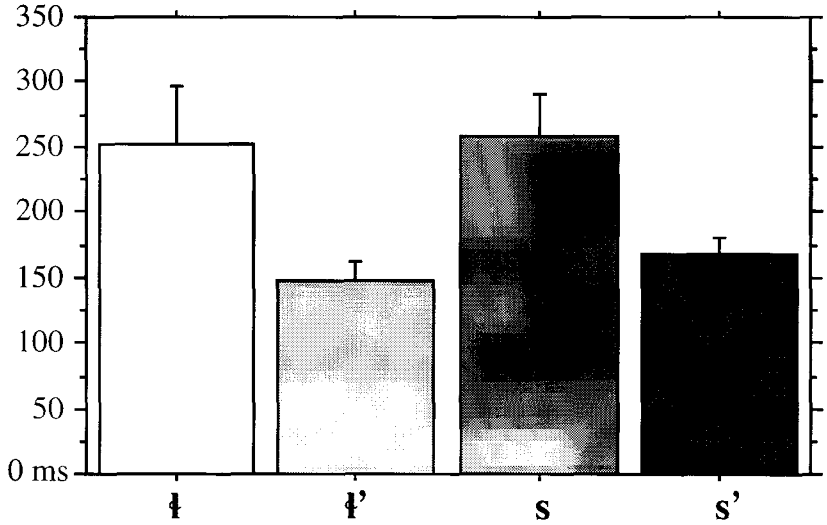


Figure 17. Mean frication duration from aerodynamic experiment (see Table 9).

The duration of detectable frication noise was measured from expanded displays of the audio waveforms recorded at the same time as the aerodynamic data collection. The means of the resulting measurements are shown in Figure 17. Analysis of variance showed a significant main effect for the different fricatives (F (3, 106) = 16.815 $p < .0001$). Post-hoc comparison of means showed that there is no significant difference between /ɬ'/ and /s'/, nor between /ɬ/ and /s/ but that any comparison between an ejective and a non-ejective fricative is significant at better than the .0001 level. These results agree quite closely with the findings from the audio recordings.

The means plotted in Figures 16 and 17 are given in Table 9 together with the number of tokens represented by each mean. Across the data set the frication duration and elevated pressure duration measures are quite highly correlated with each other ($R^2 = .711$). Neither of these duration measures show a high (negative) correlation with the peak pressure measures.

Table 9. Mean elevated pressure duration in ms and frication duration (measured from the audio waveform) for the four fricatives /ɬ', ɬ, s', s/ in the aerodynamic experiment. (Number of tokens measured in parentheses.)

	Elevated pressure duration (n)	**Frication duration (n)**
ɬ'	175 (25)	148 (25)
ɬ	262 (25)	251 (23)
s'	209 (31)	168 (31)
s	270 (32)	258 (31)

Interpretation of the results

The fricatives that we have been referring to as ejectives have the following characteristics: they are flanked by glottal closures (demonstrated by preceding and following 'silent intervals' following glottal releases, etc); they have high intraoral pressure equal to a level of pulmonic pressure that would normally be generated only in fairly loud shouting; they have quite long frication duration, though shorter than the pulmonic fricatives; they have a tendency for 'scrapiness' or pulsing during the frication. The indications that a glottal closure overlaps the entire frication duration coupled with the high intraoral pressure and the violent effects noted in the laryngograph signals suggest that they are indeed ejective. The high pressure is unlikely to be due to a brief increase in subglottal pressure to shouting levels plus a rapid opening of the vocal folds, but could be due to vigorous raising of the larynx while the glottis is closed. Equally, the pulsing pattern seen in the ejective fricatives is unlikely to be due to tense vocal fold vibration even though the visual impression given by the pattern on waveforms and spectrograms is reminiscent of laryngealized voicing. Given the very high intraoral pressure in these segments, an improbably high subglottal pressure would be required to drive vocal fold vibration. We conclude that solely the pressure generated by reducing the volume of the oral cavity above the larynx produces the outward flow required to generate the frication in these segments.

Nonetheless the frication duration is quite long, about 140-150 ms in the admittedly carefully spoken words we have studied. We have no comparable data on ejective fricative duration in other languages but have the impression that it is much shorter in Hausa /s'/ with which we are familiar, and the frication in Tlingit ejective fricatives is certainly longer than that which occurs in the fricative portion of the ejective affricates which occur in several other languages we are familiar with, including Tsez (see this volume). How is the length of frication sustained when a finite small volume of air is all that is available? We suggest that Tlingit ejective fricatives are produced by using a narrower articulatory constriction than is used in their pulmonic counterparts. Forcing the air through a narrower constriction would slow down the rate at which the available air was consumed. It would also be consistent with the tendency to scrapiness and pulsing in the fricative noise, as either the articulators or saliva would be more likely to intermittently obstruct a narrower constriction. A more constricted escape channel would also contribute to generating the high intraoral pressure observed in these segments. (A very small sample of palatographic data comparing just four words suggests that this idea is plausible, though it needs greater study.)*

We suggest that Central Tlingit speakers use a deliberate strategy of reducing the size of the escape channel in fricative articulations in order to produce the particular type of ejective fricative sounds that are so distinctive of the language. Understanding this strategy could assist instructors teaching the language to help learners master the production of these sounds .

Acknowledgments

The work reported here would not have been possible without the enthusiastic support and hospitality provided in Juneau by the Sealaska Heritage Foundation and its director Rosita Worl, who provided us with both a warm welcome and a base from which to do our research. Our thanks go to the Sealaska Corporation Board and all members of the Foundation staff for their generous assistance. For specific help with our research we owe an enormous debt to Nora and Richard Dauenhauer, as well as to other members of their family. The biggest acknowledgment of all, however, must go to the speakers who shared their knowledge of the Tlingit language with us. To Nora Marks Dauenhauer, John Marks, Amy Nelson, Selena Everson, Amos Wallace, Dorothy Wallace, Eunice Akagi, Anita Lafferty, Joseph Kinch, and Shirley Kendall we say, inadequately, *Gunalchéesh.*

References

Baken, R. (1987). *Clinical Measurement of Speech and Voice.* College Hill Press, Boston.

Bessell, Nicola J. (1996). "Tlingit F': ejective or glottalised?". Paper presented at the annual winter meeting of the Society for the Study of Indigenous Languages of the Americas, San Diego.

Boas, Franz. (1917). *Grammatical Notes on the Language of the Tlingit Indians (Anthropological Publications, Vol 8, No 1).* University Museum, University of Pennsylvania.

Dauenhauer, Nora Marks, and Dauenhauer, Richard. (1984). *Tlingit Spelling Book.* (Third edition). Sealaska Heritage Foundation, Juneau.

Dauenhauer, Nora Marks, and Dauenhauer, Richard. (1987). *Háa Shuká, Our Ancestors: Tlingit Narratives.* University of Washington Press, Seattle.

Dauenhauer, Nora Marks, and Dauenhauer, Richard. (1990). *Haa Tuwunáagu Yís, forHealing Our Spirit: Tlingit Oratory.* University of Washington Press, Seattle.

Dauenhauer, Nora Marks, and Dauenhauer, Richard. (1994). *Haa K̲usteeyí, Our Culture: Tlingit Life Stories.* University of Washington Press, Seattle.

---

* The acoustic data for one of the speakers discussed in Bessell (1966) — a speaker who is also included in this study — is entirely consistent with the results reported here. The second speaker shows longer friction durations and no 'silent interval' separating his "ejective" fricatives from a following vowel. This speaker is apparently doing something different from the speakers we studied. He is from the Southern Tlingit area and thus seems to confirm the impression that there are noticeable differences between the Central and Southern varieties of the language.

Dauenhauer, Nora Marks, and Dauenhauer, Richard. (1991). *Beginning Tlingit.* Sealaska Heritage Foundation, Juneau.

Dauenhauer, Richard and Dauenhauer, Nora Marks. (1995). "Oral literature embodied and disembodied." In *Aspects of Oral Communication*, U. M. Quasthoff (ed.), de Gruyter, Berlin and New York: 91-111.

Davis, Henry (editor). (1976) *English-Tlingit Dictionary: Nouns.* Sheldon Jackson College, Sitka. Reprinted 1996, with Sealaska Heritage Foundation, Juneau.

Emmons, George Thornton. (1991). *The Tlingit Indians,* edited with additions by Frederica de Laguna (Anthropological Papers of the American Museum of Natural History, 70). University of Washington Press, Seattle, and American Museum of Natural History, New York.

Kan, Sergei. (1989). *Symbolic Immortality: Tlingit Potlatch of the Nineteenth Century.* Smithsonian Institution Press, Washington, D.C.

Krauss, Michael E. (1973). "Na-Dene." In *Current Trends in Linguistics 10: Linguistics in North America,* T. Sebeok (ed.), Mouton, The Hague: 903-978.

Krauss, Michael E. (1979). "Na-Dene and Eskimo-Aleut." In *The Languages of Native America: Historical and Comparative Assessment*, L. Campbell and M. Mithun (eds.), University of Texas Press, Austin and London: 803-901.

Ladefoged, Peter and Maddieson, Ian. (1996). *The Sounds of the World's Languages.* Blackwells, Oxford.

de Laguna, Frederica. (1972). *Under Mount Saint Elias* (Smithsonian Contributions to Anthopology, Volume 7, in 3 parts). Smithsonian Institution Press, Washington, D.C.

Leer, Jeff. (1978). "Introduction" to *Tongass Texts* by Frank Williams and Emma Williams. Alaska Native Language Center, Fairbanks: 6-17.

Leer, Jeff. (1990). "Tlingit: a portmanteau language family?". In *Trends in Historical Linguistics*, P. Baldi (ed.), Mouton, Berlin: 73-98.

Leer, Jeff. (1991). *The Schetic (sic) Categories of the Tlingit Verb.* Unpublished Ph.D. dissertation. University of Chicago, Chicago.

Lindau, Mona. (1984). "Phonetic differences in glottalic consonants." *Journal of Phonetics* 7: 147-155.

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

Maddieson, Ian. (1996, to appear). "Phonetic universals". In *Handbook of Phonetic Sciences*, J. Laver and W. Hardcastle (eds.), Blackwells, Oxford.

Naish, Constance M. and Story, Gillian L. (1963). *English-Tlingit Dictionary: Nouns.* Summer Institute of Linguistics, Fairbanks.

Naish, Constance M. and Story, Gillian L. (1973). *Tlingit Verb Dictionary.* Summer Institute of Linguistics and Alaska Native Language Center, Fairbanks.

Nyman, Elizabeth and Leer, Jeff. (1993). *Gágiwdul.àt: Brought Forth to Reconfirm: The Legacy of a Taku River Tlingit Clan.* Yukon Native Language Center, Whitehorse, and Alaska Native Language Center, Fairbanks.

Pinnow, Heinz-Jürgen. (1966). *Grundzüge einer historischen Lautlehre des Tlingit: ein Versuch.* Harrassowitz, Wiesbaden.

Sapir, Edward. (1929). "Central and North American languages." *Encyclopaedia Britannica, 14th Edition*, volume 5: 138-142.

Swanton, John R. (1908). "Social conditions, beliefs and linguistic relationships of the Tlingit Indians." Bureau of American Ethnology, Annual Report 26, Smithsonian Institution, Washington, D.C.: 391-512.

Swanton, John R. (1909). *Tlingit Myths and Texts.* Bureau of American Ethnology, Bulletin 39, Smithsonian Institution, Washington, D.C.

Notes on the phonetics of Sele, with particular attention to vowels.

Ian Maddieson and Matthew Gordon

1. Introduction

Sele (sɛ-lɛ́) is a language of Ghana spoken by the people usually known officially and to outsiders as the Santrokofi. This name, a place-name for a former settlement of the people, is often also used as the name of the language. Sele is spoken by about 6000 speakers primarily in the three villages Benua, Bume and Gbodome, which lie along the road leading north from Hohoe toward Jasikan in the Volta Region of Ghana. The map in Figure 1 indicates the location where Sele is spoken.

Figure 1. Map of South-Eastern Ghana, showing the location where Sele is spoken.

Sele is a member of the group of languages which have been variously called the 'Togo Remnant', 'Central Togo' or 'Togo Mountain' languages (Heine 1968, Kropp Dakubu 1988, Ring 1987). These labels are all attempts to reflect the socio-geographic nature of this grouping. All of the languages concerned are spoken by relatively small populations living mostly in or near the highlands which lie near the present-day international border between Ghana and Togo. These populations largely stood aside from the major political and military events involving the larger peoples in the area, such as the Akan and the Ewe. The languages are all members of the Kwa branch of the Niger-Congo family (Stewart 1989) and are obviously related, but not such that all the languages within the grouping are more closely related to each other than to any language outside the group. However, the closest genetic relations of Sele are with other 'Togo Remnant' languages, namely with Sekpele (Likpe) and Siwu (Lolobi-Akpafu).

The data reported in this paper was collected in Bume and Gbodome in November 1994 by the first author, with the major analysis completed at the UCLA Phonetics Laboratory by the second author. The focus of this report will be on the vowels of Sele, particularly on the basis of its vowel harmony system, but since previous descriptions of the language are not easily available or are inaccurate in some respects we first present a summary of the overall phonological system.

2. Phoneme Inventory

2.1. Consonants

Previous published and unpublished descriptions of Sele phonology (Funke 1911, Heine 1968, Allen 1974), differ in the number of consonants they recognize. This is largely because they differ in how they approach the various phonological processes which create additional phonetic segments. Our interpretation of the consonant phonemes of Santrokofi, which is the same as that suggested by Allen (1974) but independently verified in the field, is as shown in Table 1. In this view there are 15 distinctive consonants in Sele.

Table 1: The consonants of Sele

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Labial velar
Plosives	p b		t	ɖ		k	k͡p
Affricates					tç		
Fricatives		f	s				
Nasals	m		n		ɲ	ŋ^{w}	
Approximants					j		w

Except at the bilabial place, the voiceless stops and affricates in Table 1 do not have voiced counterparts, although such segments are listed among the consonant phonemes by Heine (1968). Funke (1911) also lists voiced fricatives /v, z/, which do not occur.

The consonant transcribed in Table 1 as /ɖ/ is pronounced as a voiced post-alveolar plosive before the high vowels /i, u/ but is realized as an alveolar lateral approximant, [l], before mid and low vowels. Other analyses (e.g. Heine 1968) have chosen to represent the lateral as the 'basic' variant, perhaps because it occurs in a larger range of positions. One important case in which the allophony between [ɖ] and [l] is apparent is in the noun class prefix /ɖi-/ (cognate with the Proto-Bantu class 5 prefix *di 4 or *li 4). Before stems with a high vowel in their first syllable (and an oral initial consonant) this prefix has the form [ɖi-], but before stems whose first vowel is non-high the prefix has the form [le-] or [lɛ-]. (Stems with a high first vowel and an initial nasal consonant show the further variant [ni-].) Variations of this kind provide important clues to the reconstruction of earlier stages of the phonological system of Sele and its relatives (Ford 1973, Maddieson 1995). These matters will be discussed in more detail in the section on vowel harmony below.

Heine (1968) treats the palatal nasal [ɲ] as an allophone of a velar nasal phoneme /ŋ/. In Heine's view [ɲ] occurs in most vowel environments, with a velar [ŋ] (with labialization) occurring only before /o, u/ and /w/. But this is not quite correct. Phonetic [ɲ] occurs at least before /u/, and there is no warrant for the independent segmental status of /w/ in occurrences of [ŋʷ]. Rather, there seems to be a contrast between /ɲ/ and /ŋʷ/ before most if not all vowels, as in examples such as /kà-ɲà/ 'mouth', /bà-ŋʷàkɛ/ 'bees' and /ɲì/ 'drink (V)', /òŋʷi/ 'one'. Nasal prefixes become homorganic with the consonant they precede. Hence the additional phonetic segments [ɱ], [ɳ], [ŋ], [ŋ͡m] are created. These are not contrastive elements of the Sele consonant system.

An additional consonant /ʔ/ might be added to the inventory, but the glottal stop does not serve independently to distinguish lexical forms, but rather seems to be an aspect of the realization of tones and of features of the phrasal phonology. A similar situation is reported in the nearby language Avatime by Schuh (1995). We do not at present know the precise distribution or full range of functions of [ʔ] in Sele.

2.2. Vowels

According to Heine, there are seven oral vowels in Sele, and a nasalized counterpart of each of these seven. It is true that vowels which immediately follow nasals are quite markedly nasalized, but this is a contextual allophonic process. Non-allophonically nasalized vowels are not common, and it appears that not all seven qualities occur distinctively nasalized. Allen (1974) recognizes only two distinctive nasalized vowels, the two high vowels /ĩ/ and /ũ/. These occur in words such as /kú-tĩ/ 'thigh', /ku-fĩ́/ 'Sun' and /ká-tũ/ 'forehead', /sṹsũ/ 'housefly'. She suggests that there are only three other nasalized vowel qualities, /ɛ̃, ã, ɔ̃/, with the higher and lower mid qualities being neutralized under contextual nasalization. Our own observations support this neutralization. But there are at least a handful of other words where non-high nasalized vowels occur initially or after an oral consonant. We noted only one example of each: /ɛ̃ɛ̃/ 'yes', /ɔ-fã́/ 'half', /o-bísɔ̃/ 'child' (and not all speakers nasalize the final vowel in 'child'). Heine's lexicon includes a few additional words with non-allophonic /ã/.

We therefore suggest that the vowels of Sele are as in Table 2 with the proviso that the non-high nasal vowels have a somewhat marginal status. At present we represent the vowel system as if it contrasted four levels of vowel height. An alternative possibility will be considered below.

Table 2: The vowels of Sele

Oral Vowels		Nasalized Vowels	
i	u	ĩ	ũ
e	o		
ɛ	ɔ	ɛ̃	ɔ̃
a		ã	

There is one further point to be made about nasalized vowels in Sele. As Allen noticed, they do not follow either of the two voiced oral obstruents, /b, ɖ/, nor the approximants /j, w/. A conceivable alternative analysis would therefore be to regard [m, n, ɲ, ŋʷ] as the allophones of /b, ɖ, j, w/ which occur before nasalized vowels, instead of treating vowel nasalization as allophonic after nasals. This would recapitulate what is clearly the historical origin of some occurrences of [m, n] and possibly all cases of [ɲ, ŋʷ]. However, since speakers do not seem to be aware of this distributional pattern, and nasals occur before consonants as syllabic prefixes, we prefer the more conservative analysis presented here.

2.3. Tones

Sele is a tonal language with lexical contrasts both in nouns and verbs depending on tonal differences. In addition, tone alternations in verbal paradigms play an important role in signaling tense/aspect and negation. According to Heine, there are five tones in Sele: high, medium and low level tones, as well as falling and rising contours. These all occur on single vowels, but there seem to be few stems with a lexical contour tone, and we suspect that many of the contours could be analyzed as the result of processes of tone-shifting and coalescence, and the tone inventory simplified to only the level tones. Nonetheless, surface contrasts such /fá ɲà/ 'you (sg) didn't see (near past)' vs /fà ɲâ/ 'you (sg) saw (remote past)' establish that levels and contours must be distinguished by the listener. Additional contours, such as mid-high, mid-low and low-mid occur on two-vowel sequences.

In addition to the level tones mentioned above, Allen (1974) also distinguishes an extra low tone, but there seems to be a complementarity of realization between low and extra low. Allen makes clear that only a limited set of the possible sequences of tones occur on nouns. Much remains to be done to advance our understanding of the Sele tone system. In this paper, an impressionistic transcription of the tones is given. In the field little time was devoted to checking these transcriptions, as the focus was on the consonant and vowel segments of the language, and they should be treated with some caution, not to say scepticism.

2.4 Vowel Harmony

In common with most of the Niger-Congo languages of West Africa, Sele has harmonic rules governing the distribution of vowels in stems and clitics. The lower mid vowels /ɛ, ɔ/ are restricted in their co-occurrence with other vowels within stems, so that they do not co-occur with high or with higher mid vowels, i.e. with the set /i, e, o, u/ (except in some loanwords and in compounds). The nasalized high vowels also appear not to co-occur with /ɛ, ɔ/. The low vowel /a/ is unrestricted in co-occurrence within stems.

Clitics with mid vowels which occur before verb stems follow the same restrictions as on mid vowels within stems; that is, a verb prefix with a mid vowel will have a variant with /ɛ/ or /ɔ/ before a verb with a lower mid vowel. In addition, the same lower mid vowel variant occurs before verbs with /a/.

Nouns show a slightly more complex interaction between their stems and prefixes. The three noun class prefixes with high vowels — the singular class prefixes /ɖi-/ and /ku-/, and the plural class prefix /si-/ — have variants with higher mid and lower mid vowels. The singular mid vowel prefix /o-/ as well as the rare plural class prefix /e-/ have lower mid variants. Prefixes with the vowel /a/ do not vary. The patterns of the major noun class genders are laid out in Table

3, with the singular prefix to the left of the slash and the plural prefix to the right. The numbering is that used by Allen. Words illustrating these patterns are given in Table 4.

Table 3. Noun class genders of Sele.

Noun class genders	I	II	III	IV	V
before /i, u/	o-/ba-	ɖi-/a- [nĩ-] before N	o-/si-	ka-/N-	ku-/a-
before /e, o/	o-/ba-	le-/a-	o-/se-	ka-/N-	ko-/a-
before /ɛ, ɔ, a/	ɔ-/ba-	le-/a-	ɔ-/se-	ka-/N-	kɔ-/a-

Table 4. Words illustrating the noun class genders of Sele.

I	II	III	IV	V
ò-tíi/bà-tíi 'person(s)'	ɖi-si/á-si 'head(s) ní̃-nu/ánu 'eye(s)'	ó-kũ/si-kũ 'steel trap'	ká-tũ/ń-tũ 'forehead'	kú-tu/á-tu 'soup(s)'
o-té/ba-té 'father(s)'	lè-tòo/a-tòo 'load(s)'	ò-kpoo/sè-kpoo 'town'	ka-tçé/n-tçé 'night(s)'	kó-to/á-to money, 'coin(s)'
ɔ̀-ka/bà-ka 'chief(s)'	lɛ́-tçá/á-tça 'roof(s)'	ɔ-tɔ́/sɛ-tɔ́ 'fire(s)'	ká-ma/ḿ-ma 'back(s)'	kɔ-pá/a-pá 'cutlass(es)'

The alternation between high and mid vowels in prefixes illustrated in Tables 3 and 4 demonstrates that vowel height is a factor in the vowel harmony system of Sele. The question that arises is whether the alternations among mid vowels also simply reflect vowel height, or whether the phonetic basis of these alternations is tongue root position. In a number of the languages in West Africa with vowel harmony systems it is known that independent adjustments of the tongue root position form the basis of the harmony, with vowels with the tongue root advanced co-occurring in one harmony set, and vowels with a more retracted tongue position forming the other. Quite commonly low vowels do not differ in tongue root position and may co-occur with either set. It is almost certain that tongue root position did play a role in the vowel system of earlier stages of Sele and its closely related neighbors. The former system distinguished a larger number of vowels, at least nine and perhaps 10, but has clearly undergone a reorganization (Ford 1973).

One of the goals of this paper is to examine if it is possible to determine from the principal acoustic characteristics of the vowels whether the Sele vowel harmony pattern remains based on an ATR distinction between the two sets of mid vowels or is better described solely as a height harmony system, as it has been to this point in the paper. If it is an ATR system /e, o/ in Table 2 would be classified as [+ATR] and correctly transcribed as /e̘, o̘/ while /ɛ, ɔ/ would be [-ATR] and transcribed as /e̙, o̙/. If the system is based on height, the transcription in Table 2 would be the appropriate one. For simplicity we will continue to represent the mid vowels as /e, ɛ, o, ɔ/.

3. Acoustic characteristics of Sele vowels

3.1. Vowel Formants

Formant values for the first three formants of the seven oral vowel were measured in a selected set of words spoken by 14 speakers of Sele, eight men and six women. The audio recordings were digitized onto the Kay Elemetrics Computerized Speech Lab (CSL) at 10 kHz. Formants were calculated for a steady state portion of each vowel at about the mid-point of the vowel, as determined from a spectrogram. Formants were obtained from the LPC analysis function, using 14 coefficients for most of the data. Superimposed LPC and FFT spectra, calculated over 30 and 25.6 ms frames, respectively, beginning with this halfway point were displayed and the formant values calculated by the LPC analysis were checked against the FFT spectrum and spectrographic display. Corrected values were measured from the FFT display where the LPC analysis appeared discrepant. The number of tokens measured varied depending on vowel and speaker, with the most tokens for /a/ and the fewest for /e/ for all speakers.

The mean values for the first three formants of each vowel were calculated for each speaker separately. These results are shown in the Appendix. The first two formants are plotted against each other for the male speakers in Figure 2 and for the female speakers in Figure 3. Each vowel symbol on these figures represents a single speaker's mean for a given vowel. The distances along the axes correspond to a Bark-scale transformation of the linear Hz values. The origin is in the upper right corner of the display.

An ellipse with a radius of two standard deviations along the axes of the first two principal components is drawn around the distribution of each set of speaker-means for a given vowel. The individual means for all three formants, and the number of tokens measured for each speaker are given in an appendix at the end of the paper.

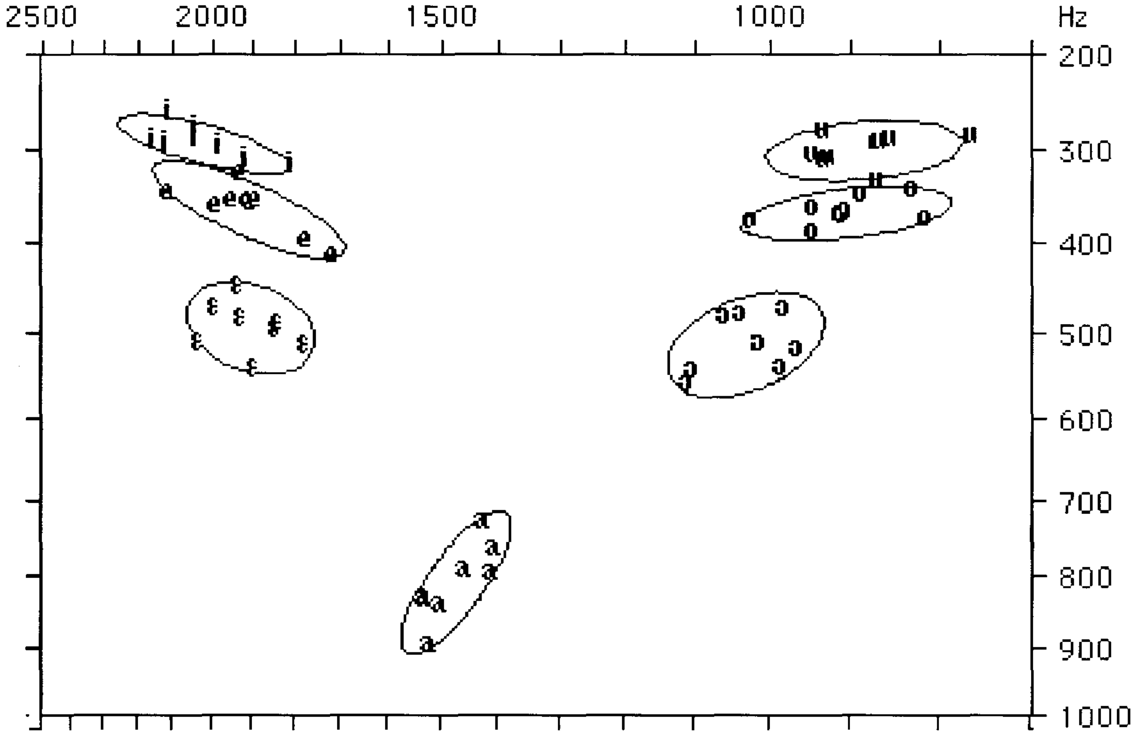


Figure 2. Plot of the means of the first two formants of each vowel for each male speaker.

None of the ellipses in the F1/F2 space calculated in this way for any vowel overlaps another for either the male or female speakers. Nonetheless, for the male speakers the high vowels and higher mid vowels have values that are quite close to each other. The low vowel is quite distant from any other vowel. In contrast the means for the female speakers' vowels are more equidistantly spaced. (We suspect that this does not represent a real gender difference but in some way is an artifact of the analysis algorithm, which may be underestimating the height of the first formant of the male speakers.)

The higher mid and lower mid vowels are clearly distinct from each other in their first formant values. The vowel /o/ also quite consistently has a lower F2 value than /ɔ/. F2 is quite similar for /e/ and /ɛ/ for speakers F2, F3, F5, M2, M3, M7; for the other speakers, /e/ has higher F2 values than /ɛ/. The third formant does not reliably distinguish between the members of these mid vowel pairs.

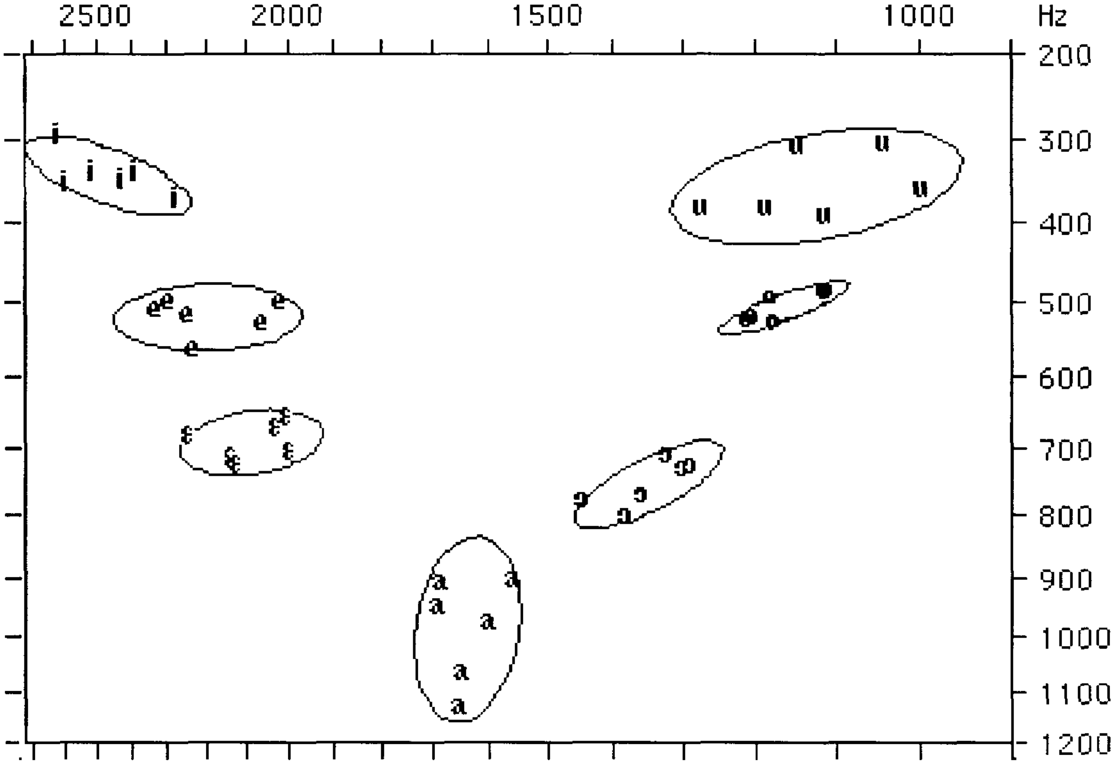


Figure 3. Plot of mean of the first two formants of each vowel for each female speaker.

3.2. Formant Bandwidth

In addition to the formant frequencies, the bandwidths of the first two formants of the mid vowels were measured in the words in Table 4. The measured vowels are shown in bold. Bandwidth values used were obtained from the LPC analysis on the Kay CSL system using a 30ms window centered over a point of the vowel judged to be not influenced by transitions from an adjacent consonant, generally at the midpoint. Bandwidths for the second formant for all mid vowels were quite large and highly variable. For both the front and back mid vowel pairs, the members of the vowel pairs did not differ significantly from each other in the bandwidth of their second formants. Bandwidth values for the first formant were less variable. Table 5 shows the values of the F1 bandwidth for the mid vowel pairs for each speaker.

Table 4: Corpus from which bandwidth measures were taken in Sele

Front Mid Vowels		Back Mid Vowels	
/e/	**/ɛ/**	**/o/**	**/ɔ/**
kōfé ‘farm’	kɔ̀tɛ̄ ‘mud’	ōté ‘father’	ntɔ̄ ‘ash’
ōtẹ́ ‘father’	kàfɛ̄ ‘axe’	ōtĩi ‘person’	ɔ̀tɔ́ ‘fire’
		kōtó ‘money’	kɔ̀tɛ̄ ‘mud’

Table 5: Bandwidth values for first formant of the mid vowels (in Hz)

Speaker	/e/	/ɛ/	/o/	/ɔ/
SF1	23.9	44.7	38.4	62.0
SF2	25.7	32.8	18.7	42.4
SF3	33.5	53.6	53.9	143.8
SF4	29.2	42.4	33.0	163.6
SF5	16.2	13.0	52.7	77.1
SF6	20.4	45.1	108.6	73.8
SM1	27.8	67.7	35.1	109.3
SM2	17.1	46.4	45.7	81.3
SM3	36.3	48.4	44.6	59.3
SM4	18.6	75.0	35.0	112.9
SM5	12.1	28.8	43.4	123.8
SM6	5.8	36.6	47.9	76.7
SM7	14.7	38.7	71.92	73.4
SM8	14.8	25.9	64.8	58.2
Grand Means	**21.2**	**42.8**	**49.6**	**89.0**

Two patterns are observable in Table 5. First, bandwidths are wider in general for the back mid vowels than for the front ones. This difference may be attributed to the relatively close proximity of the first two formants in the back mid vowels, which would widen the bandwidths of both formants for these vowels. Second, the bandwidths are generally narrower for /e/ than for /ɛ/, and for /o/ than for /ɔ/ (There are three exceptions to this pattern: the front mid vowels for speaker SF5, and the back mid vowels for speakers SF6, and SM8.) Analysis of variance showed a highly significant difference between the vowels, with post-hoc comparison of means using Fisher's PLSD test showing the members of each vowel pair differing at a level of significance better than .0001.

In sum, the differences between the mid vowel pairs are the following; in both pairs there is a robust difference in the frequency of the first formant, with /ɛ, ɔ/ having higher F1 than /e, o/. There is virtually no difference in F2 frequency between the front vowels, but /ɔ/ has a somewhat higher F2 than /o/. The first formant bandwidth is substantially greater for /ɛ/ than for /e/ and for /ɔ/ than for /o/. Bandwidth comparisons for F2 were inconclusive, as were all comparisons regarding F3.

Discussion

What is common among ATR vowel systems is generally a marked difference in F1 values between the [+ATR] and the [-ATR] vowels, the [+ATR] vowels having the lower F1 values (see Ladefoged and Maddieson 1996: 305 for formant values from six ATR systems). The low F1 of

[+ATR] vowels is attributable to the overall enlargement of the oral cavity in the pharyngeal region resulting from the tongue root advancement and supplementary movements of the larynx and other structures (Lindau 1975). F2 does not necessarily differ greatly between [+ATR] and [-ATR] counterparts, but the tongue root advancement may cause the body of the tongue to shift forward, producing a higher F2.

The difference in F1 values observed in Sele does not demonstrate that this vowel system is in fact based on ATR, since a harmony system based on height would display F1 differences in the same direction. It is the relation between F1 and F2 values, particularly for the back vowels, which needs to be examined. In a height based vowel system, lower vowels are most frequently acoustically centralized relative to higher vowels. That is, F2 for back vowels is higher the lower the vowel, but for front vowels is lower the lower the vowel. Ladefoged and Maddieson suggest that it is indicative of the use of ATR if the members of any pair of vowels believed to contrast on this parameter are located in a vowel space so that the 'higher' ([+ATR]) member of the pair is above and to the left of the 'lower' ([-ATR]) member when the axes of the plot are F1 and F2-F1 and the origin in the upper right, as in Figure 4. This figure plots a single point for each Sele vowel, which is the mean of each of the speaker means.

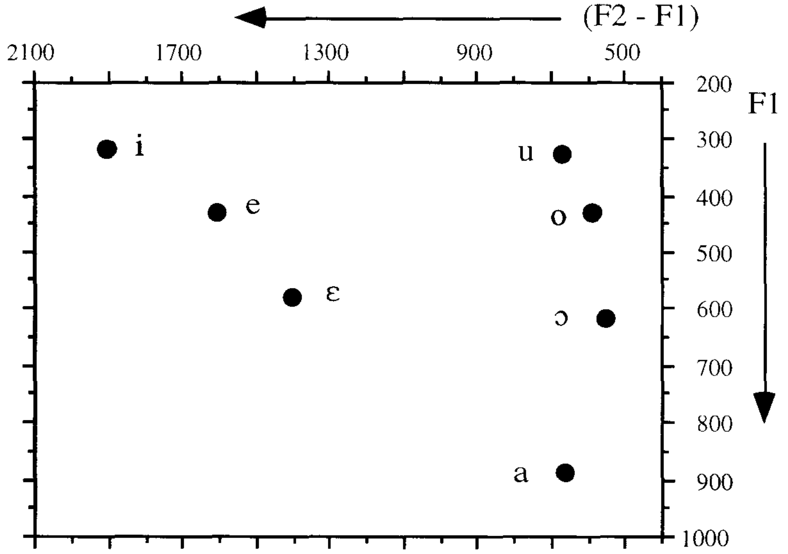


Figure 4. Mean values of Sele vowels in F1/F2-F1 space.

Figure 4 shows that by a small margin the Sele system meets this criterion, but with a large difference between the relative positions of the front mid vowel pair and the back mid vowel pair. The means are plotted in Figure 5 without subtraction of F1 from F2. In this plot a more symmetrical placement of the two pairs is seen, with the 'lower' member slightly more central than the 'higher' member in each pair. Figure 4 is thus weakly suggestive of an ATR system. In more "robust" ATR systems the unadjusted value of F2 can be lower for the [-ATR] member of a back vowel pair, but it should be noted that the magnitude and direction of difference in raw F2 values between [+ATR] and [-ATR] vowels is not consistent across languages. For example, Hess (1992) reported lower F2 values for [+ATR] /u/ than for [-ATR] /u̙/ in Kwawu Akan.

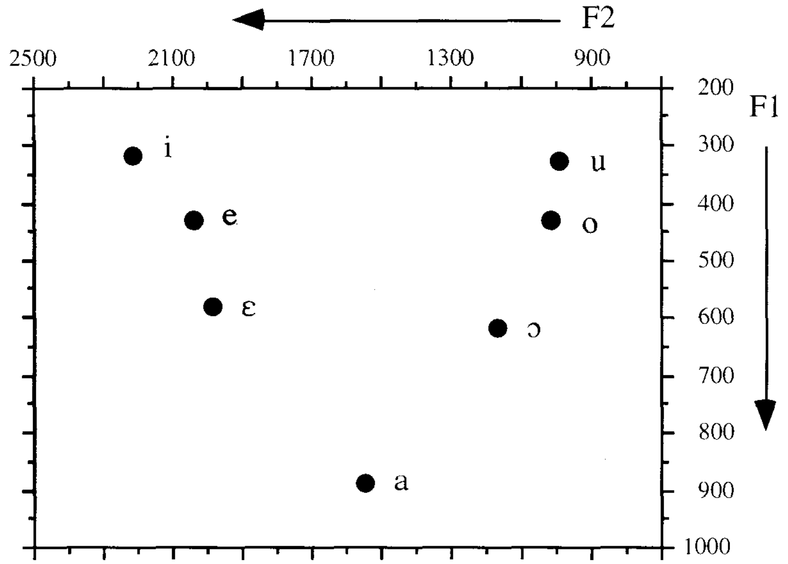


Figure 5. Mean values of Sele vowels in F1/F1 space.

In an ATR vowel system, the articulatory difference in pharyngeal cavity size may influence other acoustic properties besides the formant frequencies, in particular formant bandwidths (or, equivalently, amplitudes). In Akan, Hess (1992) showed that the [+ATR] vowels have smaller first formant bandwidths than their [-ATR] counterparts. Fulop (1996, this volume) found that the amplitude difference between the first and second formants is smaller for [-ATR] vowels than for [+ATR] vowels in Degema.

As reported above in Table 5, first formant bandwidths in the Sele vowels differ in the correct direction for the mid vowel contrasts to plausibly involve ATR. Also, as in Akan, the bandwidth differences in Sele are much greater than could attributed to simply the difference in the frequency of the first formant between [+ATR] and [-ATR] vowels. Fujimura and Lindqvist's (1971) estimated inherent bandwidth using external stimulation with a closed glottis. They show that bandwidth values for the first formant of vowels vary very little for frequencies between 400 and 700 for males and between 500 and 800 for females. Differences are less than 5 Hz in these frequency ranges; Above this range, the higher the frequency, the wider the bandwidth. At frequencies below these ranges, bandwidth is *wider* the lower the frequency, the opposite pattern from what is seen in Sele. Thus, the magnitude of the observed differences in bandwidth between the [+ATR] and [-ATR] vowels is presumably not due to any inherent frequency differences of the first formant between the [+ATR] and [-ATR] vowels.

Because, as Fujimura and Lindqvist point out, bandwidth values for natural speech would presumably be different due to the partially open nature of the glottis during speech compared to the closed glottis of Fujimura and Lindqvist's experiment, we measured first formant bandwidth in a language with mid vowels which are regarded as contrasting in height rather than ATR. Italian was chosen as the control language and the vowels measured were the high mid vowels /e, o/ and the low mid vowels /ɛ, ɔ/. Formant and bandwidth values were measured for the mid vowels produced by two speakers of standard Italian (one male and one female). (Data from a third speaker had to be regretfully omitted due to a confound with F0 patterns). The corpus, a series of disyllabic nonsense words obeying the phonotactics of Italian and read by the speakers as potentially real words, is given in Table 6. The measured vowel always occurred in the first syllable between consonants similar in place and manner of articulation to those in the measured Sele words in Table 4. Six or seven tokens of each word were measured.

Table 6: Italian words from which first formant bandwidth was measured

Front Mid Vowels		Back Mid Vowels	
/e/	/ɛ/	/o/	/ɔ/
eti	ɛti	oti	ɔti
teti	tɛti	koti	kɔti

The same measurement procedure was used for the Italian vowels as was used to measure bandwidth of the Sele vowels. Additionally, the frequency of F1 was measured to estimate the potential effect of frequency on bandwidth. Frequency and bandwidth values for the first formant of the Italian mid vowels are given in Table 7.

Table 7: Mean frequency and bandwidth of F1 for the mid vowels of two Italian speakers.

Speaker	/e/		/ɛ/		/o/		/ɔ/	
	Fq.	Bdwth	Fq.	Bdwth	Fq.	Bdwth	Fq.	Bdwth.
M1	346	99.7	565	89.1	382	137.9	593	179.2
F1	378	98.2	706	136.5	406	176.8	733	218.5

Though there are differences in mean bandwidth between the higher and lower mid vowels similar in magnitude to those found in Sele, these differences are not significant at the $p<.05$ level for either Italian speaker, in spite of the relatively large number of tokens involved for each speaker (13 or 14). It is worth noting, however, that all of the differences approach significance ($p<.10$) except for the difference between /e/ and /ɛ/ for speaker M1. It is also interesting that bandwidths for the Italian vowels are markedly wider overall than for the Sele vowels, as well as those in Fujimura and Lindqvist. A partial explanation for this might lie in a difference in the frequency distance between the first and second formants in the two languages. The first and second formants are closer together for /o/ and /ɔ/ in the Italian data than in Sele, which would tend to increase the amplitude of both. But there is no similar pattern to explain the bandwidth difference in the front mid vowels between the languages, and so it is also possible that the Italian and Sele vowels being compared are produced with a different degrees of tension of the vocal tract walls or with other differences in overall vocal tract configuration.

Nonetheless, our results suggest that the ATR vowel pairs in Sele are more clearly and consistently associated with differences in first formant bandwidth than the Italian mid vowels. Again, this is a weak indication in favor of the possibility that the vowel system continues to employ this articulatory parameter.

A final test of this hypothesis was conducted using the VTI software tool developed by Alain Soquet. This implements a version of the distinctive regions model of the vocal tract (Mrayati, Carré and Guérin 1988) to map acoustic-articulatory relations in vowels (Jospa, Soquet, and Saerens 1992). Inversion to an articulatory shape from acoustic data is performed by a neural network. Using a model enables the problem of simultaneously considering the import of the interacting values of several formants to be approached in a rigorous way. The grand means of the Sele mid vowel formants (Figure 5) were entered, and the resulting predicted vocal tract shapes obtained. The results are shown in Figures 6 and 7, for the front and back vowels respectively.

These two figures show the calculated vocal tract shapes that are likely to have produced vowels having the proportional relationship of the formants entered. Each tract shape is represented by a series of sections of variable cross-section, shown on the vertical axis, and length, shown on the horizontal axis. The glottis is to the left of each display, with the termination at the open end of the glottis represented by the end of the line at the right.

The calculated vowel shapes for the front vowel pair in Figure 6 show that modeling a wide pharynx for the vowel /e/ and a narrow pharynx for the vowel /E/ approximates the measured formants for these vowels. A longer overall vocal tract length is also modeled for /e/ than for /E/. These two characteristics represent an ATR contrast very precisely. The tract shapes are not consistent with a simple difference of vowel height (in fact, the model shows a more constricted aperture in /e/ than in /E/.)

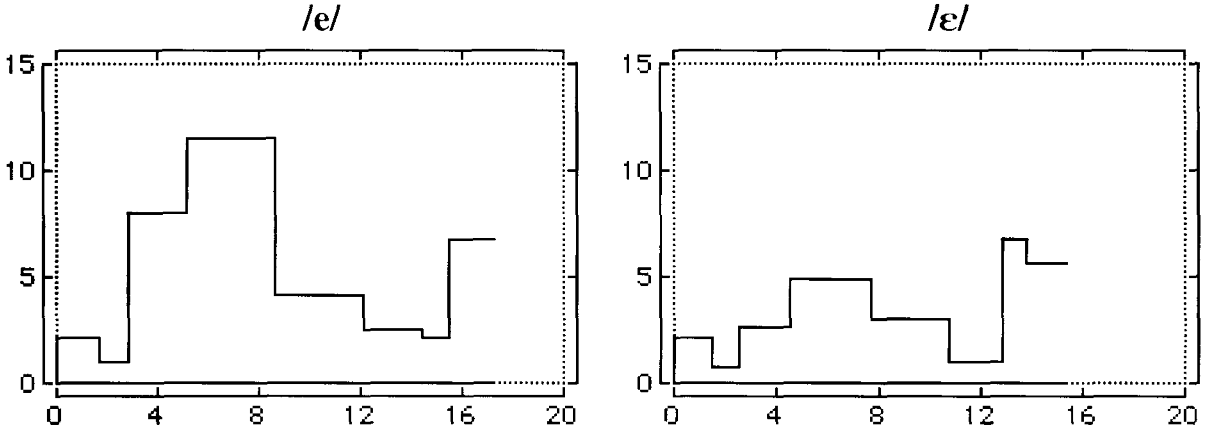


Figure 6. VTI models for the mid front vowel pair in Sele.

Figure 7, on the other hand, shows that the contrast between /o/ than for /ɔ/ was modeled as one primarily of tongue height. The constriction in the front part of the tract and at the lips is considerably narrower in the modeled /o/ than in /ɔ/. Both of these vowels are shown with relatively small pharyngeal cavities, and overall tract length does not differ.

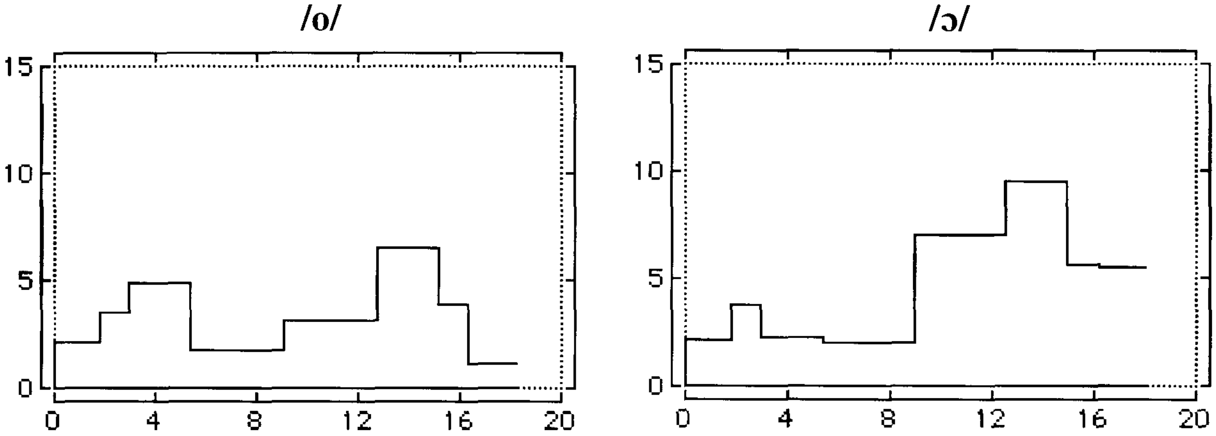


Figure 7. VTI models for the mid back vowel pair in Sele.

The modeling results continue to suggest the possibility that some part at least of the opposition between the members of the mid vowel pairs in Sele is borne by differences in the position of the tongue root. Obviously, however, because the results do not point unanimously in one direction, and the indications from the other comparisons only point weakly towards a conclusion, this question is not yet resolved.

It is tempting to think that Sele may be in the late stages of a transition away from the use of ATR in its vowel system, and the inconclusiveness of our analyses reflects this transitionality of the language. It is also quite probable that the failure to find a clear answer to the question of whether the vowels differ in ATR reflects continuing shortcomings in our overall understanding of the range of possible mechanisms involved in vowel production, and their acoustic consequences.

Acknowledgments

This research is supported by a grant from the National Science Foundation for "Phonetic Studies of Endangered Languages" to Peter Ladefoged and Ian Maddieson. It was also facilitated by support from the Faculty Exchange program between UCLA and the University of Ghana. We are most appreciative of this essential financial assistance. In Ghana, we owe a great debt of thanks to the Ghana Institute of Linguistics, Literacy and Bible Translation for moral and logistical support which made the fieldwork possible. Andy Ring and Sammy Ntumi were more than companions and hosts during this trip. The greatest thanks must go to the many speakers of Sele who willingly shared their knowledge of their language, especially Messrs Akiti, Raleigh and Atakora. Particular gratitude is owed to Mr. George Abodza who devoted many hours as translator and co-ordinator. At UCLA we thank Marco Baroni for generously sharing his data on Italian with us.

References

Allen, Christine. (1974). *Studies in the Phonology of Sɛlɛ, the Language of Santrokofi.* Unpublished M. A. Thesis, Department of Linguistics, University of Ghana, Legon.

Ford, Kevin C. (1973). "On the loss of cross-height vowel harmony." *Research Review* (Institute of African Studies, University of Ghana) 4: 50-80.

Ford, Kevin C. (1988). "Structural features of the Central Togo languages." In *The Languages of Ghana*, edited by M. E. Kropp Dakubu, Kegan Paul International for the International African Institute, London: 126-153.

Fujimura, Osamu and Lindqvist, Jan. (1971). "Sweep-tone measurements of vocal-tract characteristics." *Journal of the Acoustical Society of America* 49 (2), 541-58.

Fulop, Sean. (1996, this volume). An acoustic study of the tongue root contrast in Degema." UCLA Working Papers in Phonetics 93: 13-46.

Funke, Emil. (1911). Die Santrokofi: Beitrag zur Kenntnis der Sprachen Togos. *Mitteilungen des Seminars für Orientalischen Sprachen* 19: 116-128.

Heine, Bernd. (1968). *Die Verbreitung und Gliederung der Togorestsprachen.* Dietrich Reimer, Berlin.

Hess, Susan. (1992). "Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan." *Journal of Phonetics* 20 (4), 475-92.

Jospa, P., Soquet, A. and Saerens, M. (1992). "Acoustical sensitivity functions and the control of the vocal tract model". In *Signal Processing VI: Theories and Applications*, J. Vanderwalle, R. Boite, M. Moonen & A. Oosterlinck (editors), Morgan Kaufmann Publishers: 319-327.

Kropp Dakubu, Mary E., editor (1988). *The Languages of Ghana.* Kegan Paul International for the International African Institute, London.

Ladefoged, Peter and Maddieson, Ian. (1996). *The Sounds of the World's Languages*, Blackwells, Oxford.

Lindau, Mona. (1975). "[Features] for Vowels." *UCLA Working Papers in Phonetics* 30, UCLA.

Maddieson, Ian. (1995). "Collapsing vowel harmony and doubly-articulated fricatives-- two myths about Avatime phonology." *UCLA Working Papers in Phonetics* 91, 67-84.

Mrayati, M., Carré, R. and Guèrin, B. (1988). "Distinctive regions and modes: a new theory of speech production," *Speech Communication*, 7: 257-286.

Ring, J. Andrew. (1987). *Planning for Literacy: A Sociolinguistic Survey of Multilingualism in Ghana.* Ph. D. dissertation, Georgetown University, Washington, D.C.

Schuh, Russell G. (1995). "Aspects of Avatime phonology." *Studies in African Linguistics* 24: 31-67.

Stewart, John M. (1989). "Kwa". In *The Niger-Congo Languages*, edited by J. Bendor-Samuel. University Press of America, Lanham MD: 216-245.

Appendix

Mean values of the first three formants for the vowels of fourteen Sele speakers. Speakers are identified by gender (M or F) and an arbitrary number.

Speaker & vowel		F1	F2	F3	**Speaker & vowel**		F1	F2	F3
SF1	n				**SF2**	n			
u	19	384	1274	2794	u	19	382	1185	2681
o	14	528	1175	2829	o	13	520	1205	2944
ɔ	12	804	1380	2599	ɔ	13	711	1321	2940
a	44	949	1693	2659	a	42	906	1558	2640
ɛ	21	659	2005	2727	ɛ	24	668	2031	2973
e	5	564	2231	2829	e	4	527	2060	2922
i	19	351	2426	3060	i	18	340	2390	2940
SF3					**SF4**				
u	18	311	1146	2860	u	20	362	998	2620
o	13	498	1178	2927	o	13	487	1110	2849
ɔ	11	725	1287	2950	ɔ	10	770	1357	2845
a	38	1132	1653	2737	a	39	1064	1646	2734
ɛ	23	681	2243	2894	ɛ	17	721	2122	2862
e	4	500	2298	2931	e	4	511	2330	2963
i	17	296	2619	3131	i	18	355	2598	3192
SF5					**SF6**				
u	16	394	1112	2455	u	16	307	1042	2668
o	13	525	1210	2533	o	12	486	1108	2608
ɔ	12	778	1448	2472	ɔ	11	732	1299	2756
a	41	975	1603	2482	a	42	910	1685	2776
ɛ	21	705	1996	2677	ɛ	21	715	2128	3031
e	3	501	2017	2729	e	4	518	2247	2912
i	15	375	2277	2876	i	16	342	2517	3213

SM1					**SM2**				
u	12	285	764	1850	u	19	289	853	2040
o	11	346	889	2253	o	13	374	813	2187
ɔ	12	478	1040	2322	ɔ	15	472	984	2216
a	29	833	1533	2540	a	46	839	1506	2364
ɛ	12	515	1782	2653	ɛ	23	449	1931	2540
e	4	353	1953	2733	e	4	353	1894	2495
i	8	261	2113	2998	i	18	294	2127	3024
SM3					**SM4**				
u	11	280	935	2028	u	17	291	867	2091
o	14	376	1026	2096	o	14	343	830	2424
ɔ	14	483	1061	2169	ɔ	15	513	1014	2417
a	45	792	1461	2318	a	47	897	1523	2515
ɛ	24	483	1932	2624	ɛ	23	510	2035	2631
e	4	325	1931	2614	e	4	344	2119	2706
i	17	289	2043	2912	i	20	292	2159	2831
SM5					**SM6**				
u	11	308	929	1981	u	20	332	870	1969
o	14	362	948	2256	o	13	380	904	2046
ɔ	12	542	1106	2225	ɔ	14	519	969	2305
a	43	829	1533	2520	a	45	763	1412	2491
ɛ	24	489	1841	2697	ɛ	24	495	1844	2447
e	4	358	1989	2871	e	4	413	1715	2366
i	17	297	1982	3070	i	20	317	1813	2743
SM7					**SM8**				
u	16	310	933	2058	u	15	304	946	2046
o	13	368	908	2177	o	14	388	948	2255
ɔ	12	539	989	2230	ɔ	13	557	1113	2492
a	48	727	1431	2308	a	44	795	1417	2470
ɛ	24	471	1990	2645	ɛ	23	540	1894	2786
e	4	362	1926	2568	e	6	395	1778	2617
i	18	310	1915	2865	i	17	285	2047	3208

The Phonetic Structures of Hupa

Matthew Gordon

1. Introduction

Hupa is an Athabaskan language spoken on the Hoopa Valley Indian reservation in northwest California bordered on the north by the Yurok Indian reservation and approximately 30 miles northwest of Eureka, California. The language name is typically spelled with a "u" in the first syllable, while the reservation and the town of the same name (located on the reservation) are spelled with an "oo". The name "Hupa" is actually borrowed from a Yurok word **/hupʼoː/**which is equivalent in meaning to the Hupa term **/naːtɪnɪxʷ/** (pronounced **naːˈtʰɪnuxʷ**), which translates "where the trails (lead) back." (Golla 1970:9). Hupa belongs to the California subgroup of the Pacific Coast branch of Athabaskan (Hoijer 1960), which includes languages formerly spoken throughout northwest California and southwest Oregon. Most Pacific Coast Athabaskan languages, with the exception of Hupa, are either extinct or spoken by only 164icult to determine the exact number of Hupa speakers. Based on extrapolations from a 12% sample of the population, the 1990 census estimates there to have been 93 speakers at the time of the census (IJAL 1995). The actual number of current speakers is probably smaller than this figure would indicate; virtually all speakers are over the age of seventy.

Early work on the Hupa language was conducted by Pliny Earle Goddard and published in a number of papers at the turn of the century. Goddard (1907) includes a brief description of the sounds of Hupa and includes static palatography and kymographic data. Goddard (1905) is a description of the morphology, while Goddard (1904) is a set of texts . Later work by Goddard (1928) presents pitch measurements which demonstrate that Hupa, like other members of the Pacific Coast subgroup of Athabaskan, but unlike members of the Northern and Southern branches of Athabaskan, is not a tonal language. More recently, Golla (1970) provides a comprehensive analysis of the phonology and syntax of Hupa, as well as a detailed qualitative description of the phonetics. A qualitative description of Hupa phonetics also appears in Woodward (1964). Golla (1964) is an etymological study of Hupa noun stems. Golla (1977) contains a discussion of verb stem morphology, while Golla (1985) is a condensed pedagogical grammar of Hupa.

This paper will provide a general phonetic description of the consonants and vowels of Hupa, and will also examine a number of salient phonetic properties found in Hupa. Hupa is of interest from a phonetic and historical/comparative standpoint for a number of reasons. Hupa is virtually the only extant Pacific Coast Athabaskan language, and as such, provides valuable material for comparing the Pacific Coast branch with the northern and southern branches of Athabaskan. In certain respects, Hupa is the most conservative of the Athabaskan languages, preserving a number of phonemic contrasts which have been lost in other languages. At the same time, Hupa possesses several interesting phonetic contrasts which are later innovations. Hupa also possesses an interesting process of laryngeal spreading with creates surface contrasts in the relative timing of laryngeal and supralaryngeal features (Golla 1977).

The present paper is based on fieldwork carried out during the month of September, 1995 in Hoopa, California. Data is based on audio recordings made of three speakers, two men and one woman. Data was digitized on Kay CSL at a sampling rate of 12kHz.

2. Consonants

Like other Athabaskan languages, Hupa is characterized by a large number of consonants, which are presented below in table 1. [̹] is used to indicate a more rounded segment.

Table 1: Consonants of Hupa

	bilabial	denti-alveolar	palato-alveolar	palatal-ized velar	velar	uvular	labial-velar	glottal
unaspirated stops	**(p)**	**t**		**kʲ**		**q**		**ʔ**
aspirated stops		**tʰ**		**kʲʰ**				
ejective stops		**t’**		**kʲ’**		**q’**		
unasp. affricates		**t͡s**	**ṭ͡ʃ**					
asp. affricates		**t͡sʰ**	**ṭ͡ʃʷʰ**					
eject. affricates		**t͡s’ t͡ɬ’**	**ṭ͡ʃ’ (ṭ͡ʃʷ’)**					
fricatives		**s ɬ**	**(ʃ)**		**x xʷ x̹ʷ**			**h**
glottalized fric.		**(s̰)**						
nasals	**m**	**n**	**ɲ**		**ŋ**			
glottalized nasal		**n̰**			**ŋ̰**			
liquids		**l**						
glottalized liquid		**l̰**						
voiceless liquid		**(l̥)**						
glides		**y**					**w**	
glottalized glides		**y̰**					**w̰**	

Unaspirated, aspirated and ejective non-affricated stops contrast at the denti-alveolar, velar and uvular points of articulation, with the exception of the lack of a contrast between aspirated and unaspirated uvulars. Tracings of palatography from Goddard (1907) in figure 1 show that the denti-alveolar stops have a broad area of contact between the tongue and the roof of the mouth which extends from the back of the alveolar ridge to the front of the teeth. The area of contact for the palatal affricates is narrower than the contact area for the denti-alveolar stops, but does not extend any farther back on the hard palate, suggesting that the palato-alveolar affricates in table 1 might more accurately be described as alveolar affricates. However, it should be mentioned that the palato-alveolar affricate appears in a front vowel environment in figure 1, which could be partially responsible for its fronted articulation. Given this possible confound, they will continue to be referred to as palato-alveolar affricates in this paper. The contact for the sounds labeled palato-alveolar affricates does not extend forward onto the teeth, unlike for the denti-alveolar stops. The velar stops in Hupa are palatalized; the area of contact extends forward from the primary velar constriction along both sides of the hard palate and molars.

Unlike the palatalized velar in Russian which is realized with a much more distinct glide into the following vowel than out of the preceding vowel (Ladefoged and Maddieson 1996:364), the Hupa palatalized velar is realized with a pronounced glide both going into the oral closure as well as coming out of the closure. For example, phonemic **/naːkʲ’ɪne/** ‘mountain quail’ is pronounced as **[naːʲkʲ’ɪne]** with a palatal onglide both preceding and following the closure for the palatalized velar. This palatal onglide is illustrated in the spectrogram in figure 2.

/t/ in /**taw**/ 'no'   /t͡ʃ'/ in /**t͡ʃ'e**/ 'out'   /**kʲʰ**/ in /**kʲʰaʔ**/ 'dress'

Figure 1: Tracings of palatography from Goddard (1907) illustrating the denti-alveolar, palato-alveolar and palatalized velar stops in Hupa.

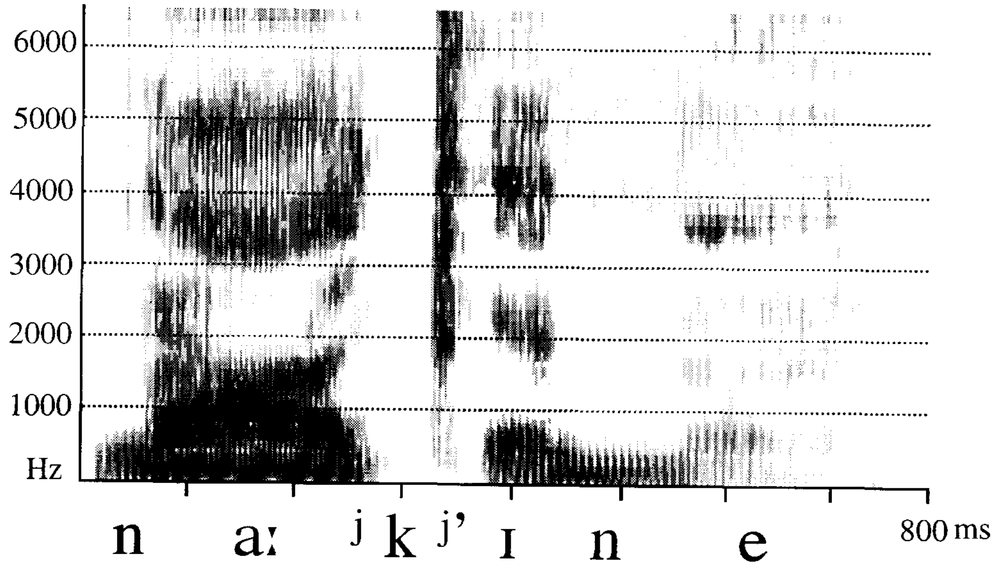


Figure 2: Spectrogram illustrating the palatalized velar in /**naːkʲ'ɪne**/ 'mountain quail' (speaker M1)

The palatal off-glide following the closure can be more clearly seen in the spectrogram in figure 3.

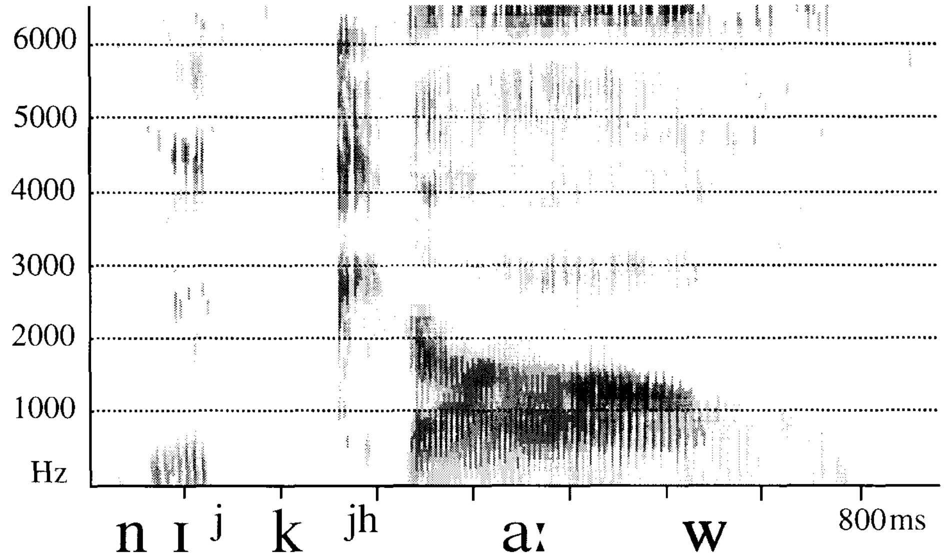


Figure 3: Spectrogram illustrating the palatalized velar in **/niʲkʲʰaːw/** 'it is big' (speaker M1)

The oral closure of palatalized velar stops is occasionally lost leaving only a glottal stop with a palatal on- and off-glide in its place. A spectrogram illustrating this phenomenon is found in figure 28 in section 2.3.

A three-way laryngeal contrast is also found in the affricated stops. The aspirated palato-alveolar affricate is labialized, and the only affricate with a lateral release is an ejective with a denti-alveolar closure. As in other Athabaskan languages, unaspirated and glottalized stops (affricated and non-affricated), but not aspirated stops, occur in stem-final position. Oral stops and also glottal stop are released in utterance final position (see figures 19, 22 and 23). Voice-onset-time and closure duration of the stops, both affricated and unaffricated are discussed in sections 2.1 and 2.2, respectively.

In addition to the stops appearing in the table 1 there is a fourth set of unaffricated stops which are velar but not palatalized. These, however, occur only in certain special speech styles (see Golla 1970: 27-28) and are not phonologically conditioned. They were not produced by the speakers examined for this paper. **/p/** is confined to loan words, while **/ʃ/** is limited to a few exclamations. The rounded **/t͡ʃʷ'/** is attested only in certain archaic forms, typically as a variant of the non-labialized palato-alveolar ejective **/t͡ʃ'/** (Golla 1970:29). The glottalized fricative **/s̰/** is attested in a single word **/xʷɪs̰/** 'mother's brother' which did not appear to be distinguish phonetically from non-glottalized **/s/** in the data collected.

Note that the segments classified as glottalized sonorants in table 1 could alternatively be treated as clusters of glottal stop plus sonorant (not necessarily in that order; see below). The decision to treat them as single segments rather than clusters is based on two distributional facts. First, while tautosyllabic biconsonantal clusters are fairly common, the only candidate clusters consisting of three tautosyllabic consonants would be clusters consisting of glottal stop plus sonorant plus obstruent (Woodward 1964). If glottalized sonorants are treated as single segments rather than clusters, the maximum number of consonants in a tautosyllabic cluster is limited to two. Along similar lines, tautomorphemic clusters are exceedingly rare in stem-final position in Athabaskan (Leer 1979, Krauss and Leer 1981), but glottalized sonorants are more common in this position. The analysis of glottalized sonorants as single segments is also consistent with the phonetic observation that glottalization associated with sonorants is often realized as creakiness rather than a full glottal stop, especially in the case of preglottalized sonorants. The contrast between modal voiced and glottalized sonorants is limited to stem-final position, where the ordering of glottalization and supralaryngeal features in glottalized sonorants is contrastive. In the text, this difference in timing will be indicated by placing the glottalization diacritic before the sonorant in the case of preglottalized sonorants and after the sonorant in the case of postglottalized sonorants. In stem-final position, Hupa also contrasts pre-glottalized stops with stops which are both pre- and post-glottalized, and preaspirated stops with unaspirated stops. These laryngeal contrasts are considered in more detail in section 2.3.

In addition to the glottalized sonorants, there is a voiceless allophone of **/l/** which occurs after **/h/** in the word **/tʃ’ahl̥/** ‘frog’. Spectrograms illustrating the voiceless lateral approximant as produced by speakers M1 and F1 appear in figures 4 and 5.

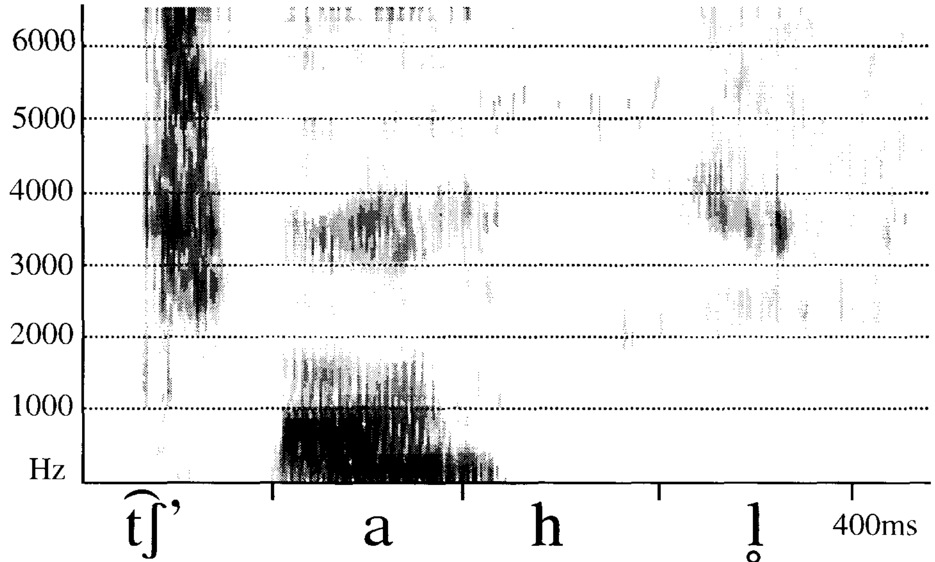


Figure 4: The voiceless lateral approximant in the word **/tʃ’ahl̥/** ‘frog’ (speaker M1)

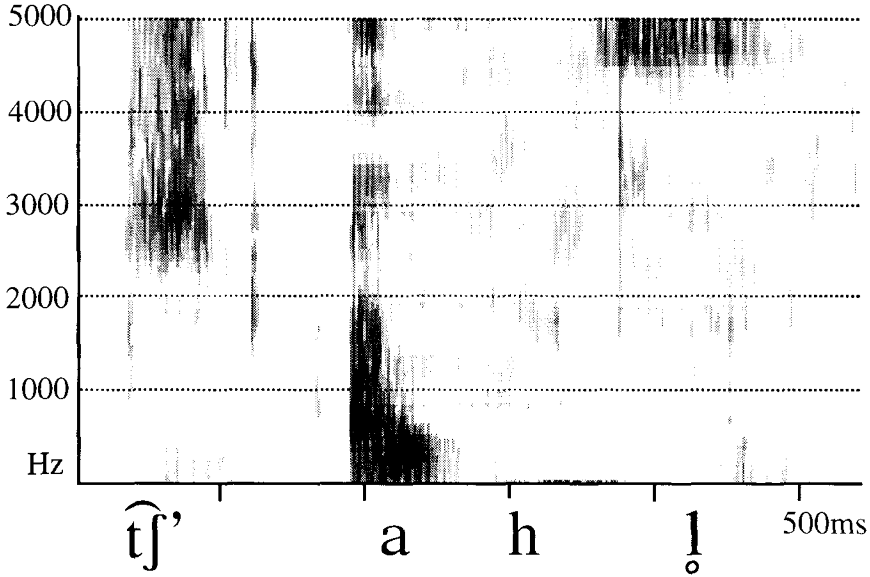


Figure 5: The voiceless lateral approximant in the word **/tʃ’ahl̥/** ‘frog’ (speaker F1)

This voiceless **/l̥/** is involved in near minimal contrasts with the voiceless lateral fricative after **/h/**: e.g. **/nɪtʃ’ahl̥/** ‘your frog’ vs. **/nɪqahɬ/** ‘you go along’. Spectrograms illustrating the voiceless lateral fricative in word-final position as produced by speakers M1 and F1, respectively, appear in figure 6 and 7.

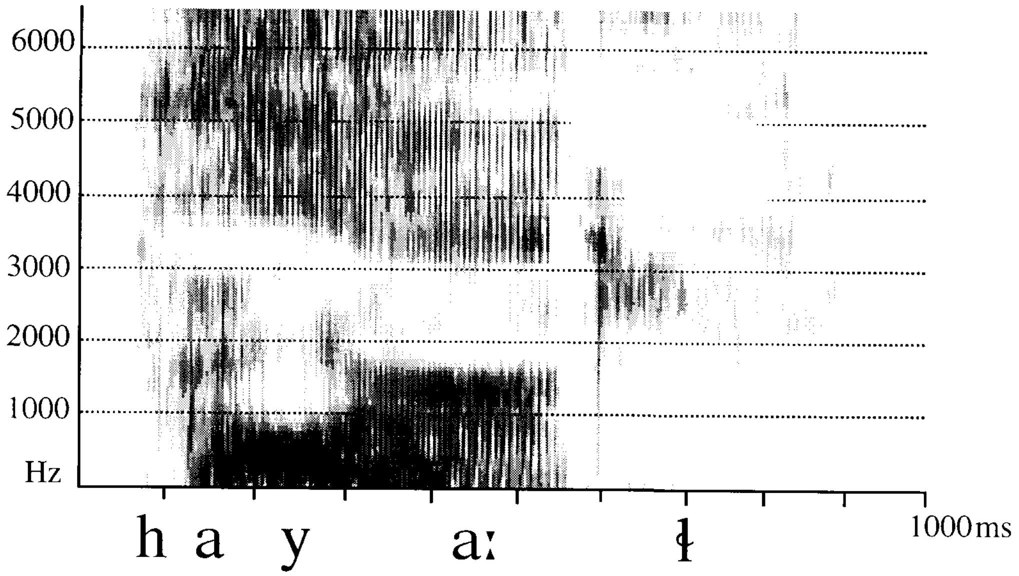


Figure 6: The voiceless lateral fricative in the word **/hayaːɬ/** ‘and then’ (speaker M1)

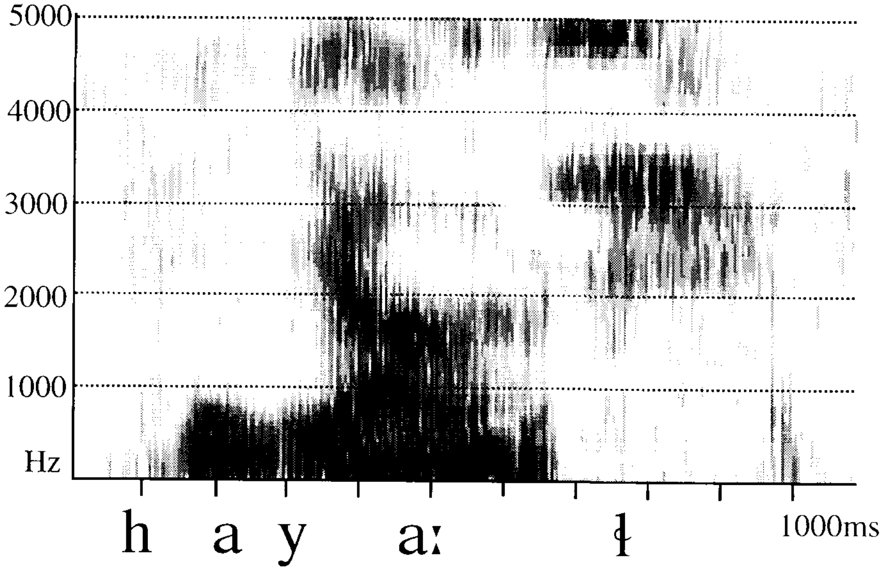


Figure 7: The voiceless lateral fricative in the word **/hayaːɬ/** ‘and then’ (speaker F1)

The voiceless lateral approximant is shorter in duration than the lateral fricative and is characterized by more intense noise. The lateral fricative also appears to involve a higher tongue body position than the voiceless lateral approximant. This difference manifests itself acoustically by an increase in the amplitude of noise at lower frequencies for the lateral fricative relative to the lateral approximant, a difference which can be seen in the spectrograms of speaker F1.

One of the more interesting contrasts found in Hupa is the contrast between the segments identified in table 1 with the symbols /xʷ/ and /x̹ʷ/, where the diacritic under the second segment indicates a greater amount of lip rounding than for the first one. These two segments are illustrated in the word in figure 8 as spoken by speaker M1. As a point of reference, an example of a non-labialized velar fricative appears in figure 9.

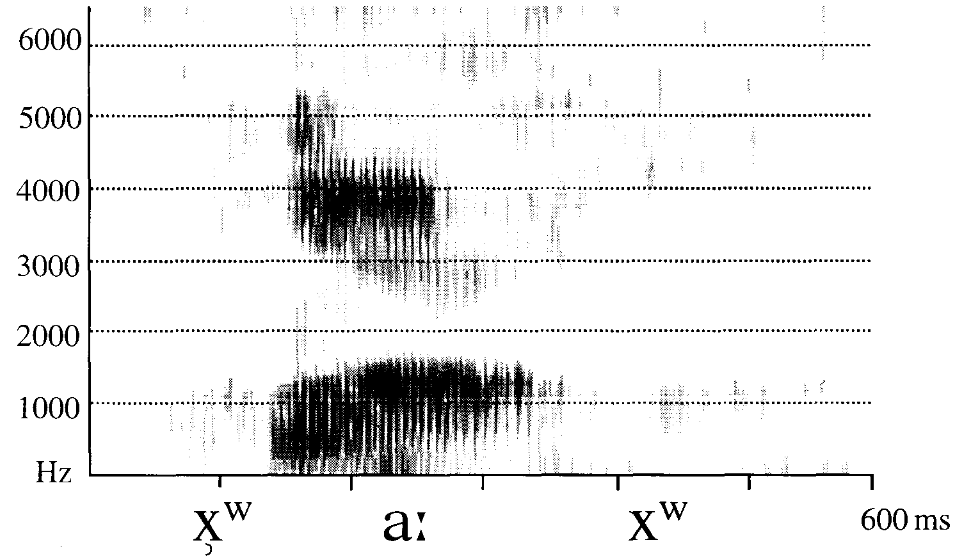


Figure 8: Spectrogram illustrating the segments /**x̹ʷ**/ and /**xʷ**/ in the word /**x̹ʷaːxʷ**/ 'his brother's son' (speaker M1)

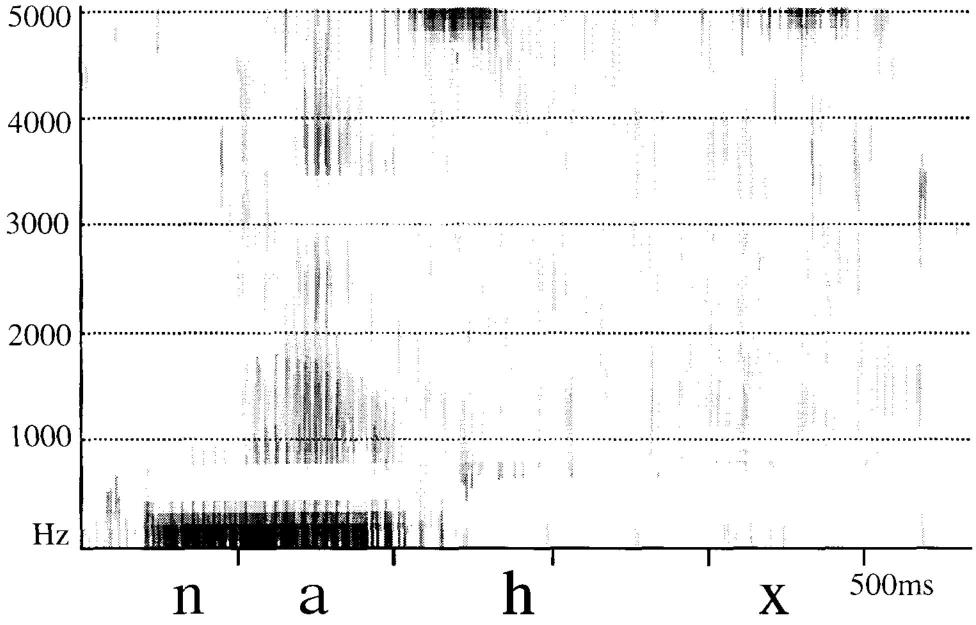


Figure 9: Spectrogram illustrating the non-labialized velar fricative in the word /**nahx**/ 'two' (speaker M1)

Figures 10 and 11 illustrate further examples of the two sounds produced by speaker F1.

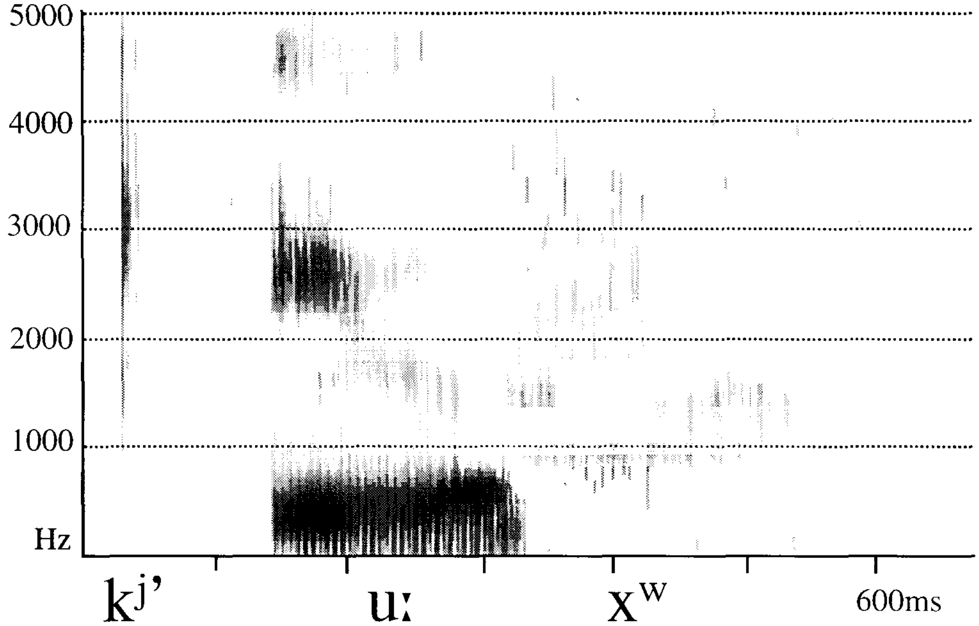


Figure 10: Spectrogram illustrating the segment **/xʷ/** in **/kʲ’uːxʷ/** ‘lightning’ (speaker F1)

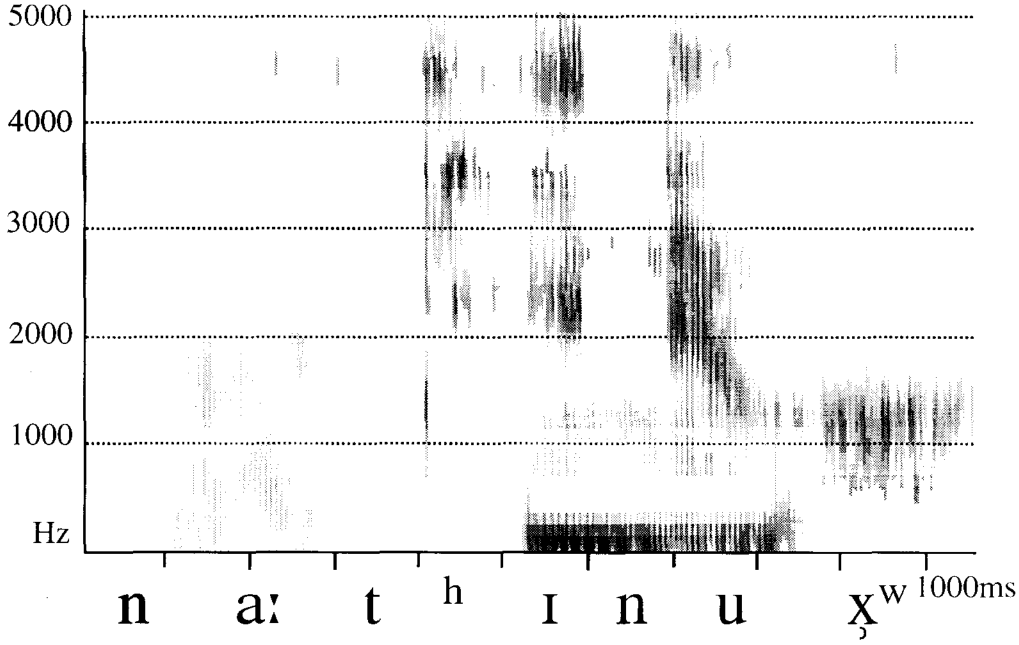


Figure 11: Spectrogram illustrating the segment **/x̹ʷ/** in **/naːtʰɪnux̹ʷ/** ‘Hupa (where the trail (leads) back’ (speaker F1)

The primary articulatory constriction for both segments seems to be in the velar region and both sounds have lip rounding, thus both may be classified as labialized velar fricatives. Differences in the acoustic properties of these two segments suggest, however, that one of the labialized velar fricatives is characterized by more lip rounding than the other. As can be seen in figures 8, 10, and 11, the principal acoustic difference between these two sounds seems to lie in the frequency at which the loudest frication noise is found. The center of noise is slightly higher for the segment transcribed as **/xʷ/** than for **/x̹ʷ/**. The difference in degree of lip rounding between the two labialized velar fricatives can also be seen in the vowels adjacent to the fricatives in spectrogram 8. Formant frequencies are lower at the beginning of the vowel adjacent to the more rounded velar fricative than at the end of the vowel adjacent to the less rounded velar fricative. There is also slightly more noise at higher frequencies during the less rounded labialized velar fricative due to increased damping of higher frequencies attributed to the greater degree of lip rounding. This latter difference can be seen most clearly in comparing figures 10 and 11. Both the lowered center of frication in terms of frequency and the lesser amount of noise at higher frequencies suggest a greater degree of lip rounding in the segment marked as /**x̹ʷ**/. In comparison to both of the labialized velar fricatives, the non-labialized velar fricative in figure 9 has more noise throughout the spectrum. The area of most intense noise is also at a higher frequency in the labialized velar fricatives than the locus of noise in the labialized velar fricatives. Phonologically, both of the labialized velar fricatives function as rounded segments (Golla 1970). When in coda position of the syllable, both sounds trigger rounding and backing of a preceding short high front vowel. Additionally, the sequence /**x̹ʷɪ**/ may optionally be realized as /**xo**/. (/**x**/ and /**x̹ʷ**/ do not contrast before /**o**/.) This option of realizing rounding on the vowel rather than the fricative is not available, however, for the less rounded labialized velar fricative in the sequence /**xʷɪ**/.

Historically, the more rounded labialized velar fricative, which occurs in certain adverbial suffixes and as third person singular possessive and subject prefixes, appears to be descend from proto-Athabaskan ***xʷ** (Hoijer 1971). The less rounded labialized velar fricative descends from two sources (Hoijer 1960): a front (palatalized) velar, transcribed by Hoijer as ***xʸ**, and a palato-alveolar fricative ***ʃ**. The ***ʃ** has been preserved elsewhere in Pacific Coast Athabaskan except for the now extinct language Mattole where, interestingly, its reflex is a velar fricative /**x**/ (Li 1930, Hoijer 1960). It is interesting to note that there is a general tendency for palato-alveolar obstruents to undergo labialization in Hupa. The aspirated affricate ***tʃʷʰ** of proto-Athabaskan has also been labialized in Hupa, as has the ejective affricate ***tʃ'** sporadically.

Another interesting phonetic aspect of Hupa is the phonetic realization of the segments transcribed as velar nasals in table 1. This contrast is limited to stem-final position. Often there is only an incomplete contact between the back of the tongue and the roof of the mouth during the production of the velar nasal. Often the constriction is only approximated at the same time the velo-pharyngeal port is open. The resulting sound is a nasalized velar approximant which is found in free variation with the realization of the same phoneme as a velar nasal stop. In both realizations, the vowel preceding the nasal/nasalized segment is also nasalized, though the degree of nasalization is not as great as in many languages with contrastively nasalized vowels. Both the velar nasal and nasalized velar approximant realizations are also found for the glottalized counterparts of the velar nasal. The spectrograms in figures 12 and 13 illustrate the two realizations of the glottalized velar nasal phoneme as produced by the same speaker, F1.

In the nasalized velar approximant in figure 12, there is a greater amount of energy at higher frequencies than in the velar nasal stop in figure 13. In both tokens, the nasal/nasalized segment is abruptly truncated by the glottal stop with some glottalization of the preceding segment. It is interesting to note that the velar nasal in Hupa is the reflex of the proto-Athabaskan segment reconstructed by Krauss and Leer (1981) as a palatal nasalized approximant (transcribed by them as /ỹ/) in stem-final position. Though the original palatal articulation has been backed to a velar one in Hupa, its approximant-like character is optionally preserved. The realization of this segment as a nasalized approximant with concomitant nasalization of the preceding vowel has been preserved in a few northern Athabaskan languages: Kutchin, Han, Upper Tanana and Tanacross (Krauss and Leer 1981: 33-37). In these northern Athabaskan languages, however, the nasalized approximant is palatal rather than velar.

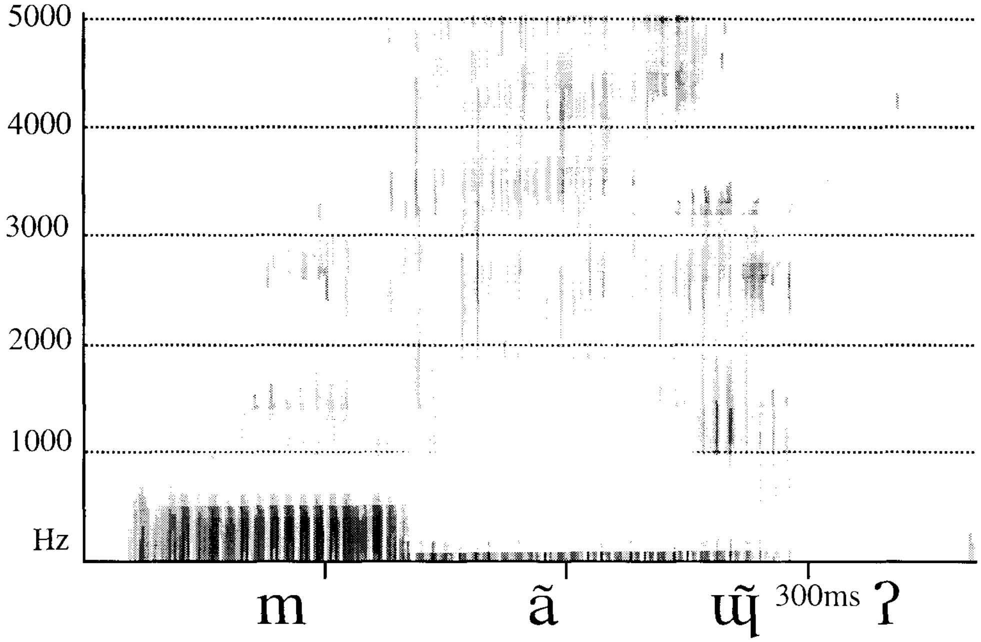


Figure 12: Spectrogram illustrating the nasalized velar approximant in the word **/maŋ/**'fly' (speaker F1)

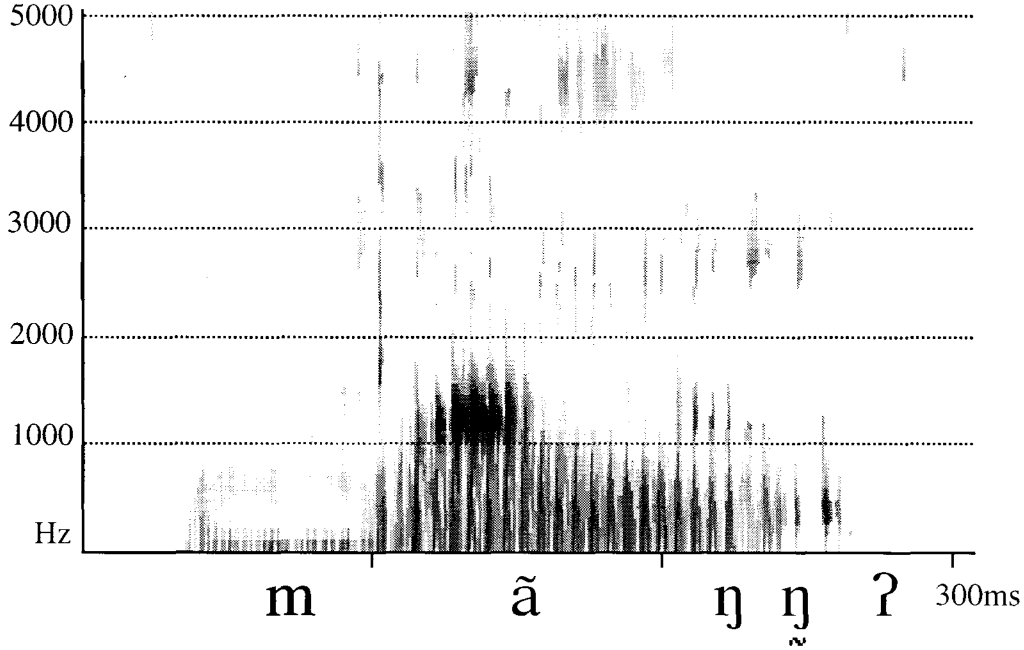


Figure 13: Spectrogram illustrating the velar nasal stop in the word **/maŋ/**'fly' (speaker F1)

Hupa contrasts open syllables, syllables closed by **/h/** and syllables closed by a glottal stop. Example spectrograms illustrating this three-way contrast are found in figures 14, 15 and 16.

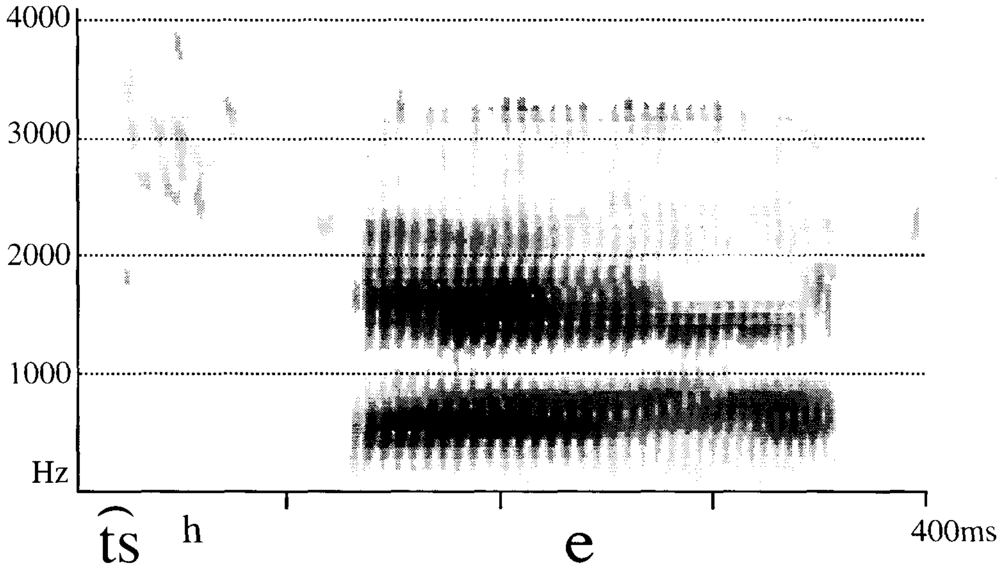


Figure 14: Spectrogram illustrating an open syllable in the word **/t͡sʰe/** ‘stone’ (speaker M2)

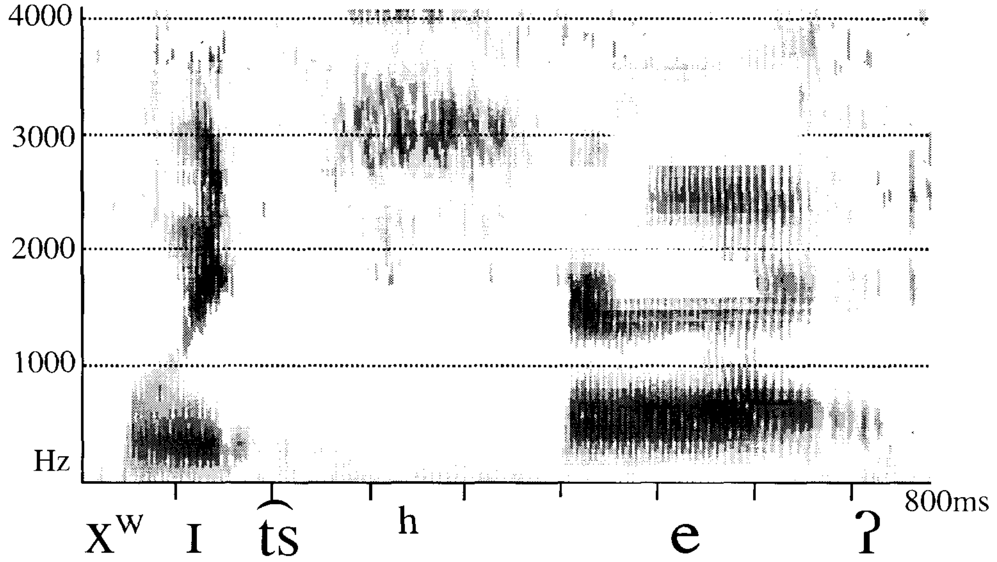


Figure 15: Spectrogram illustrating a syllable closed by a glottal stop in the word **/xʷɪt͡sʰeʔ/** ‘my daughter’ (speaker M2)

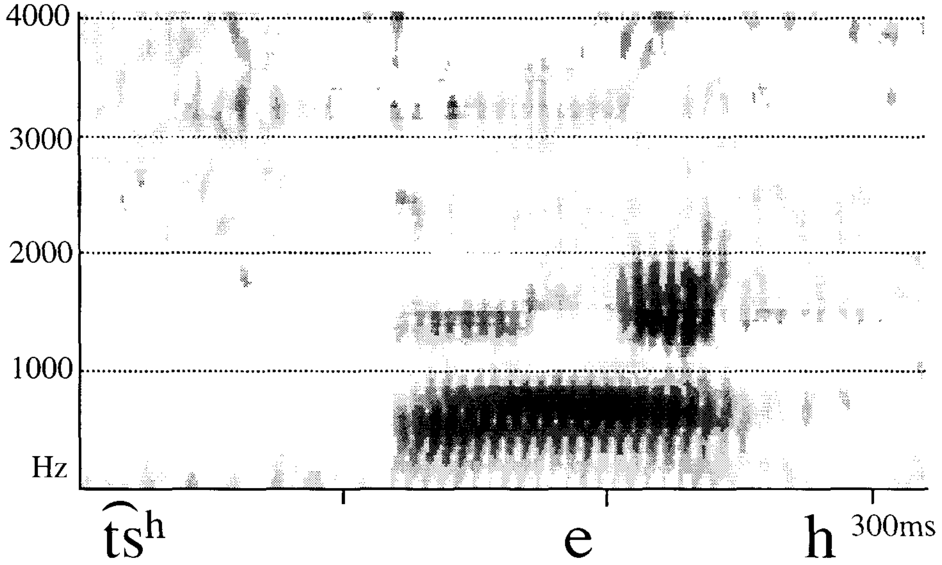


Figure 16: Spectrogram illustrating a syllable closed by an /**h**/ in the word /**t͡sʰeh**/ 'flatulence' (speaker M2)

2.1. Voice-onset-time

VOT values for both the stops and affricates were measured using a waveform display in conjunction with a spectrogram. Measurements were taken of the period from closure release to the onset of periodic voicing preceding both the high front vowels /**e**/ and /**ɪ**/ and the low vowels /**a**/ and /**aː**/. VOT values for all three environments were collapsed after Fisher's PLSD tests revealed that VOT did not differ significantly as a function of the following vowel. VOT values for each of the consonants, pooled together over the three speakers are presented in table 2, along with the number of tokens of each consonant and standard deviations.

Table 2: Mean VOT values for the stops in Hupa

Consonant	**n**	**Mean (ms)**	**Std.dev.**
p	6	10.8	1.4
t	61	15.5	5.2
kʲ	7	44.0	9.0
q	9	27.0	15.5
tʰ	35	82.2	23.2
kʲʰ	28	84.2	35.1
t'	15	92.8	28.4
kʲ'	47	80.2	45.2
q'	27	89.4	40.3
t͡s	11	70.8	19.0
t͡sʰ	9	152.3	39.2
t͡ʃʷʰ	12	167.1	51.9
t͡s'	4	120.3	17.7
t͡ʃ'	7	89.6	30.0
t͡ɬ'	13	102.0	26.9

As expected, VOT values for the unaspirated stops are much smaller than VOT values for corresponding aspirated stops at the same place of articulation. This pattern holds true of both the affricated and unaffricated stops. VOT values for the unaffricated aspirated and ejective stops are of a similar magnitude, hovering around 80-90ms. Spectrograms illustrating the unaspirated, aspirated, and ejective stops at the denti-alveolar place of articulation appear in figures 17, 18 and 19, respectively.

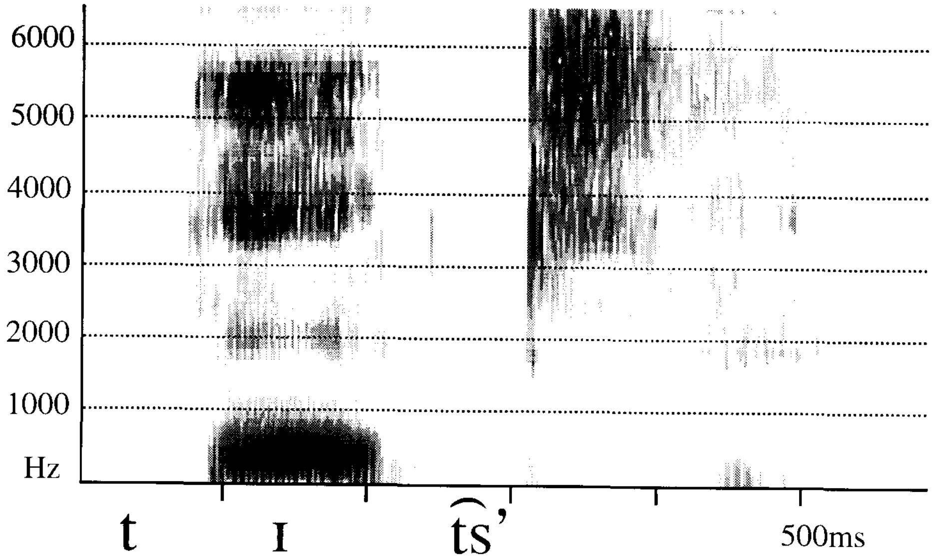


Figure 17: Spectrogram illustrating the unaspirated denti-alveolar stop **/t/** in the word **/tɪts'/** 'valley quail' (speaker M1)

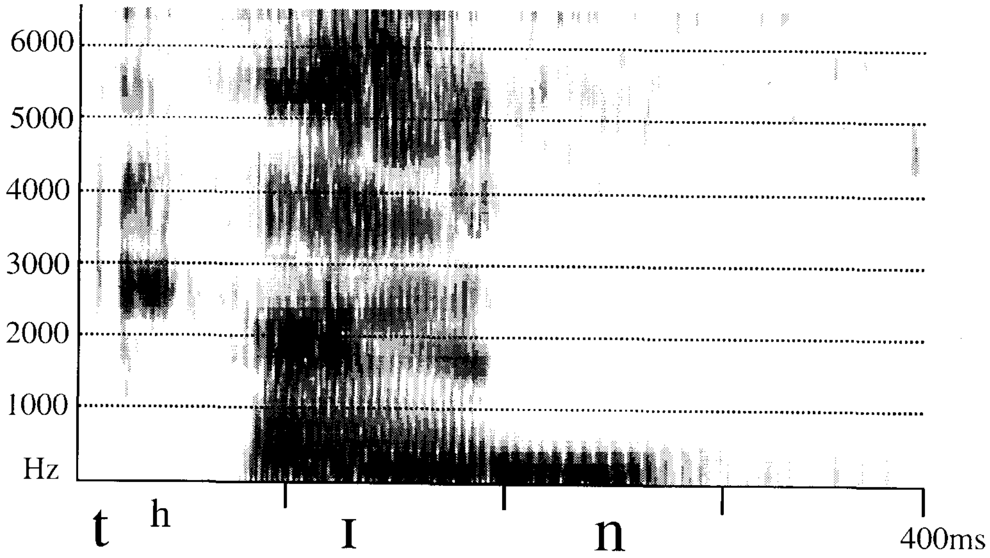


Figure 18: Spectrogram illustrating the aspirated denti-alveolar stop **/tʰ/** in the word **/tʰɪn/** 'trail' (speaker M1)

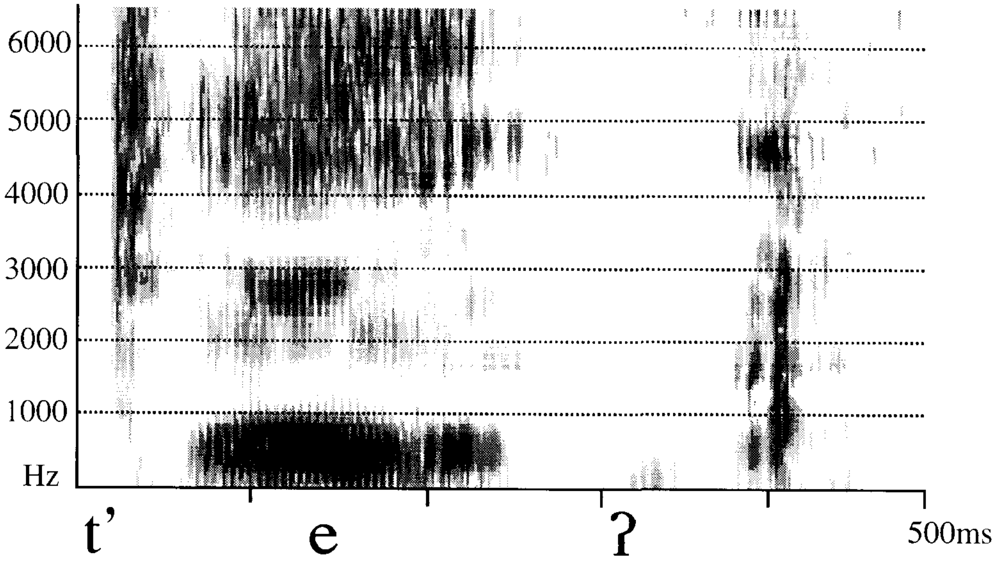


Figure 19: Spectrogram illustrating the ejective denti-alveolar stop **/t'/** in the word **/t'eʔ/** 'blanket' (speaker M1)

The only (strictly) manner of articulation contrast for the affricated stops exists in the palato-alveolar series, where VOT values for the unaspirated stops are, as expected, smaller than those for the aspirated stops. Also, unlike the unaffricated stops, the denti-alveolar ejective affricate **/t͡s'/** has a smaller mean VOT value than the aspirated denti-alveolar affricate **/t͡sʰ/**. Interestingly, the unaffricated aspirated stops in Hupa are marked by much shorter voice-onset-times than the corresponding stops of Navajo, a southern Athabaskan language (McDonough and Ladefoged 1993). For example, **/tʰ/** has a mean VOT of 130ms in Navajo compared to 82ms for **/tʰ/** in Hupa, both values being longer than the aspirated stops in English, Cantonese and Eastern Armenian measured by Lisker and Abramson (1964), but similar to those of Korean in published in the same study. VOT values for the ejective stops of Hupa are quite similar to values found in Navajo, and also far longer than those found in languages such as Hausa (Lindau 1984).

The correlation between degree of backness and VOT duration, found in many other languages of the world (Maddieson 1996), is not robustly instantiated across all manner of articulations, except for the series of unaspirated stops, where the bilabials are characterized by a shorter VOT than the denti-alveolars which in turn have shorter VOT values than the palatalized velars and uvulars. The palatalized velars have longer VOT values than the uvulars presumably due to the higher tongue body position during the palatalized velar. The higher tongue body position increases the amount of time needed for the difference in subglottal and supraglottal pressure to reach levels which allow for the transglottal air flow necessary for voicing. It should be pointed out, however, that none of the differences in VOT between stops at different places of articulation reach statistical significance at the p>.05 level.

2.2. Closure Duration

Closure duration was measured for the oral stops in intervocalic position using a waveform in conjunction with a spectrogram. The beginning of the closure was measured at the offset of the second formant of the preceding vowel. Measurements of closure duration pooled together over three speakers appear in table 3.

Table 3: Mean closure duration for the stops in Hupa

Consonant	n	Mean (ms)	Std.dev.
p	8	152.5	49.9
t	22	142.3	59.2
kʲ	3	75.4	5.0
q	3	130.6	34.7
tʰ	18	124.5	28.7
kʲʰ	18	101.4	25.1
t'	16	135.2	22.9
kʲ'	13	97.0	46.3
q'	8	49.0	17.3
t͡s	2	63.2	14.5
t͡sʰ	10	74.4	14.6
t͡ʃʷʰ	3	62.7	10.0
t͡s'	2	63.2	14.5
t͡ʃ'	7	67.3	20.2
t͡ɬ'	11	67.0	34.6

As in Navajo (McDonough and Ladefoged 1993), unaffricated stops in Hupa are characterized by longer closure durations than affricated stops, a result which is significant as revealed by an ANOVA factorial analysis at the $p<.0001$. VOT values do not differ significantly, however, for stops with different manners of articulation. Furthermore, the cross-linguistic tendency for closure duration to be correlated with the degree of frontness of the oral constriction is also found to some extent in Hupa, for at least the two frontmost places or articulations in the unaffricated stops. Within the set of unaffricated stops with the same laryngeal setting, bilabials have longer closure durations than denti-alveolars, although this difference is statistically insignificant in the case of the bilabial vs. denti-alveolar contrast. However, the unaspirated and the ejective stops display two patterns at the two backmost places of articulations. In the set of unaffricated unaspirated stops, the closure duration for the uvular /**q**/ is longer than the duration for the palatalized velar /**kʲ**/. In contrast, in the unaffricated ejective stop series, the palatalized velars have a longer average closure duration than the uvulars.

2.3. The realization of laryngeal features

One of the most interesting phonetic properties of Hupa is the timing of laryngeal features relative to supralaryngeal features in glottalized consonants and unaspirated stops in stem-final position. In stem-final position, preaspirated obstruents (indicated with an /**h**/) contrast with unaspirated stops on the surface: e.g. /**neːs**/ 'long' vs. /**ʔehs**/ 'fish dam.' Ejective stops contrast with pre-glottalized ejectives (indicated with a creaky diacritic under the latter part of the preceding vowel) : e.g. /**tʰaːq'**/ 'three' vs. /**xḛːq'**/ 'spit'. The preaspirated and the preglottalized obstruents are only found following long vowels. Example spectrograms illustrating the preaspirated and pre-glottalized obstruents appear in figures 20 and 21. As a point of a comparison, a spectrogram showing an ejective without pre-glottalization appears in figure 22.

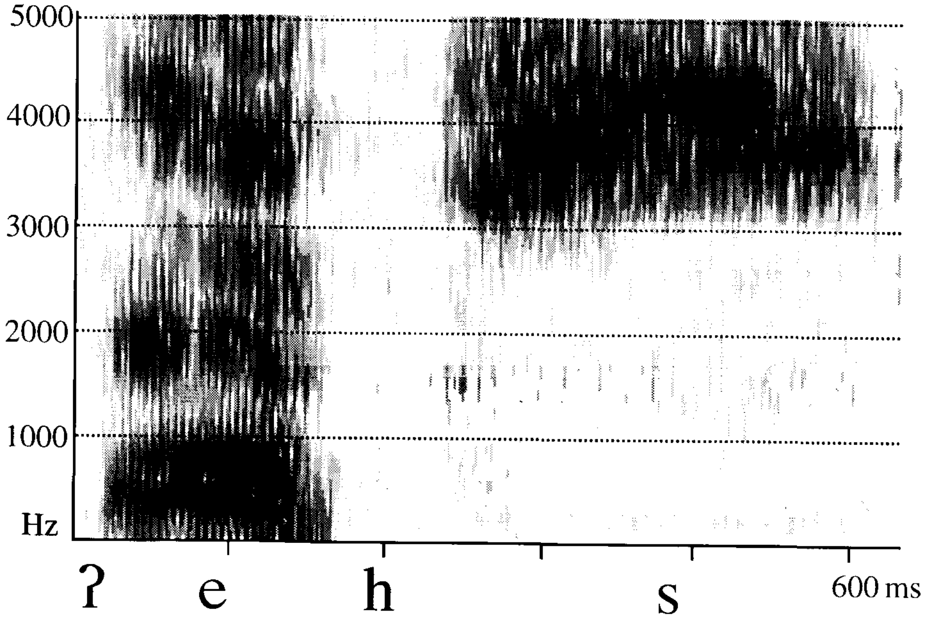


Figure 20: Spectrogram illustrating a preaspirated fricative in the word /**ʔeː̥s**/ ‘fish dam’ (speaker M1)

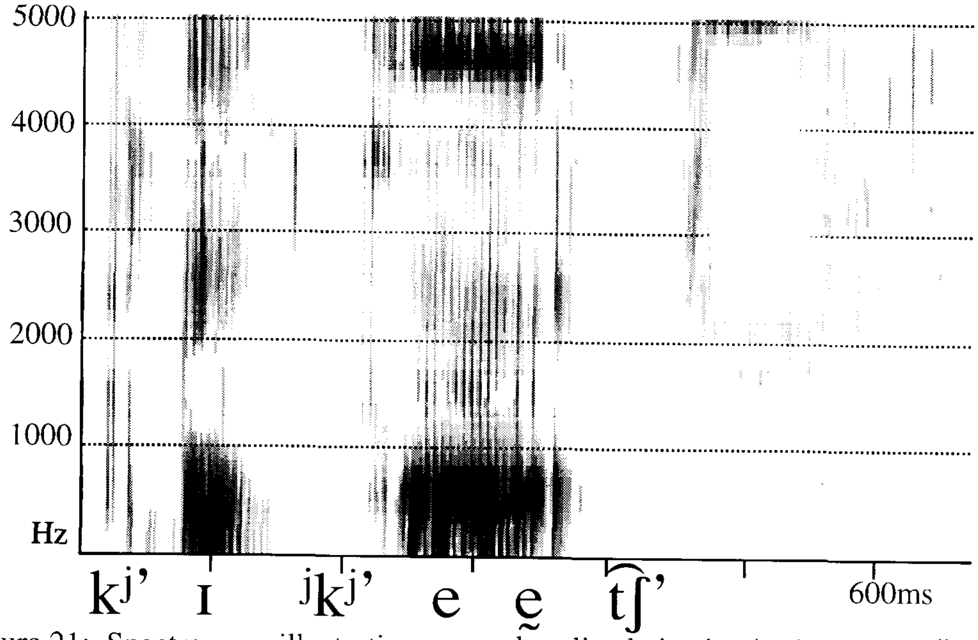


Figure 21: Spectrogram illustrating a pre-glottalized ejective in the word /**kʲ’ɪkʲ’eː̰tʃ’**/ ‘strawberry’(speaker F1)

As can be seen in comparing figures 20 and 21, the duration of preaspiration may be longer than the duration of preglottalization. In pre-glottalized ejectives, complete glottal closure is often realized before oral closure, in which case the only cues to the place of articulation lie in the release. The glottal closing gesture in ejectives without pre-glottalization may follow the oral closure in which case there may be a short period of voicing into the closure, as seen in figure 22. Alternatively, the glottal closure may occur virtually simultaneously with the oral closure. An example of this latter type of ejective can be seen in figure 23.

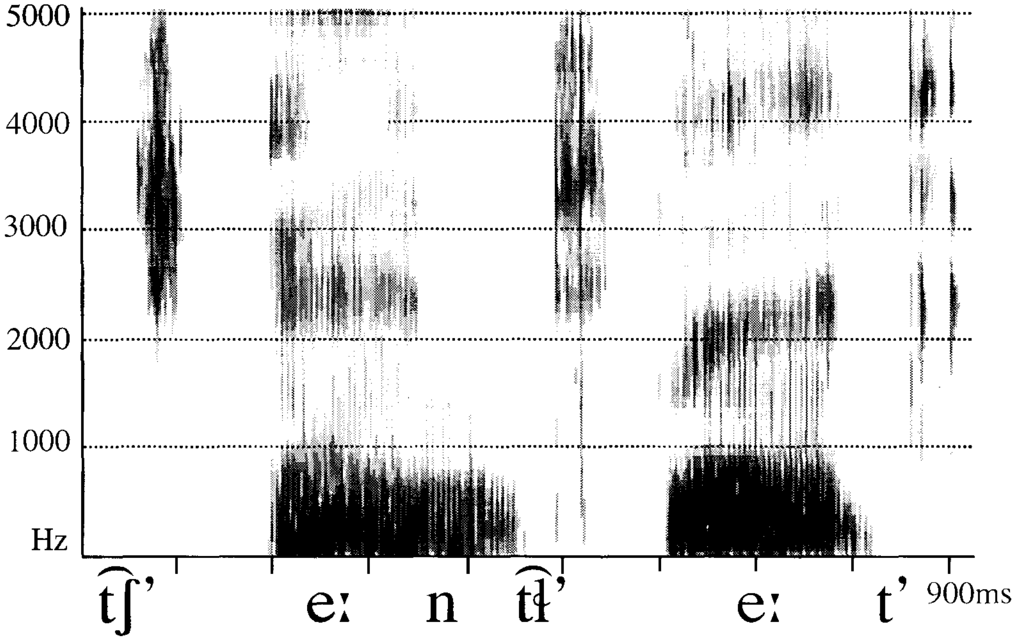


Figure 22: Spectrogram illustrating an ejective without pre-glottalization in the word **/t͡ʃ'eːnt͡ɬ'eːt'/** 'it bulges out' (speaker F1)

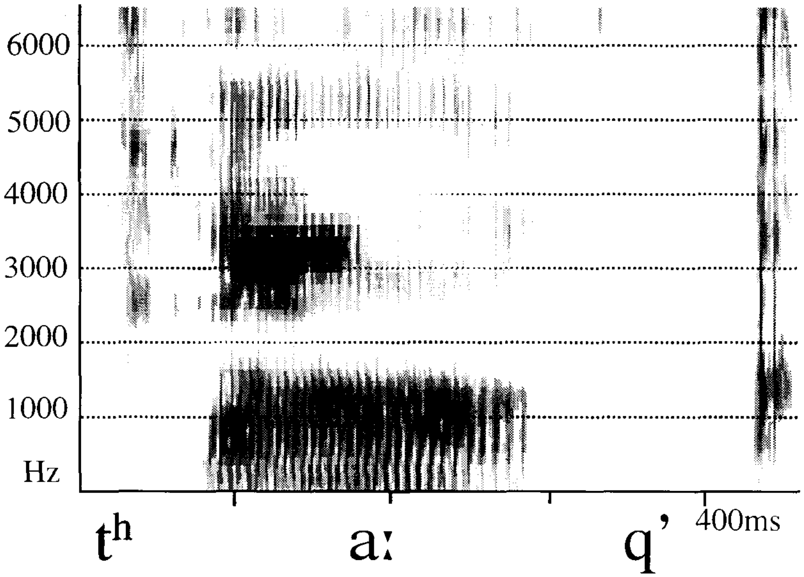


Figure 23: Spectrogram illustrating an ejective with simultaneous oral and glottal closure in the word **/tʰaːq'/** 'three' (speaker M1)

In addition to the laryngeal alternations affecting obstruents, pre-glottalized sonorants contrast with post-glottalized sonorants in stem-final position: e.g. **/han̰/**'river' **/nɪt͡ʃʷʰḛn/** 'it is bad'. Typically, in the case of pre-glottalized sonorants, glottalization is not realized as a complete glottal stop, but rather as creak on the end of the preceding vowel and on the beginning of the sonorant. This tendency to realize glottalization as creak is particularly strong in the case of non-nasal glottalized sonorants. An example of a pre-glottalized **/w/** is found in figure 24. A pre-glottalized nasal is illustrated in figure 25.

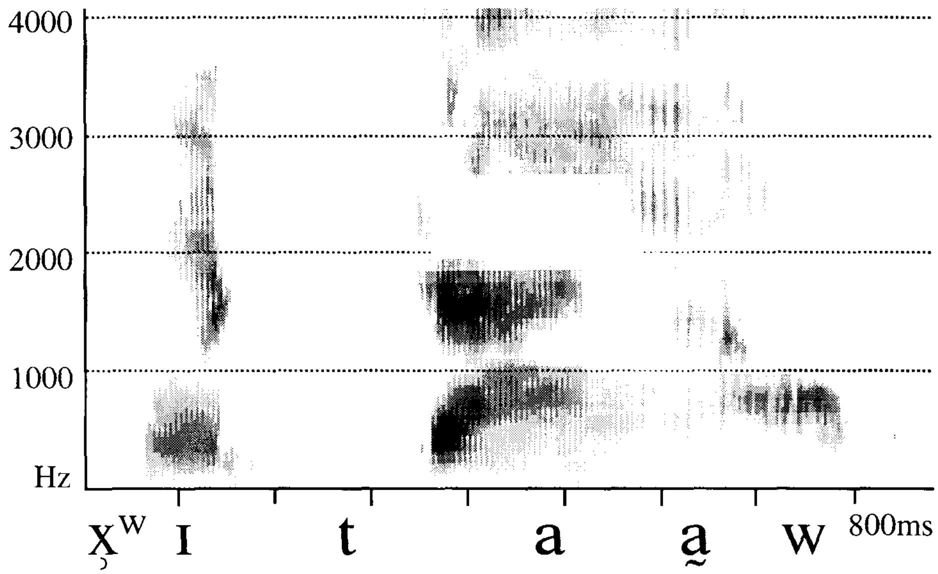


Figure 24: Spectrogram illustrating a pre-glottalized /**w**/ in the word /**χ̹ʷɪta̰ːw**/ 'his beard' (speaker M2)

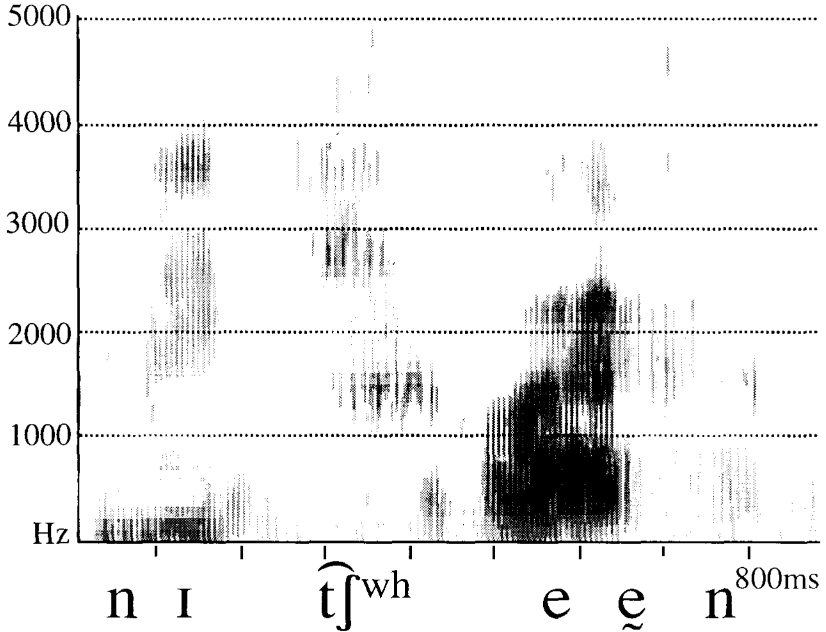


Figure 25: Spectrogram illustrating a pre-glottalized /n/ in the word /**nɪt͡ʃʷʰḛn**/ 'it is bad' (speaker M2)

Pre-and post-glottalized nasals have many different phonetic realizations which occur in free variation with one another. Common to both pre- and postglottalized sonorants is the frequent loss of the nasal consonant, with residual nasalization of the preceding vowel in the case of the post-glottalized nasals and nasalized approximants. The nasal stops in post-glottalized nasals is typically quite short and abruptly truncated by the glottal stop. Glottalization also typically spills over onto the end of the nasal. Examples of post-glottalized nasal stops and approximants can be seen in figures 12 and 13 in section 2.

Figures 26-28 depict various realizations of pre-glottalized nasals. In utterance-final position, preglottalized nasals often lose their nasal stop and instead have an oral release. This is illustrated in figure 26.

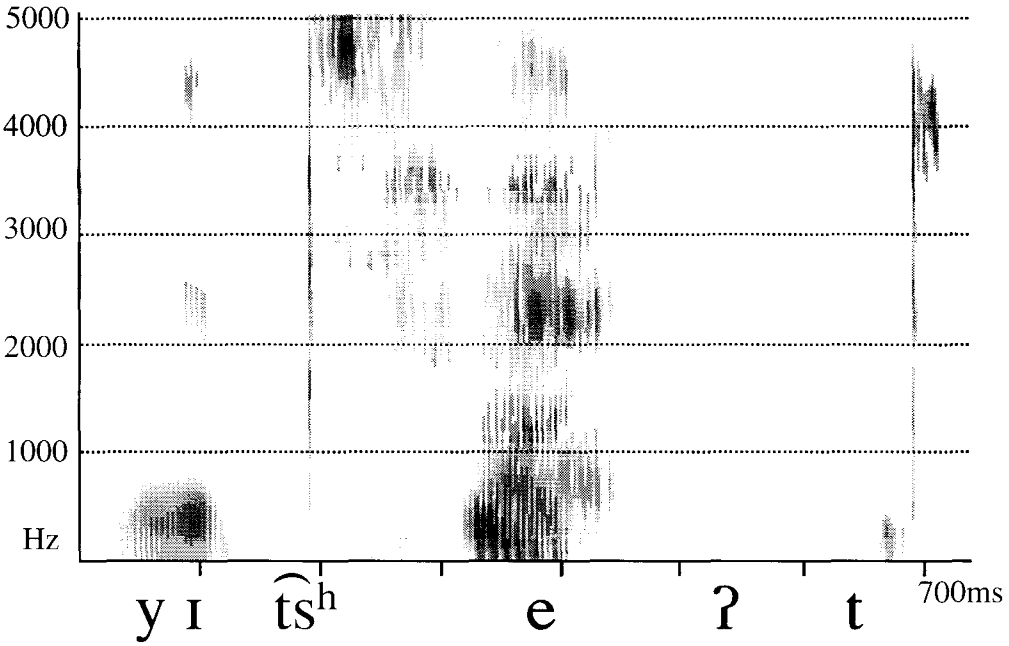


Figure 26: Spectrogram illustrating a pre-glottalized nasal with oral release in the word **/yɪt͡sʰḛːn/** ‘downhill’ (speaker M1)

Another realization of a preglottalized nasal can be seen in figure 27. In this particular token, the glottal closure is complete and the nasal following the glottal stop is voiceless. This type of realization seems to be fairly common in phrase-final position.

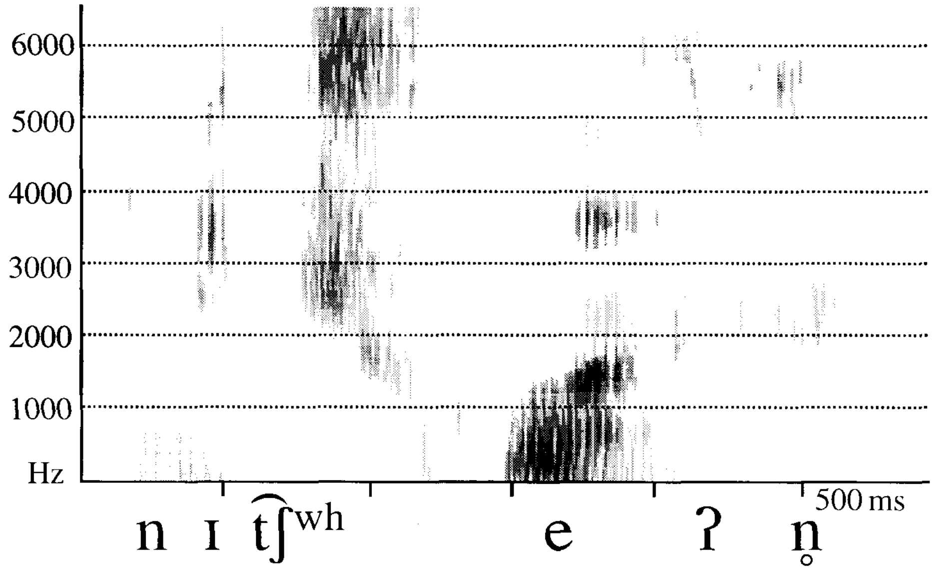


Figure 27: Spectrogram illustrating a preglottalized nasal which is voiceless in the word **/nɪt͡ʃʷʰe_̰n̥/** ‘it is bad’ (speaker M1)

The preglottalized nasal may also be voiced as in figure 28.

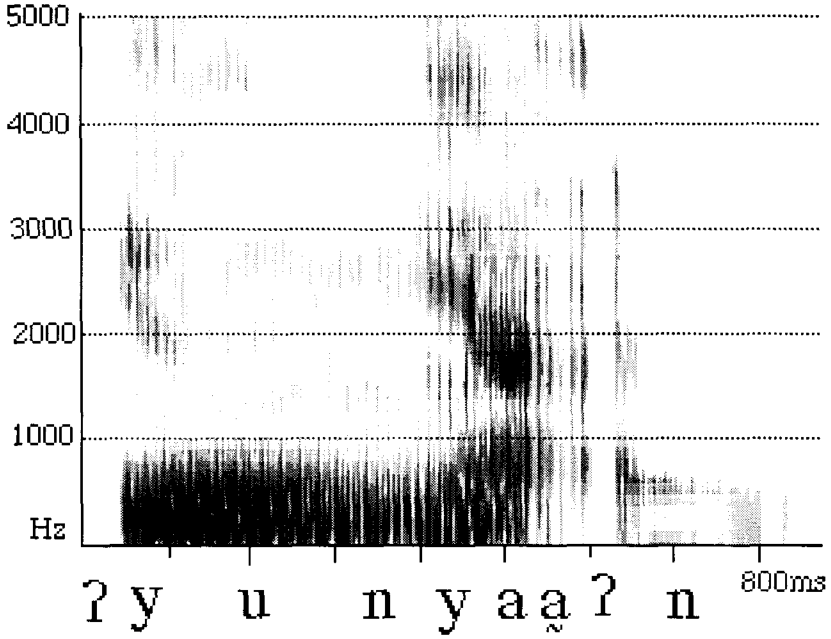


Figure 28: Spectrogram illustrating a preglottalized nasal which is voiced in the word [**ʔyunya ̰n**] (underlyingly /**kʲ'ɪwɪya ̰n**/) 'unshelled acorns' as produced by speaker F1

In this token, the palatalized velar stop at the beginning of the word has been lost leaving only the palatal glide in its place. There is also a lengthy period during which the vowel preceding the final nasal is glottalized. Full glottal closure is only achieved for a very brief period of time immediately prior to the beginning of the nasal.

As pointed out by Golla (1970, 1977), these surface alternations in laryngeal timing are the result of differences in the underlying representation of forms with different laryngeal realizations. Pre-glottalized and pre-aspirated obstruents occur in positions where they underlyingly precede a consonant or a word-boundary, while obstruents without preaspiration or preglottalization underlyingly precede a vowel which is lost on the surface unless a suffix follows. The effect of the underlying vowel on the ordering of glottalization and oral features in sonorants is opposite to the pattern seen for obstruents. Preglottalized sonorants underlyingly precede a vowel, while post-glottalized sonorants underlyingly precede a consonant or word boundary.

3. Vowels

Hupa has six phonemic vowels, three full (or long) vowels and three reduced (or short) vowels. The three full vowels are markedly longer than the reduced vowels. Both the set of full vowels and the set of reduced vowels contain a front vowel, a low central vowel and a back vowel. Formant values for the full and reduced centralized vowels are quite similar to each other, as are values for the full and reduced back vowel pair. Given the length distinction between the full and reduced vowels and their similar formant values, the full vs. reduced distinction is analyzed below as a distinction primarily involving length for the central and back vowels, but one of quality as well as length for the front vowels. The six vowel phonemes appear in table 4.

Table 4: Vowel phonemes of Hupa

Short		Long	
ɪ	**o**	**eː**	**oː**
a		**aː**	

In addition to the six vowels in table 4, a retracted and rounded allophone of /**ɪ**/ approximating /**u**/ occurs before /**x̧**^{w}/, /**x**^{w}/ and /**w**/. A sequence of /**ɪwɪ**/ is realized as the single long vowel /**uː**/. The reduced vowels, especially /**ɪ**/, have many different phonetic realizations depending on the surrounding environment. Reduced vowels are often extremely short in unstressed position, as can be seen in figures 3 and 26.

3.1. Vowel formants

Formants for the seven vowels of Hupa, the six phonemic vowels and the allophone /**uː**/, were measured from a corpus uttered by three speakers on Kay CSL. Data was sampled at 10 kHz. Formants were calculated for a steady state portion of the vowel at about the mid-point of its duration. Superimposed LPC and FFT spectra, calculated over 30 and 25.6 ms frames, respectively, centered around the midpoint of the value were displayed. Formants values are typically those determined by the LPC analysis, based on either 12 or 14 coefficients. The measured vowels occurred in stressed syllables, in either monosyllabic or disyllabic words.

Mean values for each of the three speakers, along with the number of measured tokens, appear in table 5.

Table 5. Mean values for the first three formants for the vowels of three Hupa speakers

Speaker and Vowel		**F1**	**F2**	**F3**	**Speaker and Vowel**		**F1**	**F2**	**F3**
M1	**n**				**F1**	**n**			
eː	21	554	1823	2463	**eː**	23	596	2222	2972
ɪ	71	400	1840	2533	**ɪ**	11	446	2287	3102
aː	40	729	1206	2691	**aː**	20	778	1467	3188
a	21	721	1242	2594	**a**	17	482	1634	2860
oː	18	552	961	2459	**oː**	9	565	1107	2821
o	9	542	891	2418	**o**	4	520	1140	2816
uː	7	415	842	2218	**uː**	7	778	1467	2644
M2									
eː	14	566	1750	2238					
ɪ	26	417	1940	2476					
aː	22	787	1231	2040					
a	6	754	1337	1982					
oː	5	537	862	2065					
o	3	455	804	2197					
uː	6	393	794	2170					

Plots for F1 vs. F2 for the two male speakers together appear in figure 29, while F1 and F2 values for the female speaker are plotted in figure 30. The ellipses enclose all points within two standard deviations of the mean for each vowel.

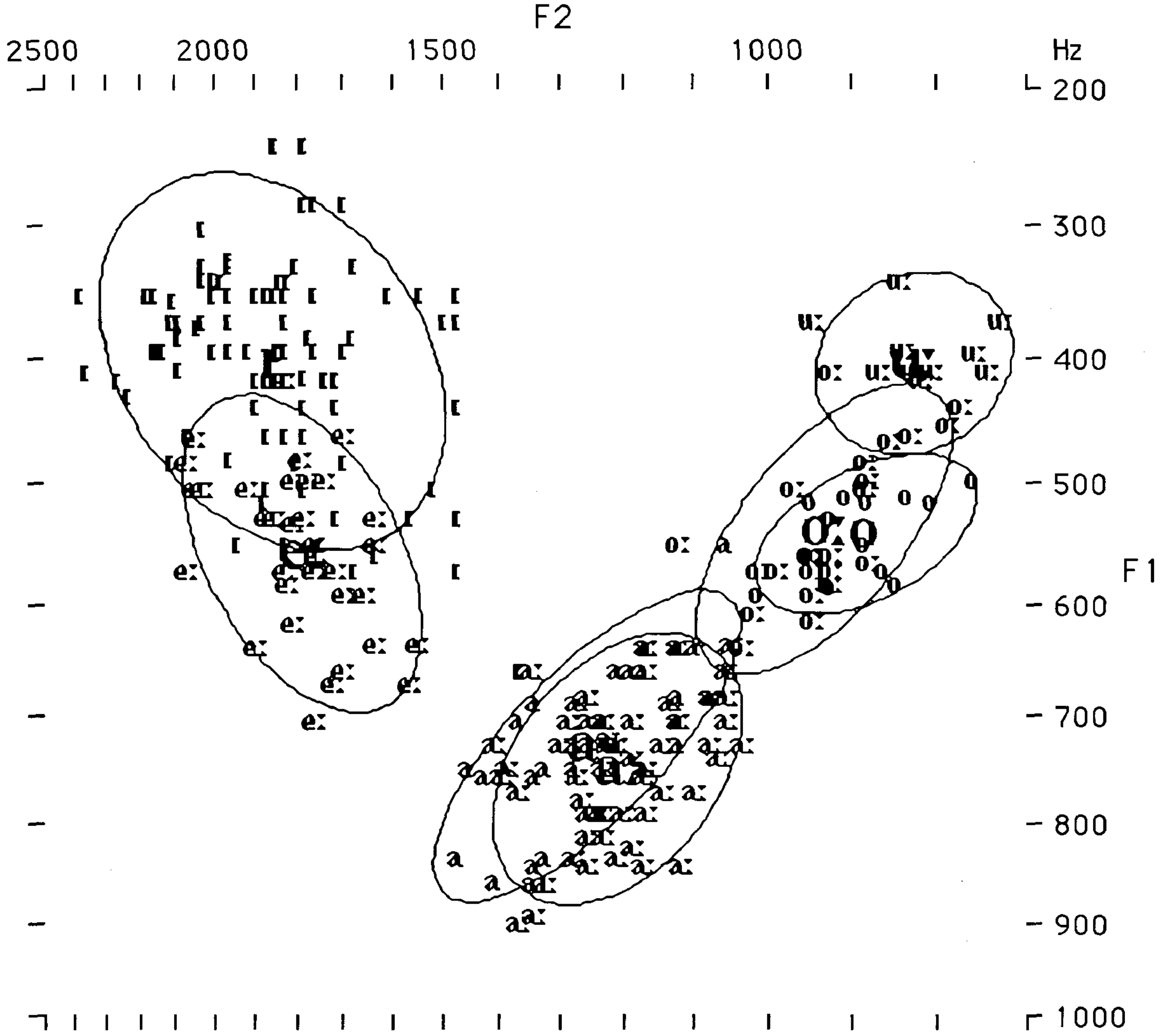


Figure 29: F1 and F2 values plotted for all tokens for speakers M1 and M2.

As is clear from both the formant measures and plots, there is a fair amount of overlap in values for the first and second formants between the short and long central vowels, and also between the short and long back vowels. However, formant values for short /a/ and long /aː/ are slightly different for each of the three speakers. Mean values for the second formant for short /a/ are greater than for long /aː/ for all three speakers, suggesting a slighter more forward articulation for short /a/ as compared to long /aː/. Note, however, that this difference is very small for speaker M1. Values for the first and third formant are also smaller for short /a/ than for long /aː/ for all speakers, though these differences are only appreciable for speaker F1, the same speaker who shows the greatest difference in second formant values for the short and long central vowels. The lower first formant values are suggestive of a higher tongue position for short /a/, while the lower values for the third formant short /a/ perhaps are indicative of some lip rounding. Values for the first and second formants are quite similar between short /o/ and long /oː/, though there is a slight trend for the long /oː/ to have higher first and second formant values than the short /o/. The backed and rounded allophone of /ɪ/, transcribed as /uː/, is more central for speaker F1 than for the two male speakers. Lowered third formant values for speakers M1 and F1 also suggest lip rounding in the production of this vowel. The long short pair /eː/ and /ɪ/ show the least overlap of the three long vs. short pairs. /ɪ/ has lower first formant and higher second formant values for all three speakers, indicative of a higher and fronter articulation for /ɪ/ than for /eː/.

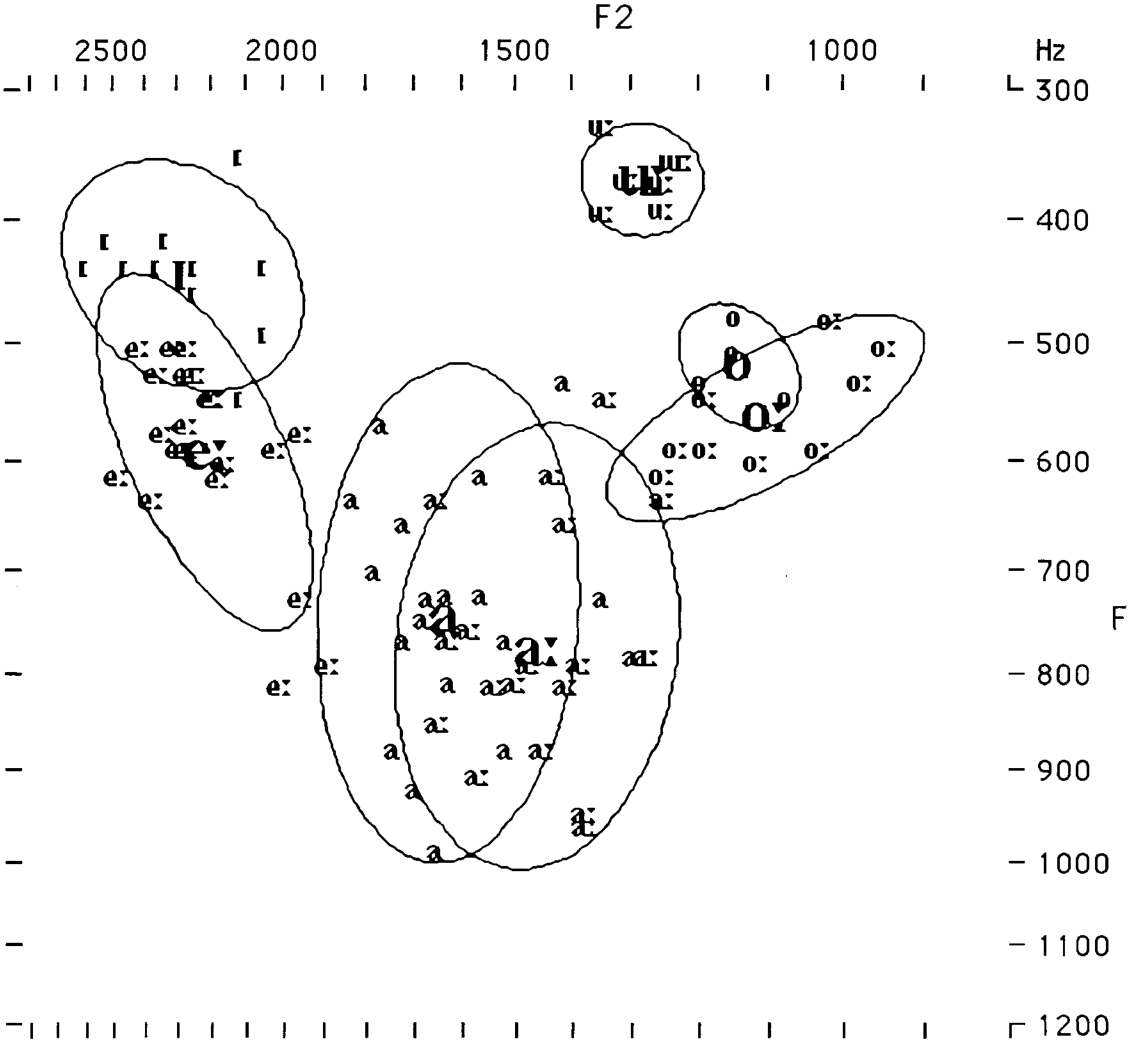


Figure 30: F1 and F2 values plotted for speaker F1.

References

Goddard, Pliny Earle (1904). *Hupa texts.* University of California Publications in American Archaeology and Ethnology 1. 89-368.

Goddard, Pliny Earle (1905). *The Morphology of the Hupa Language.* University of California Publications in American Archaeology and Ethnology 3. 1-344.

Goddard, Pliny Earle (1907). *The Phonology of the Hupa Language. Part I, The Individual Sounds.* University of California Publications in American Archaeology and Ethnology 5. 1-20.

Goddard, Pliny Earle (1928). *Pitch Accent in Hupa.* University of California Publications in American Archaeology and Ethnology 23. 333-8.

Golla, Victor (1964). Etymological study of Hupa noun stems. *International Journal of American Linguistics* 30:2. 108-117.

Golla, Victor (1970). *Hupa Grammar.* Ph.D. dissertation. University of California, Berkeley.

Golla, Victor (1977) A note on Hupa verb stems. *International Journal of American Linguistics* 43:4. 355-8.

Golla, Victor (1985). *A Short Practical Grammar of Hupa.* Hoopa: Hoopa Valley Tribe.

Hoijer, Harry (1960). Athabaskan languages of the Pacific coast in Stanley Diamond (ed.) *Culture in History: Essays in Honor of Paul Radin*. 960-76.

Hoijer, Harry (1971). Athabaskan Morphology. in Jesse Sawyer (ed.) *Studies in American Indian Languages*. Los Angeles: University of California Press. 113-48.

Krauss, Michael and Leer, Jeff (1981). *Athabaskan, Eyak, and Tlingit sonorants.* Fairbanks: Alaska Native Language Center Research Papers 5.

Ladefoged, Peter and Maddieson, Ian (1996). *Sounds of the World's Languages*. Oxford: Blackwells.

Leer, Jeff (1979). *Proto-Athabaskan Verb Stem Variation, Part One: Phonology*. Fairbanks: Alaska Native Language Center Research Papers 1.

Li, Fang-Kwei (1930). *Mattole, an Athabaskan Language*. Chicago: University of Chicago Press.

Lindau, Mona (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics* 12. 147-55.

Lisker, L. and Abramson, A (1964). Cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384-422.

Maddieson, Ian (1996, to appear) *Phonetic universals.* In Laver, J. and Hardcastle, W. (eds.) Handbook of Phonetic Sciences. Oxford: Blackwells.

McDonough, Joyce and Ladefoged, Peter (1993). Navajo stops. *UCLA Working Papers in Phonetics* 84. 151-164.

Woodward, Mary (1964). Hupa Phonemics. in William Bright (ed.) *Studies in Californian Linguistics*. University of California Publications in Linguistics 34. 199-216.

Acknowledgments

Many thanks are due to Victor Golla and Danny Ammon for their assistance and helpful advice and suggestions. A special debt of gratitude is owed to the kind speakers of Hupa who so generously gave of their time and expertise and taught me about their language. Also, more generally, a warm thank you to the people of Hoopa Valley for their hospitality and good will.