Title
WPP, No. 98: Fieldwork Studies of Targeted Languages VI

Publication Date
2000-04-01

Fieldwork Studies of Targeted Languages VI

UCLA WPP Volume 98

April 2000

The UCLA Phonetics Laboratory Group

Heriberto Avelino
Adam Albright
Mary Baltazani
Marco Baroni
Roger Billerey
Rebecca Brown
Andrew Bye
Taehong Cho
Sandy Disner
Melissa Epstein
Heidi Fleischhacker
Christina Foreman
Suzi Meyers Fosnot
Matt Gordon
Bruce Hayes
Amanda Jones
Sun-Ah Jun
Pat Keating
Sahyang Kim
Somy Kim
Jenny Ladefoged
Peter Ladefoged
Hyuck-Joon Lee
Ian Maddieson
Mira Oh
Amy Schafer
Donca Steriade
Melissa Tai
Henry Tehrani
Siri Tuttle
Motoko 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 DC 00029
NSF grant SBR 9319705
NSF grant SBR 9511118
NSF grant IIS 99960188
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 90095-15643
phonetic@humnet.ucla.edu
http://www.humnet.ucla.edu/humnet/linguistic/faciliti/wpp.htm

UCLA Working Papers in Phonetics

98

April 2000

Table of contents

Linking linguistic contrasts to reality: The case of VOT

Peter Ladefoged and Taehong Cho
[To be published in a Festschrift for Eli Fischer- Jørgensen]

In the early days of acoustic phonetics Eli Fischer-Jørgensen (1954) described a number of important properties of stop consonants, including the fact that the length of the aspiration in voiceless stops (Voice Onset Time, or VOT, as it was later termed) varied with place of articulation. Cho and Ladefoged (1999) summarize several factors underlying this universal tendency, showing how it may depend on aerodynamic circumstances, the mass and mobility of different articulators, temporal adjustment between the closure duration and VOT, and perceptual concerns. These factors are given different weights in different languages, resulting in variations across languages in the way contrasts in VOT are manifested.

This paper is based on the data reported in Cho and Ladefoged (1999). It will discuss the extent to which the phonological inventory of a language affects the VOT, and conclude by discussing how a phonological description of a single language can make explicit statements about the physical manifestation of VOT in that language, noting how it differs or is the same as that in other languages. We will not consider within languages variations in VOT due to place or articulation or other factors. It is sufficient for our purposes to consider VOT only in voiceless velar stops in a number of different languages.

Previous discussions of cross-linguistic variations in VOT (e.g. Lisker & Abramson 1964, Keating 1984) have been hampered by not being able to refer to a large body of data that had been collected in accordance with the same protocol. When comparing the phonetic properties of different languages it is important to collect data from a number of speakers of each language, so that any individual bias is discounted, and to analyze the data in the same way, so that differences in recording techniques and measurement procedures do not affect the results. We were fortunate to have access to a body of data that fulfilled these prerequisites.

The data we used consisted of measurements of VOT in 18 languages. For the last 8 years Peter Ladefoged and Ian Maddieson have been the Principal Investigators of a National Science Foundation sponsored project to study the phonetic structures of endangered languages. (This project is continuing for a further three year period under the sole direction of Ian Maddieson.) Endangered languages are no different from any other languages in their possible phonetic structures. These languages are endangered in that they are losing speakers, usually for socio-economic reasons that have no linguistic biases. Consequently a sample of the world's languages consisting solely of languages spoken by communities that are disappearing could theoretically be representative of the world's languages as a whole. Of course, as Maddieson (1997a) and others have pointed out, there are enormous problems in getting a valid sample of languages that is truly representative of the 6,703 (Grimes 1999) current languages in the world (not to mention the further problems of extending this sample to cover all human languages that ever have been or ever could be). The 18 languages that are used in the present work fail completely as a representative sample of possible languages. But they are diverse enough to be at least indicative of the range of VOT that can be found.

The 18 languages in the data set represent 12 different language families as shown in Table 1. Most of the languages are spoken by a comparatively small number of speakers, but Navajo and Apache are fairly widely spoken. Navajo is not an endangered language, but it was investigated in the same way as the other languages. Jalapa Mazatec is also not dying rapidly. It is spoken by nearly all the inhabitants of Jalapa de Diaz in Mexico, including the children. It is endangered in the sense that it is changing rapidly due to the influence of Spanish. Many distinctions are no longer made by younger speakers. Scottish Gaelic may be spoken by 70,000 people, as we have been told, but it is clearly an endangered language, spoken by very few young people. Eastern and Western Aleut are closely related, but they have a number of phonological differences.

Table 1. Languages in the data set.

LANGUAGE	FAMILY	LOCATION
Aleut (Eastern)	Eskimo-Aleut	Alaska, U.S.A.
Aleut (Western)	Eskimo-Aleut	Alaska, U.S.A.
Apache	Athabaskan	Arizona, U.S.A.
Banawá	Arawan	Northern Brazil
Bowiri	Niger-Congo	Ghana
Chickasaw	Muskogean	Oklahoma, U.S.A.
Dahalo	Cushitic	Kenya
Defaka	Niger-Congo	Nigeria
Gaelic	Indo-European	Scotland, U.K.
Hupa	Athabaskan	California, U.S.A
Jalapa Mazatec	Otomanguean	Mexico
Khonoma Angami	Tibeto-Burman	Nagaland, India
Montana Salish	Salishan	Montana, U.S.A.
Navajo	Athabaskan	New Mexico, U.S.A.
Tlingit	Athabaskan	Alaska, U.S.A.
Tsou	Austronesian	Taiwan
Wari'	Chapacuran	Northern Brazil
Yapese	Austronesian	Western Pacific

The recordings of these 18 languages were made in the field in a standardized way by one or other of the two Principal Investigators, Peter Ladefoged and Ian Maddieson, with the exception of the Hupa data, which were recorded by Matthew Gordon, at that time a graduate student in the UCLA Phonetics Lab. Material illustrating the full range of segmental contrasts in each language was recorded, but in this paper we will refer only to the data on voiceless unaspirated and voiceless aspirated velar stops. These stops were always recorded in initial position in citation forms of contrasting words before a non-high vowel. It is arguable that this is not the most appropriate data in that it does not reflect natural utterances in the languages. We thought, however, that it was preferable to ensure unity of style across languages, even at the expense of naturalness.

Several speakers of each language were recorded using high quality equipment. All the speakers were adult native speakers who used the language in their daily life. As noted above, most of the languages investigated are moribund (the children no longer speak them), but all our speakers were completely fluent. The recordings were all analyzed by Research Associates in the UCLA Phonetics Lab in the same way (for details see Cho and Ladefoged, 1999). The differences between languages that emerged are almost certainly not artifacts of the slight differences in the circumstances in which the recordings were made, nor are they due to variations in the measurement techniques of the graduate students who worked on this project, all of whom were closely supervised and trained in the same way. We can conclude that the data reflect real differences between languages.

Figure 1 shows the mean VOT of the velar stops in these 18 languages. If a language contrasts aspirated and unaspirated velar stops, both values are shown, so that there are 25 columns of mean values. There is a very wide range of values, going from20 ms for the Khonoma Angami voiceless unaspirated stops to 154 ms for the Navajo aspirated stops. The lower of these figures is not unexpected; voiceless velar stops are known to have some voicing lag, so it is no surprise that the lowest value recorded in the data set is significantly greater than zero. But, to those unfamiliar with Navajo, the value of 154 ms for aspirated velar stops may seem excessive — almost as if it were an artifact. It is not. As listeners to the language can attest, aspiration is a very salient feature of Navajo speech.

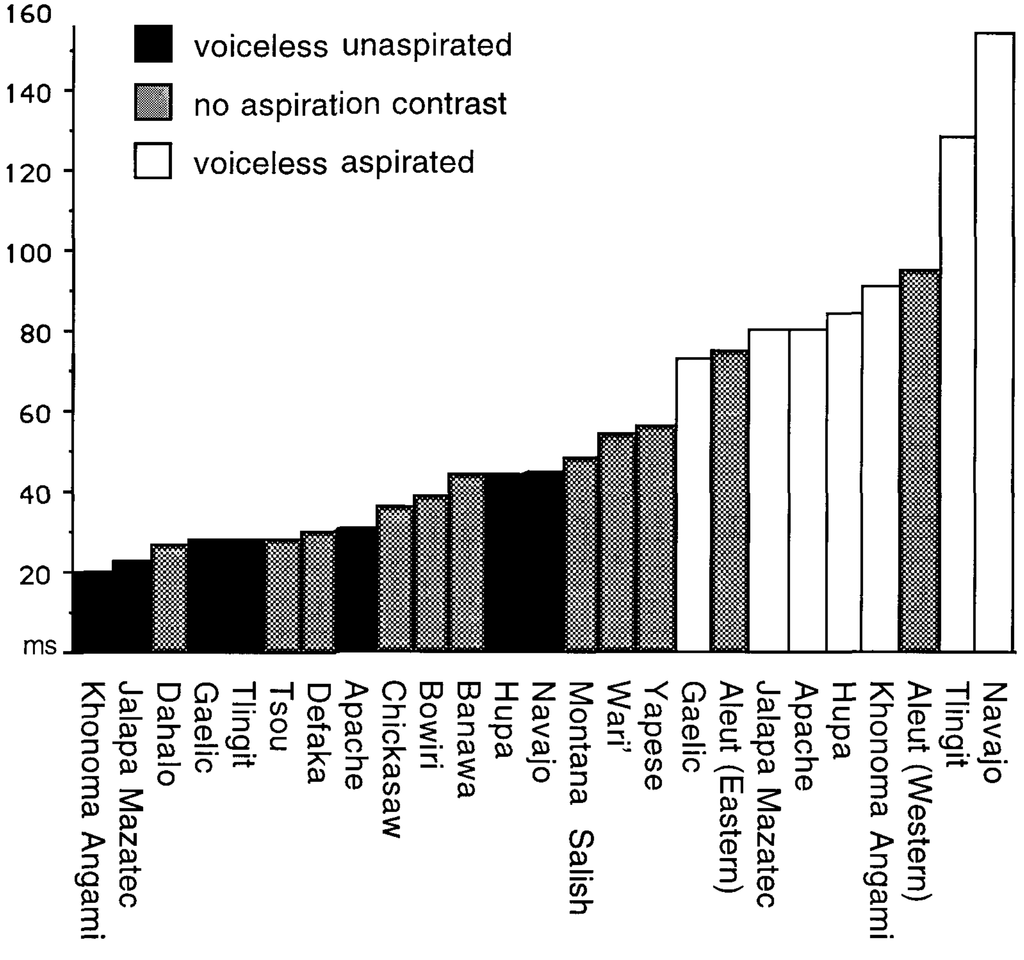


Figure 1. Mean VOTs (ms) for voiceless velar stops in 18 languages. (Values for voiced velar stops are not shown.)

Before discussing the data in detail, we must consider the set of oppositions that occurs among these velar stops in each language. The fact that in some languages there is only a single velar stop, while in others there are two, and in yet others three, might be expected to influence the VOT chosen by each language. If there is no need to make a perceptual distinction between two sounds, then one might expect a language to use the simplest articulatory gesture, what Docherty (1992) has called the low-cost option. This simplest gesture — whatever it is — is presumably the same for all human beings. So one might expect all languages that have only one velar stop to use the same gesture. When there is more than one voiceless velar stop, then each of them must be kept distinct and something other than the low-cost option will have to be used. It is important, therefore, to consider the phonological oppositions. The contrasts within each language are shown in Table 2. We have omitted labialized velars, considering them to be irrelevant to the present discussion of VOT. But we have included information on ejectives, as these sounds are perceptually similar to velar stops that are distinguished simply by VOT. None of the languages in the sample has distinctions due to variations in phonation type, such as creaky voice or breathy voice. The symbols used are those chosen by the authors of the individual studies.

Table 2. The contrasting velar stops (not including labialized stops) that occur in each of the 18 languages.

LANGUAGE	CONTRASTING VELAR STOPS	SOURCE
Aleut (Eastern)	**k**	Cho et al. 1997
Aleut (Western)	**k**	Cho et al. 1997
Apache	**k, k^h, k’**	Gordon et al. 2000
Banawá	**k**	Ladefoged et al. 1997
Bowiri	**k, g**	Maddieson p.c.
Chickasaw	**k**	Gordon et. al. 1997
Dahalo	**k, g**	Maddieson et al. 1993
Defaka	**k, g**	Shryock, et al. 1996
Gaelic	**k, k^h**	Ladefoged et al.1997
Hupa	**k, k^h, k’**	Gordon, 1996
Jalapa Mazatec	**k, k^h**	Silverman et al. 1995
Khonoma Angami	**k, g**	Blankenship et al. 1993
Montana Salish	**k, k’**	Flemming et al. 1994
Navajo	**k, k^h, k’**	McDonough & Ladefoged 1993
Tlingit	**k, k^h, k’**	Maddieson et al. 1996
Tsou	**k**	Wright & Ladefoged 1997
Wari’	**k**	MacEachern et al. 1997
Yapese	**k, k’, g**	Maddieson 1997b

There are six languages in the set that have only one velar stop. In none of them is this stop voiced, so, quite understandably from a phonological point of view, the authors of the original studies have represented each of these stops by the symbol **k.** But from a phonetic point of view, these stops vary considerably. Some might well be considered phonetically aspirated stops, and the others unaspirated. Both forms of Aleut have stops that are among the most aspirated in the set. The other four languages that have only a single velar stop, Banawá, Chickasaw, Tsou and Wari’, vary in their choice of VOT, Tsou being among the lower group and Wari’ among the higher. None of these six languages needs to make a perceptual distinction between two sounds, and they might all be expected to use the simplest articulatory gesture. But the data show that they do not necessarily choose the same simplest, low-cost, articulation.

There are only four languages, Bowiri, Dahalo, Defaka and Khonoma Angami, that have what those with a European bias might consider to be the typical velar contrasts, **k** vs **g**. A fifth, Yapese, has this contrast plus an ejective. We do not have adequate data on the voiced stops in these five languages (except that we know that they all have some voicing during the stop closure), so we cannot say which of them are more like French, with a contrast between a fully voiced stop and a voiceless unaspirated stop, and which more like English, with a contrast between a partly voiced stop and an aspirated stop.

In our pursuit of the distinction between voiceless unaspirated and aspirated stops, we can consider all 11 languages that have a single voiceless velar stop that has, from a phonetic point of view, to be called voiceless unaspirated or aspirated. When describing these languages we do not need to make the distinction for phonological reasons, but we must, within the usual techniques of phonetic description, say that these stops fall into the one phonetic category or the other. The VOT's for these 11 languages are represented by the gray columns in Figure 1. Where should we draw the line for the phonetic boundary between voiceless unaspirated stops and aspirated stops? If we look at the data in Figure 1 it might appear as if we could draw it after Banawá. But if we remove the distraction of the data from languages that contrast voiceless unaspirated and aspirated stops, and look at languages that have just one voiceless stop, as in Figure 2, the answer is not so obvious. There is a steady increase in VOT, and no obvious way of dividing the data into two phonetic categories until we get to Eastern and Western Aleut, both of which clearly have aspirated stops. The difference between Banawá and Montana Salish, the dividing line suggested by the data in Figure 1, is smaller than that between Banawá and Bowiri, and that between Montana Salish and Wari'. There is a smooth increase in VOT, and we can make only an arbitrary decision about which languages have voiceless unaspirated stops, and which have aspirated stops.

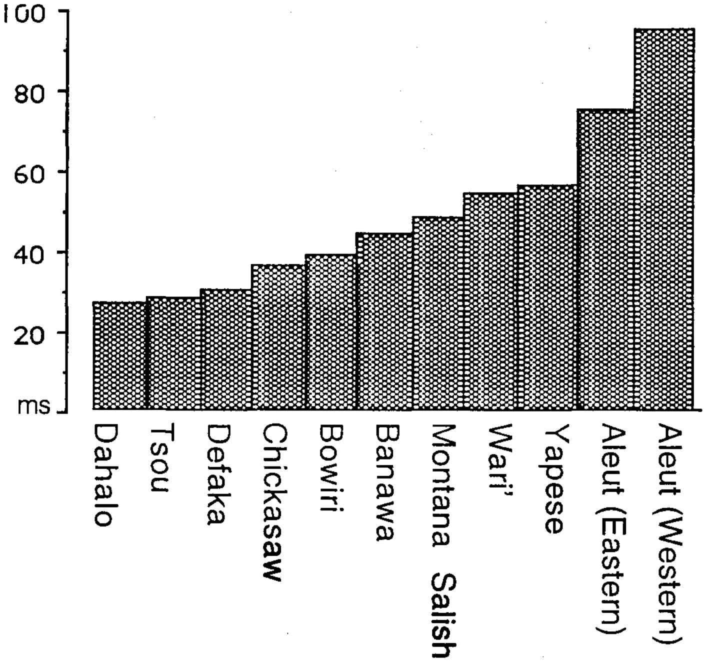


Figure 2. The 11 languages in the sample that do not distinguish voiceless unaspirated and aspirated stops.

Now consider the 7 languages in the sample that contrast voiceless unaspirated and aspirated stops, shown in Figure 3. We noted earlier that when there is no need to make a perceptual distinction we might expect that languages will choose the low-cost option, and make the simplest articulation. This expectation turned out to be wrong. Here, where there is a need to make a perceptual distinction between similar sounds, one might expect that languages would maximize the perceptual difference between them. But, as in the previous case, this expectation is not met. Languages do not behave in this way. Some languages make a large difference, others do not. Hupa has a difference of only 40 ms between voiceless unaspirated and aspirated stops. This is less than half Tlingit's 100 ms.

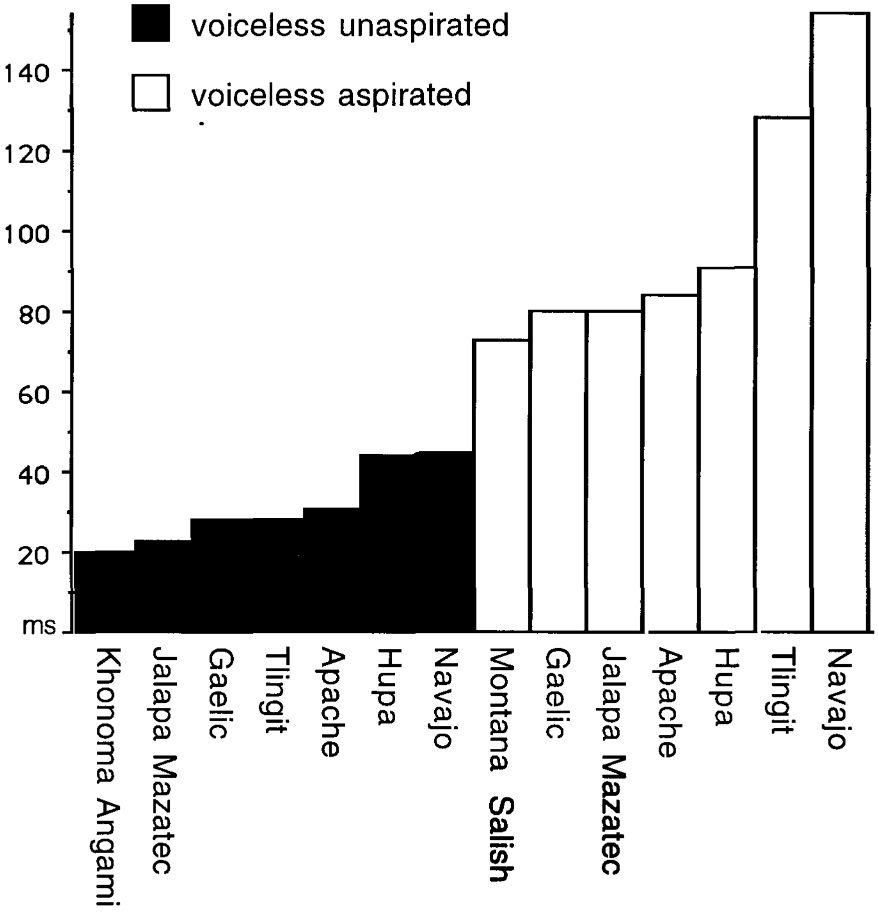


Figure 3. The 7 languages in the sample that distinguish voiceless unaspirated and aspirated stops.

There are six languages that have ejectives. Apache, Hupa, Montana Salish, Navajo and Tlingit contrast **k, k', k**^{h}. Yapese contrasts **k, k', g** and Montana Salish has just **k, k'**. These languages are interesting in that they may use VOT as a helping feature (Stevens, Keyser & Kawasaki,1986) to further the distinction between ejectives and other stops. There is a tendency for this to happen, as can be seen from the data in Figure 4, in which the shading has been arranged so as to make it easier to compare languages. In every language except Hupa the ejectives are clearly distinguished from the other stops by VOT.

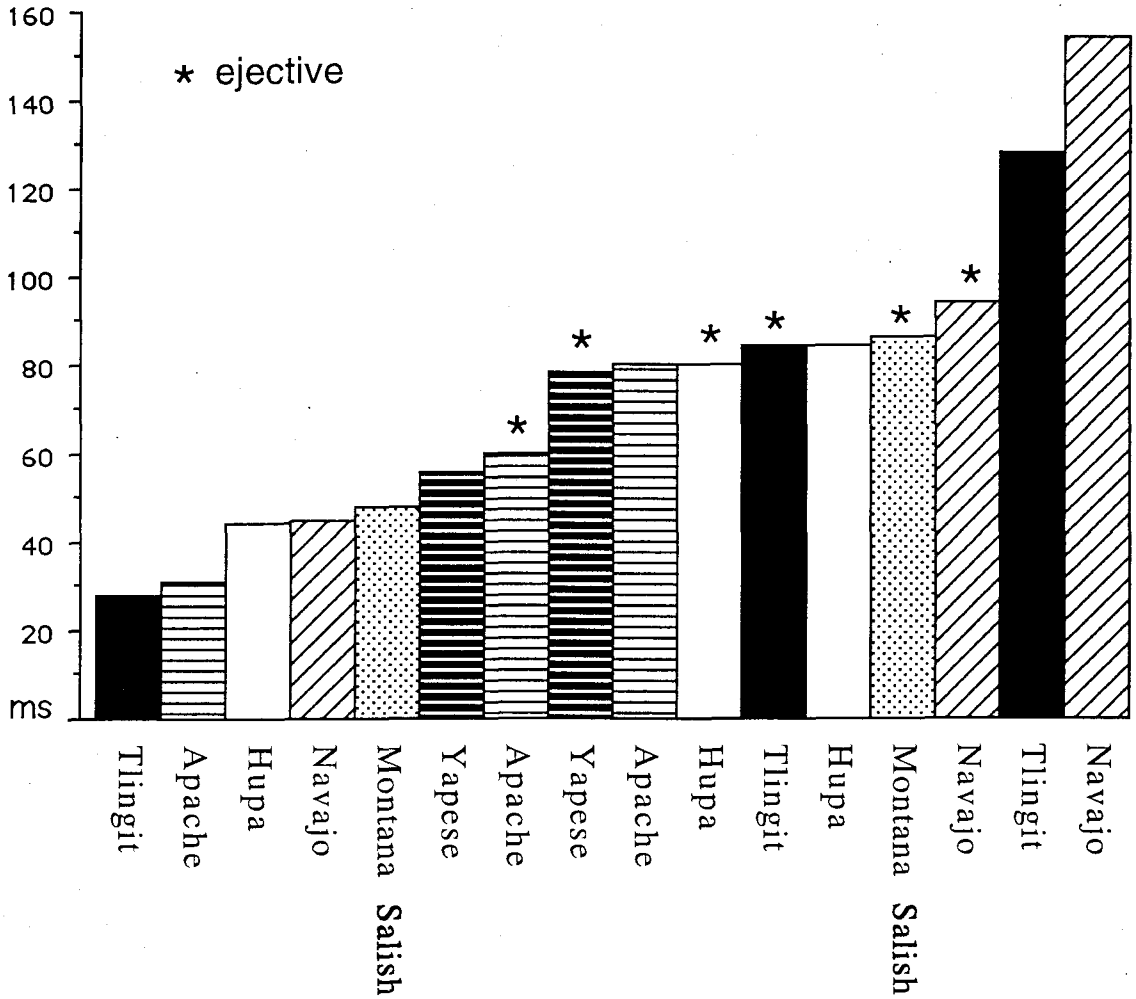


Figure 4. The 6 languages in the sample that have ejectives. Each language has been given a different shading so as to make it easy to compare the ejectives with the other stops in the language.

We will now assess the data as a whole, and consider how phonological statements about each language can be related to observable VOT differences. We presume that phonological descriptions will need to distinguish at least three possibilities [voiced], [voiceless unaspirated] and [aspirated]. We have considered only the [voiceless unaspirated] and [aspirated] categories in this paper, partly because we do not have data on voiced stops, and partly because we do not know of any claims that different degrees of negative VOT are phonologically contrastive. However, languages vary in the amount of voicing that can occur in a phonologically voiced stop, and any mechanism that we propose for the realization of phonologically voiceless stops in terms of physical phonetic variables should apply in a similar way to phonologically voiced stops.

How should the three possibilities {voiced, voiceless unaspirated, aspirated} be realized? Because we want a variable that will be constant within a language across different places of articulation, we propose making phonetic specifications not in terms of the directly observable acoustic measure, VOT, but in terms of an underlying physiological measure. We suggest that there is a phonetic parameter, which we will call Articulatory

VOT, definable in terms of the difference in time between the initiation of the articulatory gesture responsible for the release of a closure and the initiation of the laryngeal gesture responsible for vocal fold vibration. We think it likely that speakers aim for a certain timing difference between articulatory and glottal gestures irrespective of the articulatory gesture involved. This is the low-cost option suggested by Docherty (1992). Differences in VOT within a language are usually the inevitable consequence of the physiological movements and the aerodynamic forces that occur at different places of articulation and in different syntagmatic contexts. Cho and Ladefoged (1999), however, found a few cases in which a single VOT target cannot account for all the observed variations in VOT within a language. Sometimes there may be variations in VOT ascribable to aerodynamic causes (e.g. the variations due to place of articulation in unaspirated stops) that a language may choose to use in other circumstances (e.g. as an aid to the perception of places of articulation of aspirated stops in which different aerodynamic forces occur and would, unless prevented, have produced a different acoustic effect).

Even if we specify VOT in terms of an underlying physiological parameter, our data show that there is a great deal of between language variation. Moreover, it is impossible to predict the differences between languages from knowledge of the phonological contrasts within a language. It is not the case that if a language lacks a contrast between **k** and **kʰ** it will have the simplest possible VOT, with a value between the modal value for **k** and that for **kʰ** . Nor is it the case that if a language does have a contrast between **k** and **kʰ** will it make that contrast with a larger than usual VOT for **kʰ** and a smaller one for **k**, so as to make sure that the difference is easy to hear. Nor does the VOT in ejectives have any simple relation to the VOT of other phonological contrasts. We propose, as in Cho and Ladefoged (1999), that each language chooses a modal VOT value for each of the categories [voiced], [voiceless unaspirated] and [voiceless aspirated] that are specified in the phonology. The statement of these values is the link between phonology and measurable phonetic parameters. A phonological description of a language that does not include statements of this kind is incomplete.

Many thanks are owed to the members of UCLA Phonetics Lab and all the speakers who participated in the project studying ‘Phonetic structures of endangered languages’. This work was supported by NSF grant SBR 951118 to Peter Ladefoged and Ian Maddieson.

References

Blankenship, Barbara, Peter Ladefoged, P. Bhaskararao and N. Chase. 1993. Phonetic structures of Khonoma Angami. Linguistics of the Tibeto-Burman area 16.69-88.

Cho, Taehong, and Peter Ladefoged. 1999. Variations and universals in VOT: evidence from 18 languages. Journal of Phonetics 27. 207-229.

Cho, Taehong, Alice Taff, Moses Dirks and Peter Ladefoged. 1997. Some phonetic structures of Aleut. UCLA Working Papers in Phonetics 95.41-67.

Docherty, Gerry. 1992. The timing of voicing in British English obstruents. Berlin: NewYork.

Fischer-Jørgensen, Eli. 1954 Acoustic analysis of stop consonants. Miscellanea Phonetica 2.42-59.

Flemming, Edward, Peter Ladefoged and Sarah Thomason. 1994. Phonetic structures of Montana Salish. UCLA Working Papers in Phonetics 87.1-34.

Gordon, Matt. 1996. The phonetic structures of Hupa. UCLA Working Papers in Phonetics 93.164-187.

Gordon, Matt, Pamela Munro and Peter Ladefoged. 1997. The phonetic structures of Chickasaw. UCLA Working Papers in Phonetics 95.41-67.

Gordon, Matt, Brian Potter, Willem de Reuse and Peter Ladefoged. 2000. Phonetic structures of Apache. UCLA Working Papers in Phonetics 97. forthcoming.

Grimes, Barbara E. (ed.) 1999. Ethnologue: Languages of the world. Thirteenth edition. Dallas: Summer Institute of Linguistics.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. Language 60.286-319.

Ladefoged, Peter, Jenny Ladefoged and Daniel Everett. 1997. Phonetic structures of Banawá, an endangered language. Phonetica 54.94-111.

Lisker, Leigh and Arthur Abramson. 1964. Cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384-422.

MacEachern, Margaret, Barbara Kern and Peter Ladefoged. 1997. Wari' phonetic structures. Journal of Amazonian Languages 1.5-25.

McDonough, Joyce and Peter Ladefoged. 1993. Navajo Stops. UCLA Working Papers in Phonetics 84.151-164.

Maddieson, Ian. 1997a. Phonetic universals. Handbook of phonetic sciences, ed. by John Laver and William Hardcastle, 619-639. Oxford: Blackwells Publishers.

Maddieson, Ian. 1997b. The many glottalized consonants in Yapese – how, why and where? Paper presented at the Third Austronesian Formal Linguistics Association Conference, UCLA.

Maddieson, Ian, Bonny Sands, Peter Ladefoged and Sinisa Spajic. 1993). The phonetic structures of Dahalo. Afrikanistische Arbeitspapiere 36.5-53.

Maddieson, Ian, Nicola Bessel and Caroline Smith. 1996. A preliminary phonetic report on the sounds of Tlingit. UCLA Working Papers in Phonetics, 93, 125-148.

Shryock, Aaron, Peter Ladefoged and Kay Williamson. 1996. Phonetic structures of Defaka. The Journal of West African Languages. 26.3-27.

Silverman, Daniel, Barbara Blankenship, Paul Kirk and Peter Ladefoged. 1995. Phonetic structures in Jalapa Mazatec. Anthropological Linguistics 37.70-88.

Stevens, Kenneth. N., S. Jay Keyser and Hiroko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. Invariance and variability in speech processes, ed. by Joseph Perkell and Dennis Klatt, 426-463. New Jersey: Lawrence Erlbaum Associates.

Wright, Richard, and Peter Ladefoged. 1997. A phonetic study of Tsou. Bulletin of the Institute of History and Philology, Academia Sinica 68.987-1028.

Phonetic structures of Western Apache

Brian Potter*, Matthew Gordon**, John Dawson, Willem de Reuse***
and Peter Ladefoged

Linguistics Department, University of California, Los Angeles
*Inquizit Technologies
** Now at Linguistics Department, University of California, Santa Barbara
***Department of English, University of North Texas

1. Introduction

Western Apache is a Southern Athabaskan language spoken primarily on the San Carlos and White Mountain Apache Reservations in east central Arizona. Figure 1 illustrates the locations of these reservations.

Figure 1. Map indicating primary area in which Western Apache is spoken.

The first recorded use of the term 'Apache' is by Juan de Oñate at San Juan Pueblo on September 9, 1598. As discussed in Opler (1983), there are several hypotheses about the original source for this word including the Zuni and Yavapai words **[ʔaapaču]** and **[ʔawáača]**, respectively, referring to the Southern Athabaskan people in general, as well as the Spanish word *mapache*, 'racoon'. The Western Apache people refer to themselves (in their practical orthography) as the '**n̩nee**', '**n̩dee**', or '**ʔinee**', depending on dialect.

The Southern Athabaskan languages, also known as the 'Apachean' languages, additionally include Navajo, and Mescalero, Chiricahua, Jicarilla, Lipan, and Plains Apache. The classification in Figure 2 combines the proposals in Hoijer (1938) and Hoijer (1971), and is compatible with that of Hardy (1979). Note that Chiricahua and Mescalero are now considered to be mutually intelligible dialects of one language. We not know whether Lipan is still spoken.

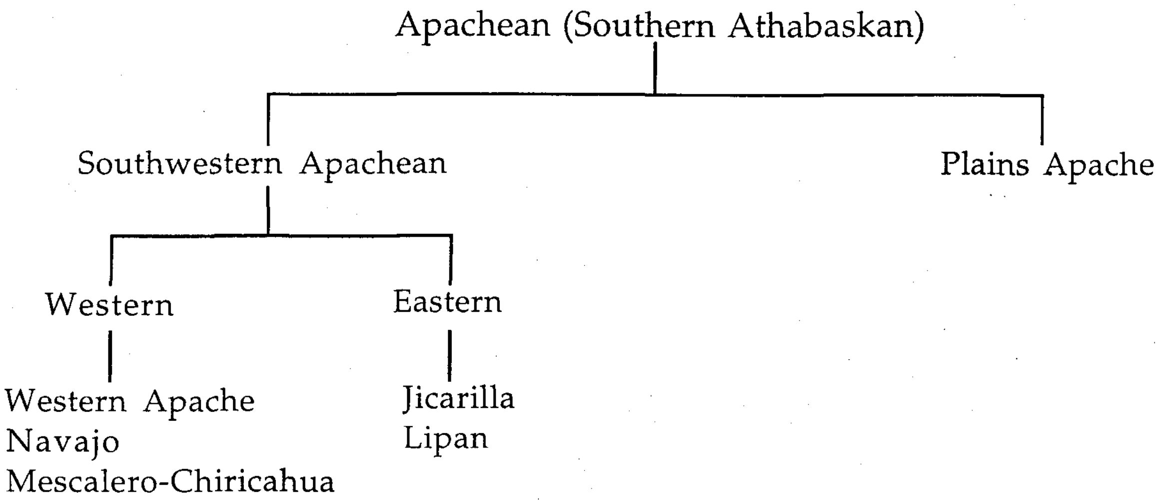


Figure 2. Classification of the Southern Athabaskan languages.

Hoijer identified the Apachean languages as 'closely related' (Hoijer 1963:6) but 'mutually unintelligible dialect groups' (Hoijer 1945:193, 1938). Western Apache is not mutually intelligible with Navajo, its geographically closest relative, but De Reuse (1994, 1992) notes that Western Apache and Navajo speakers understand much of each other's languages, and that some speakers claim mutual intelligibility. Western Apache dialects spoken in regions close to the Navajo reservation exhibit a greater degree of mutual intelligibility with Navajo than do dialects spoken in regions farther from the Navajo reservation. The Tonto dialect of Western Apache, spoken in and near Camp Verde and Payson, Arizona, is geographically and linguistically the closest dialect to Navajo, while the San Carlos dialect, the predominant dialect on the San Carlos Reservation, is geographically and linguistically the most distinct. Note that the predominant dialect on the White Mountain Apache reservation is White Mountain Apache.

The 1990 U.S. Census reports an estimated 12,390 speakers of 'Apache'. It is unclear to what extent this figure accurately distinguishes Western, Mescalero-Chiricahua, Jicarilla, Plains, and Lipan Apache. An estimated 303 respondents also identified themselves specifically as 'San Carlos' speakers. Census estimates also indicate that over 17,000 respondents identified themselves as speakers of 'Indian' or 'American Indian', and approximately 1,500 respondents identified themselves as 'Athabaskan' speakers. The particular languages spoken by these individuals are not known. A fair estimate of the total number of Western Apache speakers at present is 14,000 - 15,000. The Census figures are listed with commentary by Broadwell (1995). The figures are from an unpublished Census report: Language Spoken at Home and Ability to Speak English for United States, Regions and States: 1990 (CPH-L-133).

Relatively few materials are presently available on Western Apache phonetics. There are two dissertations on Western Apache phonetics and/or phonology: Greenfeld (1972), a study of White Mountain phonetics and phonology, and Durbin (1964), a study of San Carlos phonology and morphology. De Reuse (1994) provides an introduction to Western Apache phonetics and Hill (1963) and de Reuse (1993) discuss dialect variation within Western Apache. Several works on Navajo phonetics are relevant to the present study including deJong & McDonough (1994), McDonough & Austin-Garrison (1994), McDonough & Ladefoged (1993), and McDonough, Ladefoged & George (1993). Additionally, there is discussion of Navajo and Western Apache sounds/orthography in Young & Morgan (1987) and the White Mountain Apache Culture Center (1972a,b), respectively. (See de Reuse (1994) and Potter (1997) for more comprehensive discussion of Western Apache linguistic materials.)

2. Present study

The present study is based on a list of approximately 150 Western Apache words that was designed to illustrate several of the principal phonetic features of the language, including voice-onset-time and closure duration of the stops, spectral properties of the fricatives, and duration and quality of the vowels. The word list was checked and recorded by one male speaker, M1, in Los Angeles. The remainder of the fieldwork was conducted in Tucson, Arizona, and in the towns of Bylas, Peridot and San Carlos on the San Carlos reservation. The word list was elicited from a total of nine speakers, four women and five men. The speakers represented a number of different dialects within Western Apache including San Carlos, White Mountain, Cibecue, and two different varieties spoken in Bylas. Data from one of the female speakers could not be used; thus, this paper presents data from three female and five male speakers. Recordings were made using a headmounted noise-canceling microphone, which ensured an approximately 45 dB signal to noise ratio. Three speakers were recorded outdoors. Speakers repeated each word twice, while recordings were made on a Sony Digital Audio Tape Recorder at a sampling rate of 48 kHz. Upon return to the UCLA Phonetics Laboratory, the data that were to be used for acoustic analyses were transferred to the Kay Computerized Speech Lab and down sampled to 16 kHz for analysis.

3. Consonants

The consonant phonemes of Western Apache are shown in Table 1. IPA symbols are used in the text, except for those delimited by angled brackets, which represent the orthographic conventions prevalent in the Athabaskan literature, shown in Table 2.

Table 1. The consonants of Western Apache as investigated in the present study.

	Bilabial	Alveolar	Palato-alveolar	Velar	Glottal
Stops	**p**	**tʰ t** **t’**		**kʰ k** **k’**	**ʔ**
Affricates		**tsʰ ts** **ts’**	**tʃʰ tʃ** **tʃ’**		
Nasals	**m**	**n**			
Fricatives		**s z**	**ʃ ʒ**	**x ɣ**	**h**
Laterals		**ɬ l**			
Lateral affricates		**tɬ tl** **tɬ’**			
Approximants	**w**		**j**		

Table 2. Conventional Athabaskan orthography for the consonants of Western Apache.

	BILABIAL	ALVEOLAR	PALATO-ALVEOLAR	VELAR	GLOTTAL
STOPS	**b**	**t d** **t'**		**k g** **k'**	**'**
AFFRICATES		**ts dz** **ts'**	**ch j** **ch'**		
NASALS	**m**	**n**			
FRICATIVES		**s z**	**sh zh**	**x gh**	**h**
LATERALS		**ł l**			
LATERAL AFFRICATES		**tł dl** **tł'**			
APPROXIMANTS	**w**		**y**		

Note that several sounds and/or sound combinations present in Western Apache are not indicated in Table 1. Table 1 does not include the prenasalized voiced stops [**mb**] and [**nd**], which occur as variants of /m/ and /n/. The glottalized nasals ['**m**], ['**n**] and prenasalized ['**nd**] are not included as they occur only in words made up of more than one morpheme. Labialized [**hw**] and [**kw**] are considered as the equivalent of phonemes in the orthography, but we did not investigate them in the present project. Note also that while listed separately in Table 1, the phonemic distinction between /x/ and /h/ is of some debate (cf. de Reuse 1994), and [w] is possibly an allophone of /ɣ/ (Greenfeld 1972, de Reuse 1994). These sounds, which are distinguished in the Western Apache orthography, were included in the present study. The presence of /m/ in the table is questionable, as virtually all occurrences of this sound are in loan words. The status of [$\mathbf{p}^{h}$] will be discussed in section 3.3. Finally, note that the release of the unaspirated affricates is typically voiceless although there may be some tokens with slight voicing during the fricative phase.

3.1. Unaffricated stops

Western Apache unaffricated stop consonants are produced at three places: bilabial, alveolar, and velar. At the alveolar and velar places of articulation there are three possibilities: aspirated, ejective and unaspirated. (The voiceless unaspirated alveolars are characteristically realized as taps in intervocalic environments other than in root-initial position.) The bilabial stops are more restricted. Ejective bilabial stops do not occur, and aspirated bilabial stops are rarely attested, surfacing primarily in borrowed words. Figures 3 - 5 illustrate the alveolar stop series for one speaker.

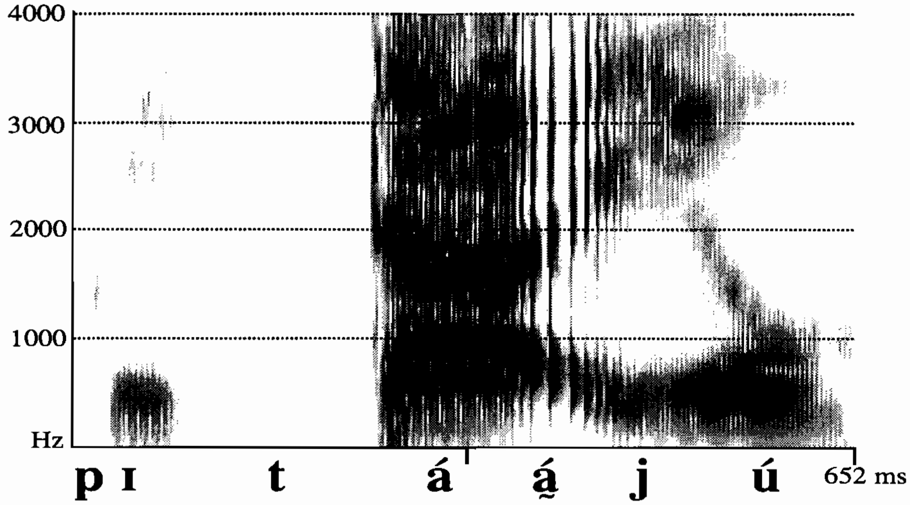


Figure 3. Voiceless unaspirated stop in the word **pitáʔju** ‘at its edge’ as produced by speaker F1.

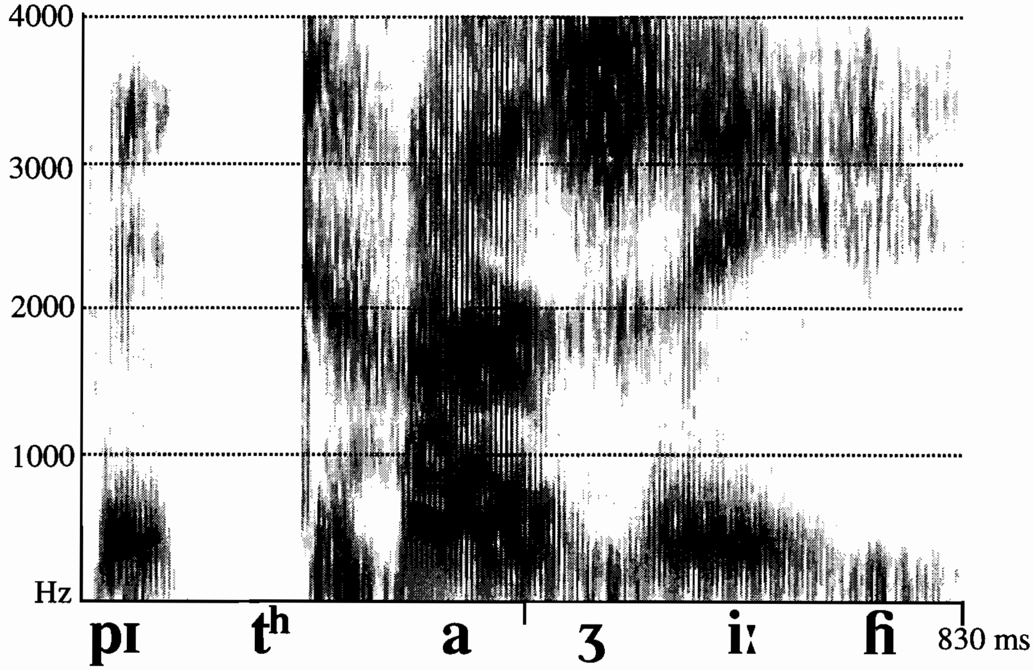


Figure 4. Voiceless aspirated stop in the word **pit^{h}aʒiːh** ‘his turkey’ as produced by speaker F1.

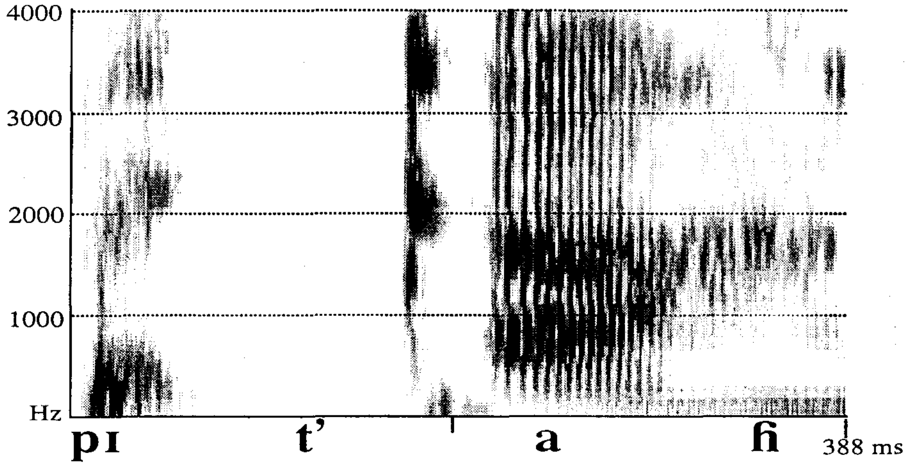


Figure 5. Ejective stop in the word **pit'ah** 'next to him' as produced by speaker F1.

In addition, although not previously recognized for Western Apache, the present study investigates the possible presence of contrastively voiced stops in the language. As reported in Section 3.3, a voiced alveolar stop, contrasting with the unaspirated alveolar stop, was identified in the speech of two Western Apache speakers.

3.1.1. Voice onset time (VOT)

Voice-onset-time (the period from consonant release to the onset of voicing of the following vowel, abbreviated VOT) was measured for the stops using the waveform in conjunction with a spectrogram. VOT measurements were made for the boldface consonants appearing in the words in Table 3. Consonants were measured in two intervocalic contexts, in both of which the stop was preceded by the high front vowel [ɪ]. The contexts are distinguished by the vowel following the stop, high front [ɪ] or [i] versus low [a].

VOT values were found not to differ as a function of whether they appeared before low or before high vowels. Mean values averaged over all speakers and all consonants were in fact 43 ms in both environments. Nor did individual speakers show substantially different patterns. For this reason, the other analyses of VOT discussed below collapse results for all speakers and both contexts in which the target stop appeared.

Table 3. Words used to measure VOT and closure duration for unaffricated stops.

Stop Consonant	Before [i] / [ɪ]		Before [a]	
p	**pɪpɪt**	**his stomach**	**ɪpan**	**buckskin**
t	**pɪtɪɬ**	**his blood**	**pɪtá'jú**	**at its edge**
k	**pɪkɪsh**	**his cane**	**pɪkat**	**his cedar**
tʰ	**pɪtʰɪsko**	**more so**	**pɪtʰaʒi**	**his turkey**
kʰ	**ɬɪkʰɪʒ**	**spotted**	**ɪkʰaɬ**	**animal hide**
t'	**pɪt'iːs**	**his cottonwood tree**	**pɪt'ah**	**near him**
k'	**pɪk'ɪsn̩**	**his brother**	**ɪk'ah**	**fat**

VOT for the three manners of stop consonants (aspirated, ejective, and unaspirated) appear in Figure 6. Bilabials were omitted from this comparison, due to the lack of aspirated and ejective bilabials in the data set. Thus, data in Figure 6 includes only alveolar and velar stops. All three manner classes differ significantly from one another in terms of VOT. VOT is longest for the aspirated stops, intermediate in length for the ejective stops, and shortest for the unaspirated stops. The effect of manner of articulation on VOT was found to be highly significant at the $p<.0001$ level according to a two-factor analysis of variance with place of articulation and manner of articulation as independent variables: F (2, 185) = 97.831, p<.0001. Additionally, all of the differences between individual manners of articulation were found to be highly significant at the p<.0001 level according to Fisher's PLSD post hoc tests.

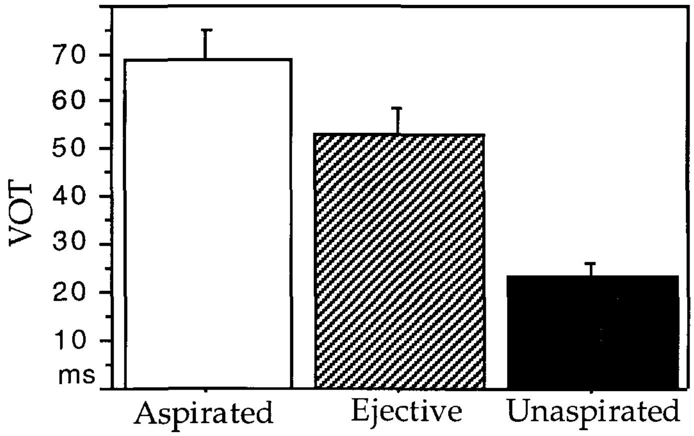


Figure 6. VOT values for aspirated, ejective, and unaspirated stops (velars and alveolars).

VOT was also highly influenced by place of articulation according to the analysis of variance: F(1, 185) = 41.485, p<.0001. In a series of unpaired t-tests, velars and alveolars were compared within the ejective and aspirated series, while velar, alveolar, and bilabial stops were compared within the unaspirated series. Results for all three manners of articulation were similar. Velars had significantly longer VOT values than alveolars (p<.0001), as shown in Figure 7, which presents results for ejective, unaspirated, and aspirated stops.

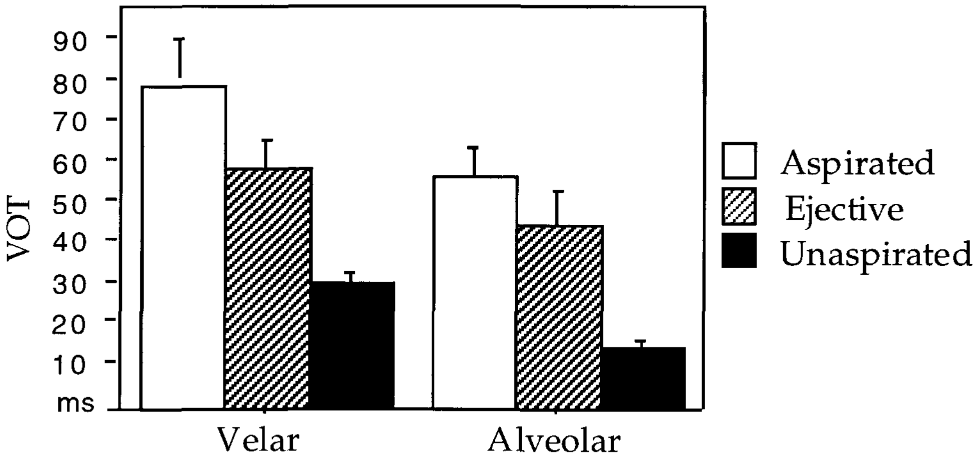


Figure 7. VOT for velar and alveolar stops classified by manner of articulation.

Alveolars, in turn, had significantly greater VOT values than bilabial stops in the unaspirated series (p=.0192), the only class of stops for which comparison could be made. Though statistically significant, this difference was quite small, only a few ms. VOT values for the three places of articulation in the unaspirated stop series are shown in Figure 8.

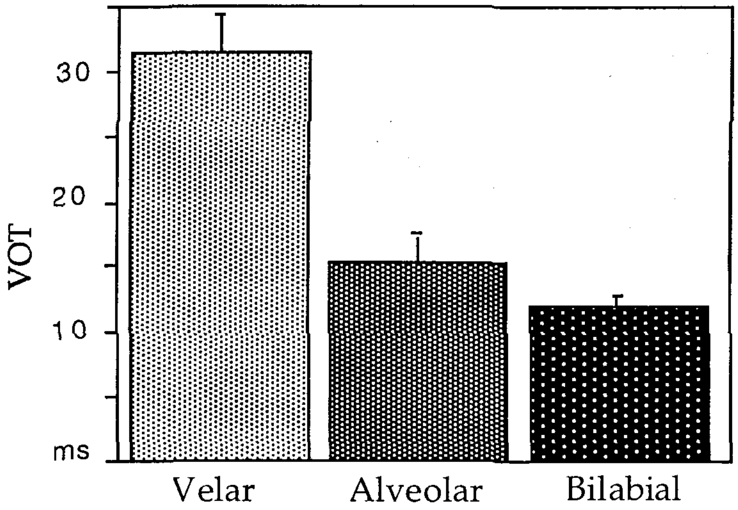


Figure 8. VOT values for velar, alveolar, and bilabial stops in the unaspirated stop series.

The VOT patterns found in Western Apache stops bear close resemblance to those found in other languages of the world (cf. Cho and Ladefoged in press, for data on 18 languages). Figure 9 shows comparative data for alveolar and velar stops in Apache and three other languages in the Na-Dene phylum. All these languages were recorded and analyzed in the same way, as part of the UCLA endangered languages project. The pattern, aspirated stops having a longer VOT than ejectives, which are in turn longer than unaspirated stops is found in Navajo (McDonough and Ladefoged 1993) and Tlingit (Maddieson et al. 1996). In Hupa (Gordon 1996), however, VOT values for the dental ejectives are slightly longer than for the dental aspirates. Although the direction of the VOT difference between aspirated and ejective stops is not entirely consistent across these four languages, the fact that ejectives and aspirated stops consistently differ suggests that VOT is perhaps used as a perceptual cue to differentiate the two classes of stops. However, further research on a greater number of genetically diverse languages, in addition to perception experiments, would be necessary to substantiate this hypothesis.

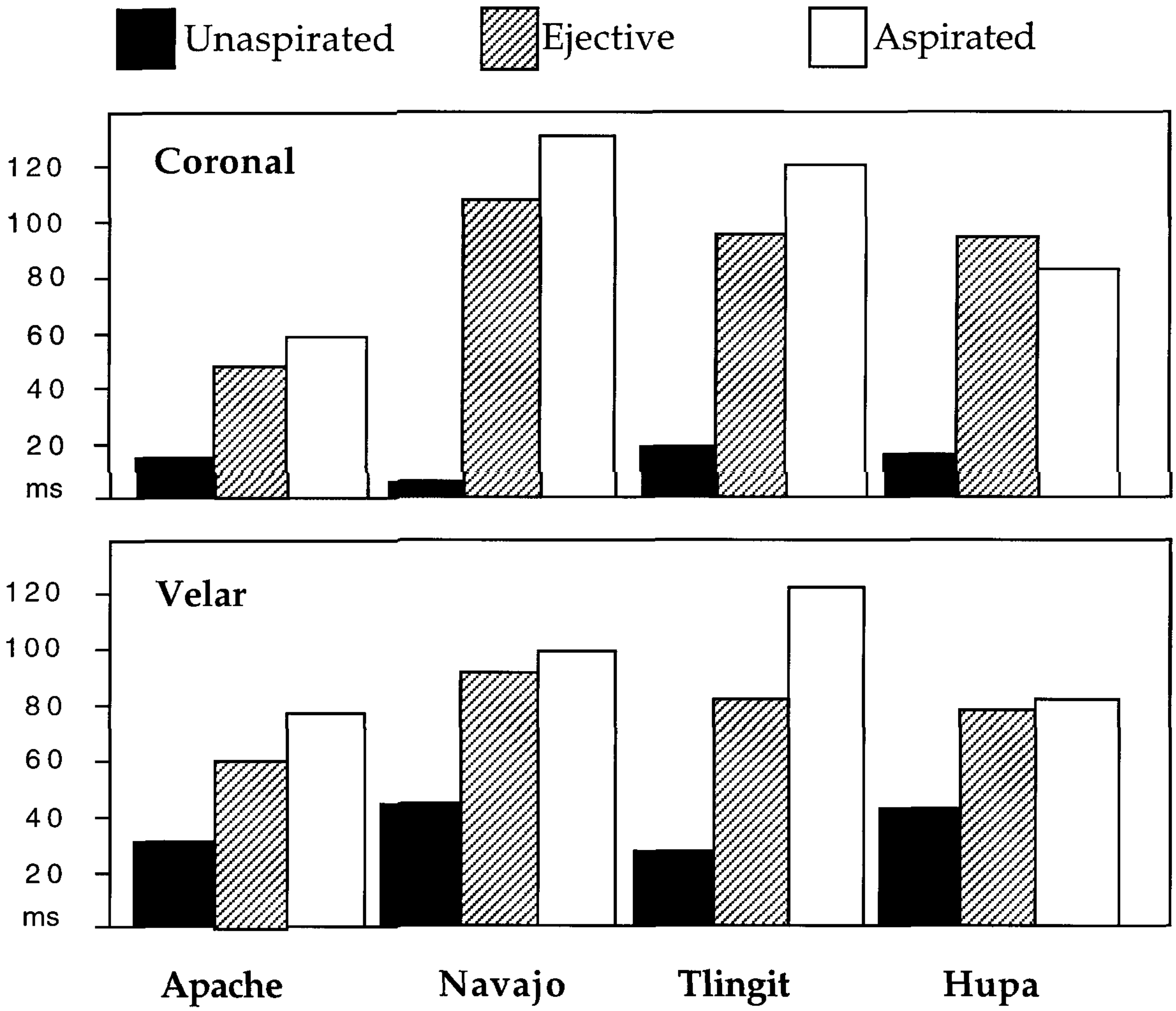


Figure 9. Comparative VOT for unaspirated, ejective and aspirated stops in four languages. Data from the present study, McDonough and Ladefoged (1993) for Navajo, Maddieson et al. (1996) for Tlingit, and Gordon (1996) for Hupa.

A further result of interest in the present study was that female speakers had significantly longer VOT values according to an unpaired t-test (p=.0003) than male speakers for both the ejective and aspirated stops: females 48 ms and males 40 ms, averaged over all places and manners of articulation. A gender dependent difference was not found in the unaspirated stops, probably because of the small range of VOT values for the unaspirated series which is involved in a phonemic contrast with the aspirated stop series. VOT values for each consonant separated by gender appear in Figure 10.

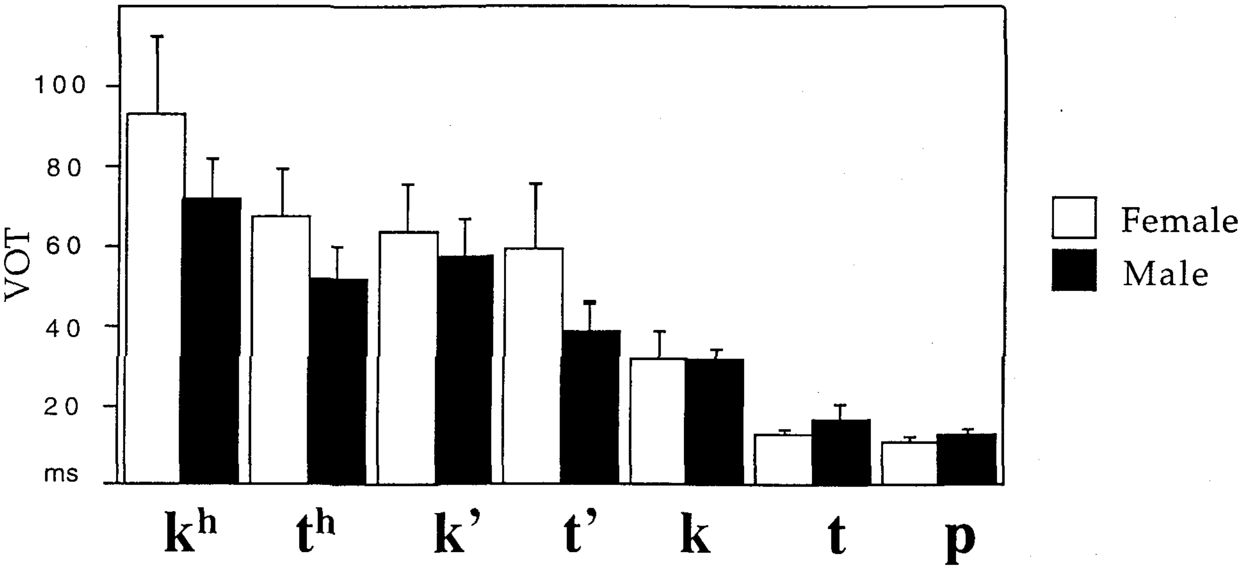


Figure 10. VOT values for all unaffricated stops divided by gender.

3.1.2. Closure duration

Closure duration values for the boldfaced stops in Table 2 were also measured from the waveforms. Measurements were taken from the onset of the stop closure to the point of the release. Closure duration values are compared for aspirated, ejective, and unaspirated stops produced at the alveolar and velar places of articulation. Bilabials were excluded from the analysis due to their absence in the ejective and aspirated series.

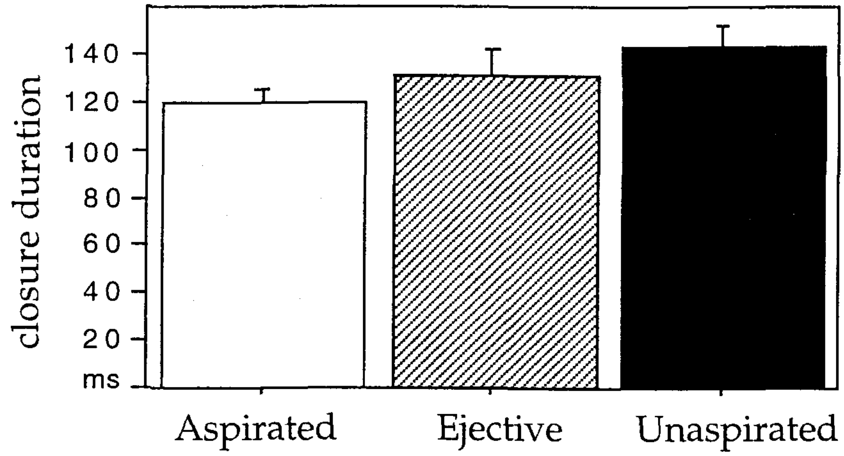


Figure 11. Closure duration for aspirated, ejective, and unaspirated stops (velar and alveolar places of articulation).

An analysis of variance for closure duration of velars and alveolars (the only places of articulation attested for all manners of articulation) with manner and place of articulation as independent variables indicated a highly significant effect of manner (F(2, 184) = 6.865, p=.0013) and a weaker but significant effect of place (F(1,184) = 5.161, p=.0243) on closure values. Closure values (shown in Figure 11) differed between all three manners of articulation according to Fisher's PLSD post hoc tests, though the only statistically significant difference was between aspirated and unaspirated stops (p=.0003), with the aspirated stops displaying shorter closure durations than the unaspirated stops. Ejectives had closure durations which were intermediate between those of the aspirated and unaspirated stops, although these differences did not quite reach statistical significance: aspirated vs. ejective, p=.0784; ejective vs. unaspirated, p=.0583.

The unaspirated series has the longest closure durations but the shortest VOT values. Conversely, the aspirated series displays the shortest closure durations but the longest VOT values. The ejectives have both VOT and closure values that are intermediate between those of the unaspirated and aspirated series. A similar inverse correlation between VOT and closure duration was found according to place of articulation. Velars had significantly shorter closure duration values than alveolars (p=.0250), as shown in Figure 12 for all three manners of articulations.

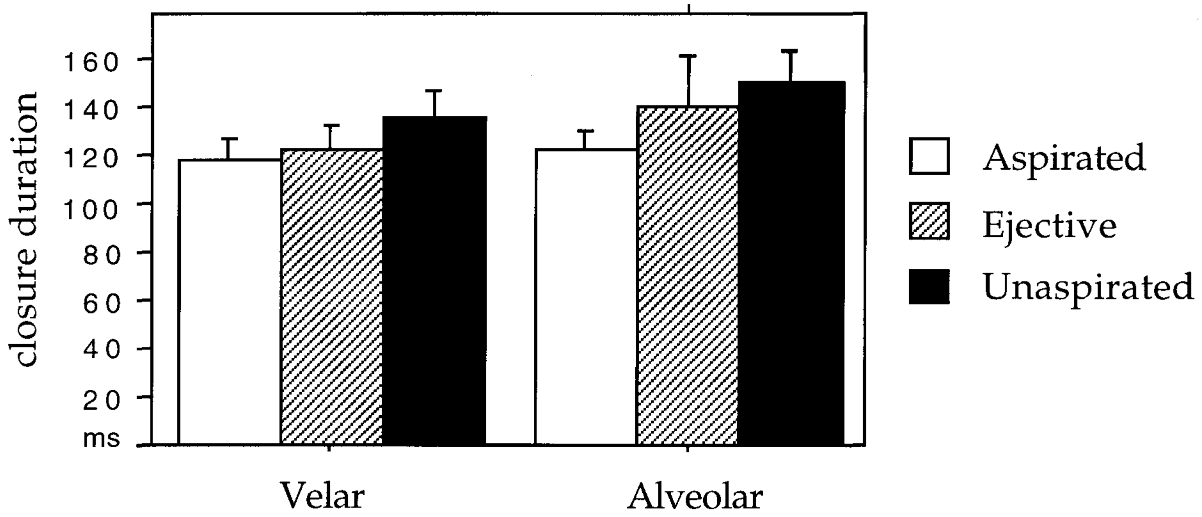


Figure 12. Closure durations for velar and alveolar stops classified by manner of articulation.

In the unaspirated stop series, the only stop series containing bilabials, the only significant difference due to place of articulation was between bilabial stops and velar stops, according to an unpaired t-tests. Results for the unaspirated stops appear in Figure 13.

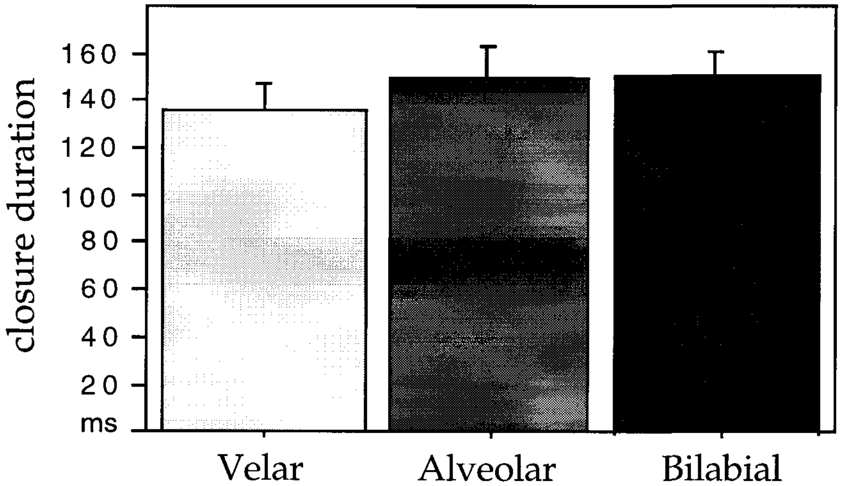


Figure 13. Closure durations for velar, alveolar, and bilabial stops in the unaspirated series.

As shown in Figures 14 and 15, if closure and VOT values are summed together for each place of articulation, the combined values for each of the three places (two in the case of ejectives and aspirated stops) of articulation turn out to be approximately the same. This is shown in Figure 14 for the velar and alveolar aspirated stops, and in Figure 15 for all three places of articulation in the unaspirated series.

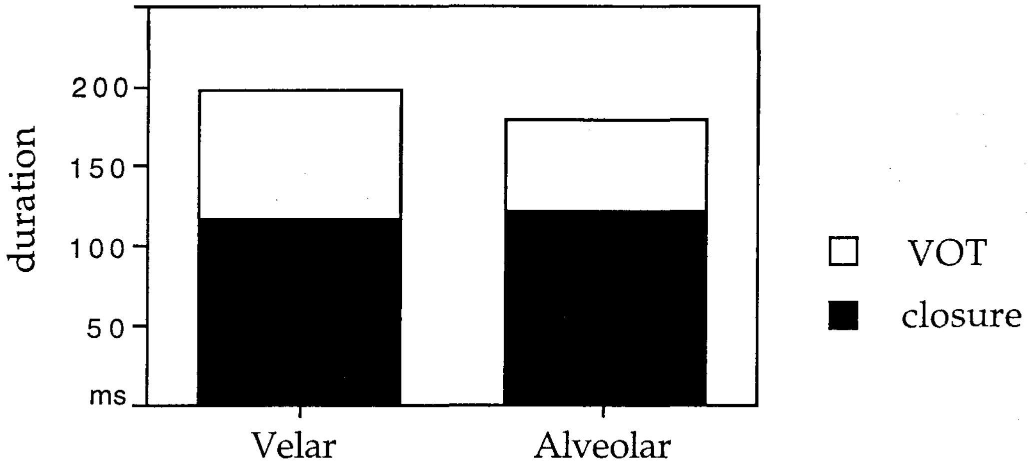


Figure 14. Closure duration and VOT values for velar and alveolar aspirated stops.

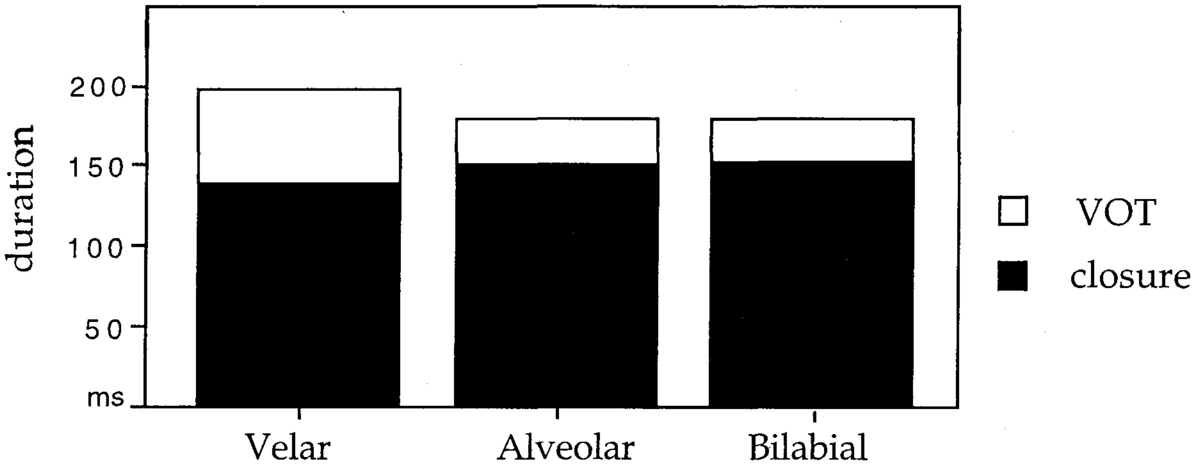


Figure 15. Combined closure duration and VOT values for unaspirated velar, alveolar, and bilabial stops.

The inverse correlation between VOT and closure duration for stops at different places of articulation appears to be a general cross-linguistic property attested in a large number of languages. Maddieson (1997) offers a possible explanation of the place-dependent nature of both the closure duration and VOT suggesting that the duration of the vocal fold opening may be fixed, with different durations for aspirated and unaspirated stops. When the closure duration is relatively longer, the VOT becomes relatively shorter (and vice versa). As he puts it: "There is an abduction-adduction cycle of the vocal cords for voiceless stops which is longer in duration than the closure and has a constant time course, anchored to the onset of closure" (Maddieson 1997: 621).

A possible reason for the shorter closure duration for the velar stop than both the alveolar and the labial stops is that the seal for more posterior stops may be more difficult to hold in the face of increased air pressure (Maddieson 1997). When the oral closure is further back, the cavity behind the closure is smaller, and the air pressure reaches a maximum more quickly. The faster the increase in pressure, the shorter the closure is held, perhaps due to a biomechanical feedback mechanism. Following this logic, the seal for a velar will be maintained for a shorter period of time than the closure for an alveolar stop which in turn will be held for a shorter period of time than the bilabial stop. The Western Apache data are compatible with this account, as velars show the shortest closures, whereas bilabials have the longest closures, though differences in both closure duration and VOT between the bilabial and alveolar stops are negligible in the language.

A weaker inverse correlation between closure duration and VOT is also found if different manners of articulation are compared, as shown in Figure 16. Those stops with the longest VOT values, the aspirated ones, have the shortest closure durations. Conversely, the stops with the shortest VOT values, the unaspirated ones, have the longest closure durations. Ejectives are intermediate in terms of both VOT and closure duration.

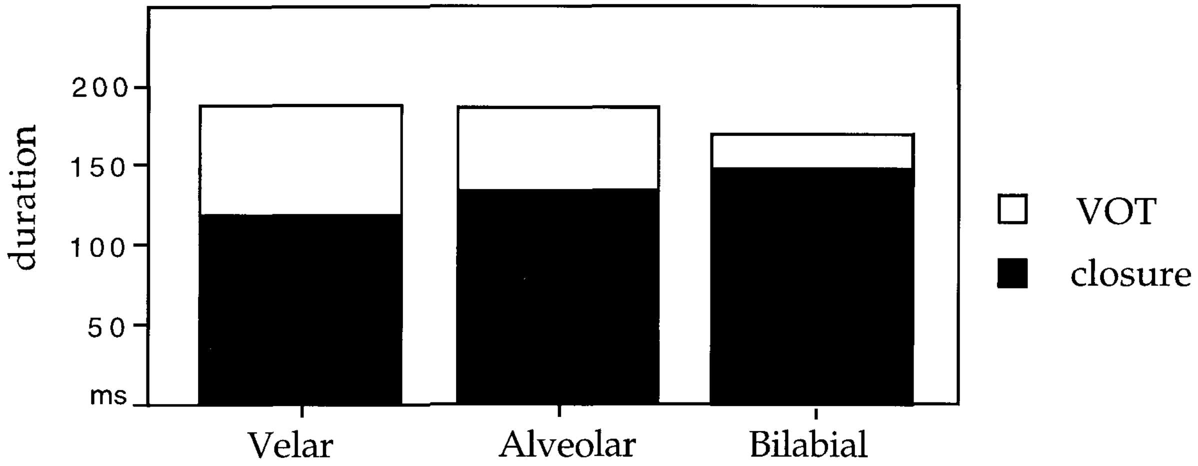


Figure 16. Combined closure duration and VOT values (ms) for aspirated, ejective, and unaspirated stops.

The manner dependent tendency for closure durations to be shortest when VOT values are longest also potentially finds an explanation similar to the one offered by Maddieson to account for the place-dependent inverse relationship between closure duration and VOT. The duration of the vocal fold opening may be fixed for each class of stops, so that lengthening the closure phase results in a compensatory shortening of the VOT phase. Conversely, shortening the closure results in a longer VOT period.

This explanation, with slight modification, will account for the inverse relationship between closure duration and VOT for the ejective stops. Ejectives require a closed glottis to be produced, unlike the aspirated and unaspirated stops which are fully voiceless in Western Apache and require an open glottis during production. If the duration of the glottal closure in ejectives is fixed, both the compensatory relationship between closure duration and VOT in the ejective series and the parallel between ejectives and voiceless stops are accounted for. Thus, it may be the case that not only is the duration of the glottal abduction-adduction cycle fixed, but that the adduction-abduction cycle is also fixed, with the cycles roughly equivalent to one another in duration.

Closure durations were found to be significantly longer at the $p<.0001$ level according to an unpaired t-test for the females than for the males: 147 ms for the females compared to 127 ms for the males. This result is similar to that found for VOT, which was also longer for female than for male speakers. Closure duration values are plotted in Figure 17 for all stop consonants separated according to the speaker gender.

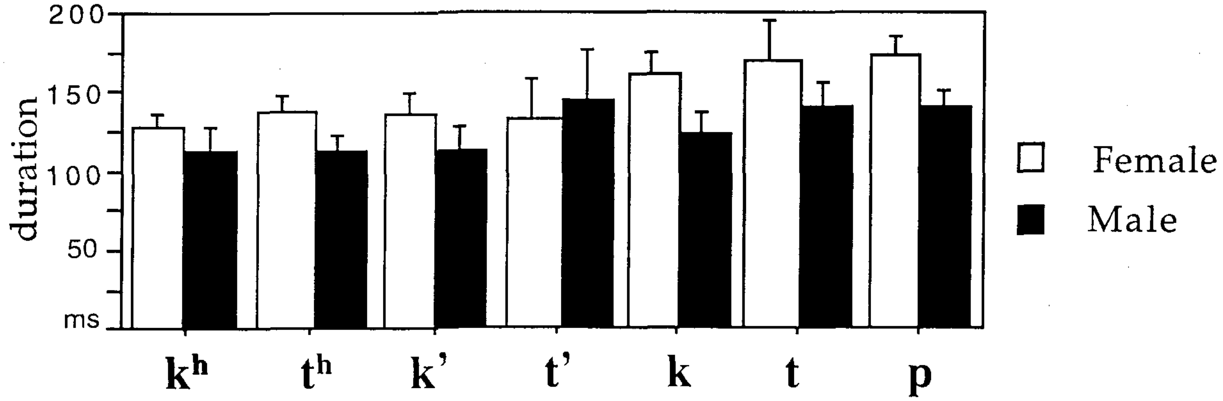


Figure 17. Closure duration values for all unaffricated stops divided by gender.

3.2. Affricated stops

The VOT and closure duration for affricated stops was measured using waveforms in conjunction with spectrograms. Measurements were made for the boldface consonants appearing in the words in Table 4. The affricates were measured in the same two intervocalic contexts as the unaffricated stops, with one exception. In the word [**najɬtɬʰat**] ‘he is greasing it’, the aspirated /**tɬʰ**/ occurs between a lateral fricative [ɬ]and [**a**].

Table 4. Words used to measure VOT and closure duration for affricated stops.

Affricate	Before [i] / [ɪ]		Before [a]	
ts	**pɪtsɪɬ**	his mountain	**naʔtiːtsaː**	he got up
tʃ	**pɪtʃɪʔɪ**	his heart	**pɪtʃat**	his leg
tɬ	**ɪk’ɛʔɪtlɪʃɛʔ**	stink bug	**tarɪtlaʔ**	lightning
tsʰ	**pɪtsʰɪʔ**	his daughter	**pɪtsʰaɬ**	his needle
tʃʰ	**pɪtʃʰɪʒ**	**his firewood**	**pɪtʃʰan**	its feces
tɬʰ	**tɪtɬʰɪd**	shaky	**najɪɬtɬʰat**	he is greasing it
ts’	**pɪts’ɪlɛ**	his cane	**tɪts’ak**	strong (hard but flexible)
tʃ’	**pɪtʃ’ɪt**	his blanket	**harɪtʃ’aːt**	it is oozing
tɬ’	**pɪtɬ’iːʃ**	his snake	**pɪtɬ’akaɬ**	her dress

Voicing for the unaspirated laterally released affricate typically commences approximately in the middle of the lateral, though sometimes even earlier, as in the exemplar in Figure 18, which also illustrates an alveolar tap. The acoustic result is a stop followed by a lateral that is voiceless for its first half and voiced for its second half. The unaspirated laterally released affricate in Navajo has a similar realization (McDonough and Ladefoged 1993).

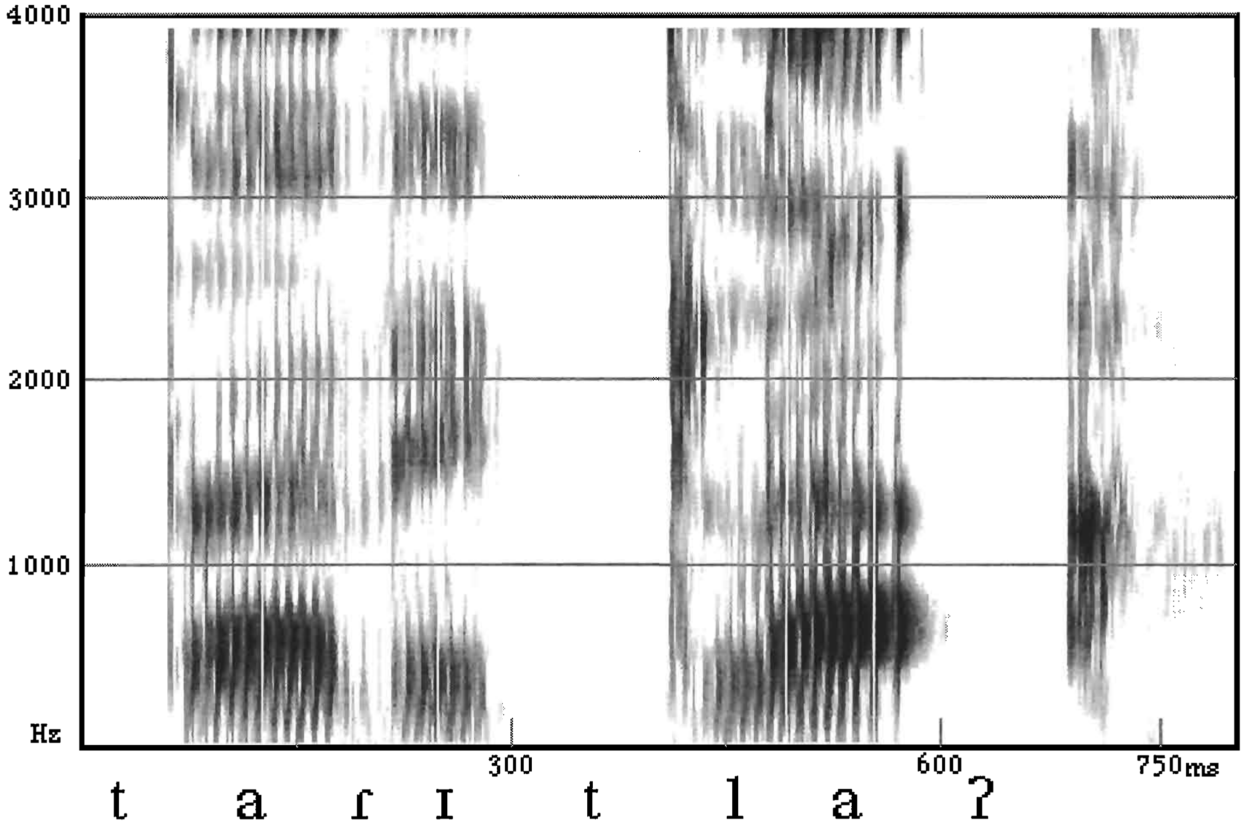


Figure 18. Unaspirated laterally released affricate in the word **taɾɪtlaʔ** ‘lightning’ as produced by speaker M3.

3.2.1. VOT of the affricated stops

An analysis of variance of VOT was performed for the affricated stops, taking place of articulation, manner of articulation, the following vowel, gender of the speaker, and the repetition number as independent variables. Four of these five factors were found to exert a highly significant influence on VOT values: place of articulation, F (2, 200) = 62.521, p<.0001; manner of articulation, F (2, 200) = 95.660, p<.0001; the following vowel, F (1, 100) = 21.933, p<.0001; gender, F (1, 200) = 26.056, p<.0001. The remaining factor, the repetition number (i.e. the first or second repetition of each word uttered by a speaker), exerted a negligible influence on VOT. None of the interactions between factors was statistically significant at the p<.05 level except for the interaction between manner and gender: F (2, 200) = 4.882, p=.0085.

Figure 19 graphically depicts the effect of manner of articulation on mean VOT values. Following the pattern of the unaffricated stops, the aspirated affricates are characterized by the longest VOT values on average, while the unaspirated affricates have the shortest average VOT values. The ejective is intermediate in terms of VOT. A Fisher’s PLSD post hoc test indicates that all manner of articulations are significantly different from one another in terms of VOT at the p<.0001 level.

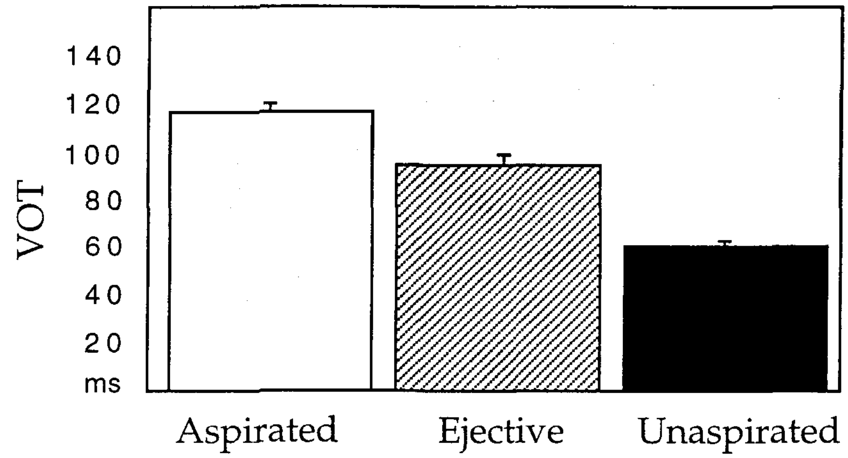


Figure 19. VOT for aspirated, ejective, and unaspirated affricates.

VOT values as a function of place of articulation are shown in Figure 20. As Figure 20 indicates, VOT is longest for the alveolar affricate, shorter for the palato-alveolar affricate, and still shorter for the laterally released affricate. All of these differences are statistically significant from one another at the p<.0001 level according to PLSD posthoc tests dependent on the ANOVA described above.

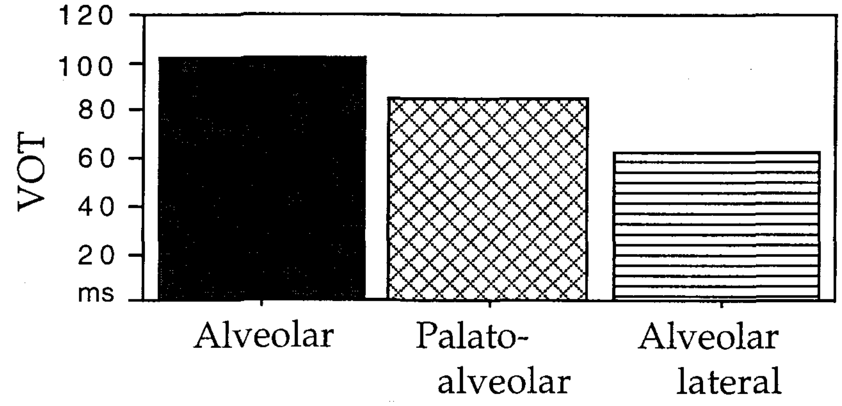


Figure 20. VOT for alveolar, palato-alveolar, and laterally released affricates.

The difference between the alveolar and the palato-alveolar is the opposite pattern from that seen in the unaffricated series, where the more front alveolar articulation has shorter VOT values than the backer velar articulation. The VOT results for the Western Apache affricates are similar to those for the unaspirated series in Navajo (McDonough and Ladefoged 1993). Interestingly, the Navajo differences are most noticeable in the unaspirated series; neither the affricated nor the ejective series of affricates show reliable differences in VOT between the alveolar, the lateral, and the palato-alveolars in Navajo. In Apache, the same VOT pattern (alveolars longest followed by palato-alveolars followed by laterals) is found for all manners of articulation, as Figure 21 demonstrates.

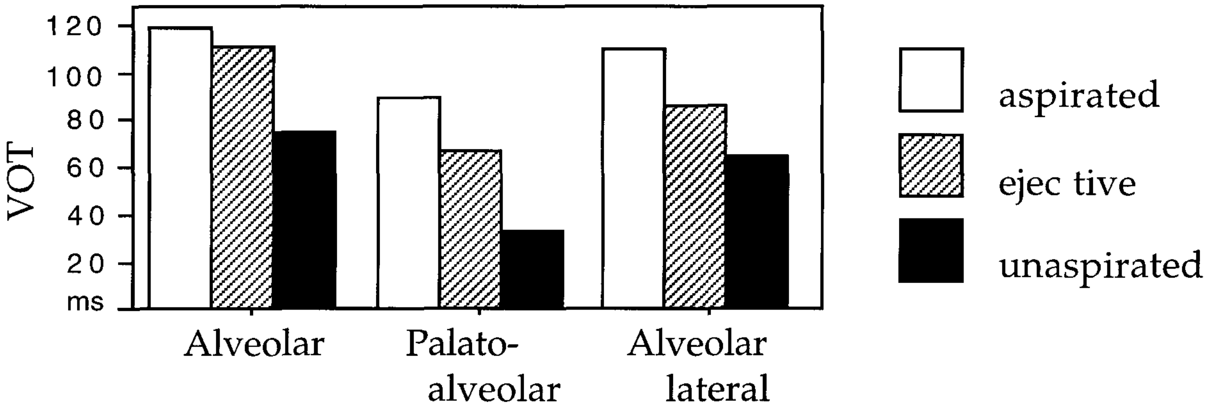


Figure 21. VOT for alveolar, laterally released, and palato-alveolar affricates classified by manner of articulation.

VOT values were slightly longer before high vowels than before low vowels, a difference of 15 ms averaged over all places and manners of articulation which was significant at p<.0001 according to a Fisher's PLSD posthoc test. VOT values were longer on average (17 ms) for female than for male speakers (p<.0001), a difference which was also seen in the unaffricated stop series in Figure 17. The first repetition was negligibly longer (6 ms) on average than the second repetition of each word. This difference was only barely significant according to a Fisher's PLSD posthoc test (p=.0337). Given the extremely small absolute difference in VOT duration and the large number of tokens in the analysis, this result does not appear to be robust.

3.2.2. Closure

For closure duration, an ANOVA was performed with place of articulation, manner of articulation, the following vowel, gender of the speaker, and the repetition number as independent variables. The largest effect was exerted by gender: F (1, 196) = 19.428, p<.0001, with smaller effects exerted by the following vowel (F (1, 196) = 13.531, p<.0003), manner of articulation (F(2, 196) = 6.717, p=.0015), and place of articulation (F (2, 196) = 4.456, p=.0128). Repetition number had no reliable effect on closure duration (F (1, 196) = .568, p=.4520). None of the interactions between factors was statistically robust except for a three-way interaction between place of articulation, the following vowel, and manner of articulation. This difference was fairly weak statistically, p=.0126.

Mean closure duration values as a function of manner of articulation appear graphically in Figure 22. Exactly opposite to the VOT results in Figure 20, closure duration is longest for the unaspirated series and shortest for the aspirated affricates. Only the difference between the aspirated and unaspirated affricates reaches statistical significance, p=.0031. The inverse relationship between closure duration and VOT as a function of manner parallels the result for the unaffricated stops and is similar to the result found for Navajo (McDonough and Ladefoged 1993).

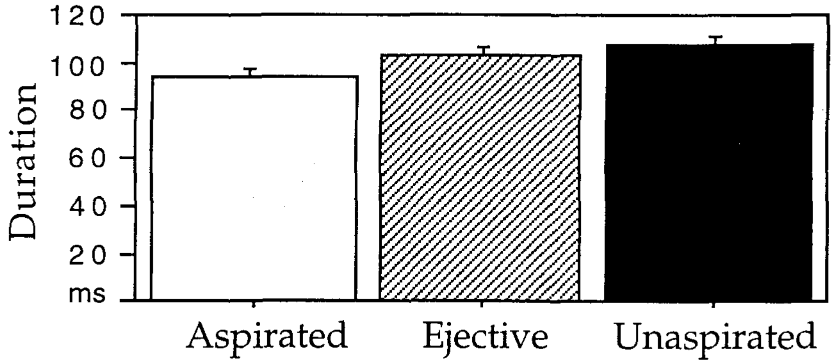


Figure 22. Closure duration for aspirated, ejective, and unaspirated affricates.

Closure durations for the palato-alveolar affricates were significantly shorter than those for both the alveolar and lateral affricates as Figure 23 shows: palato-alveolar vs. alveolar, p=.0088 according to a Fisher's PLSD post hoc test; palato-alveolar vs. lateral, p=.0028. The difference between alveolars and laterals was minuscule and statistically insignificant, p=.6434.

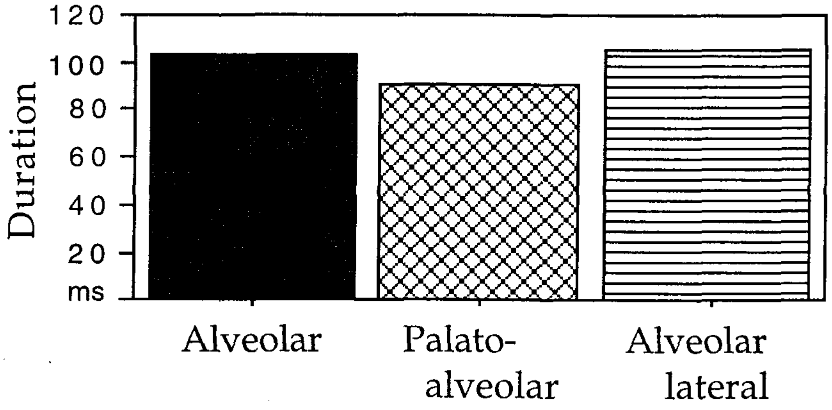


Figure 23. Closure duration for alveolar, palato-alveolar, and laterally released affricates.

These results for Western Apache differ from those for Navajo, where the laterally released affricate has much longer closure durations than both the alveolar and palato-alveolar affricates, at least in the unaspirated and aspirated series (McDonough and Ladefoged 1993).

Closure duration was slightly longer before /i/ than before /a/ (11 ms, p=.0028 in a Fisher's PLSD posthoc test), and closure duration was longer for females than for males (16 ms, p<.0001). Repetition number was shown to have no reliable effect on closure duration (2 ms, p=.5173).

3.3 Voiced Stops

Although not previously recognized for Western Apache, the present project investigated the possible presence of contrastively voiced stops in root-initial position. The words in Table 5 were identified as containing potentially voiced bilabial and alveolar stops and included in the word list recorded by the Western Apache speakers. No potentially voiced velar stops had been identified at the time the recordings were made.

Table 5. Words used to investigate the possible presence of voiced stops in Western Apache.

	Before [ɪ]		Before [a]	
Bilabial	**pɪpɪ́ʒ** (not voiced)	his knife	**ɬɪpaːh** (not voiced)	it is gray
Alveolar	**sɪdɪl** (voiced)	they are in position	**yaːyɪtah** (not voiced)	he habitually forgets it

As indicated in Table 5, only the word **<sidil>** was found to contain a voiced stop. Substantial dialect variation was present for this word, however. Two of the speakers pronounced the word as [**sɪnɪl**], the probable earliest common form (cf. Navajo **<sinil>**, Young & Morgan 1987), while one speaker pronounced it as [**sɪndɪl**]. Of the five speakers who pronounced the word with a non-nasal alveolar stop, only two had a voiced [**d**]. For these two speakers, however, the contrast between the voiced and voiceless unaspirated alveolar stops is clear. Figures 24 and 25 illustrate the contrast for speaker F3.

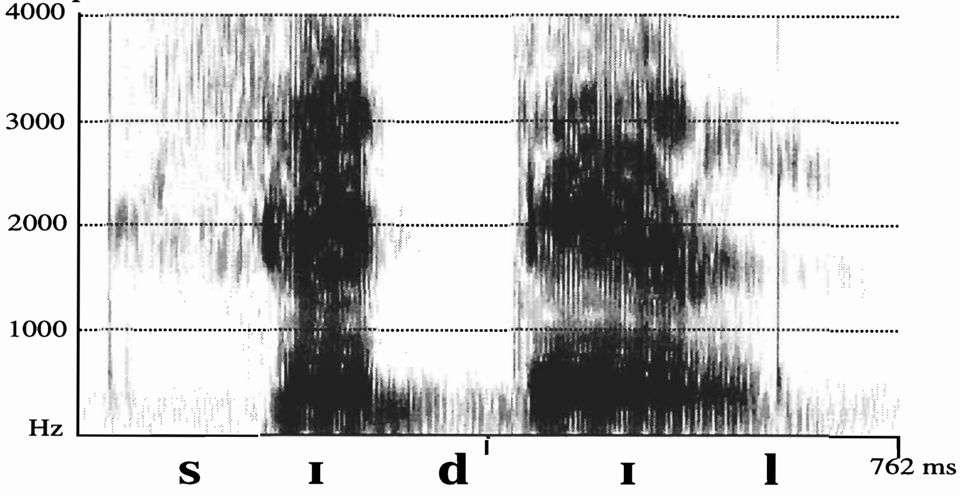


Figure 24. Voiced stop in the word **sɪdɪl** 'They (plural objects) are in position' as produced by speaker F3.

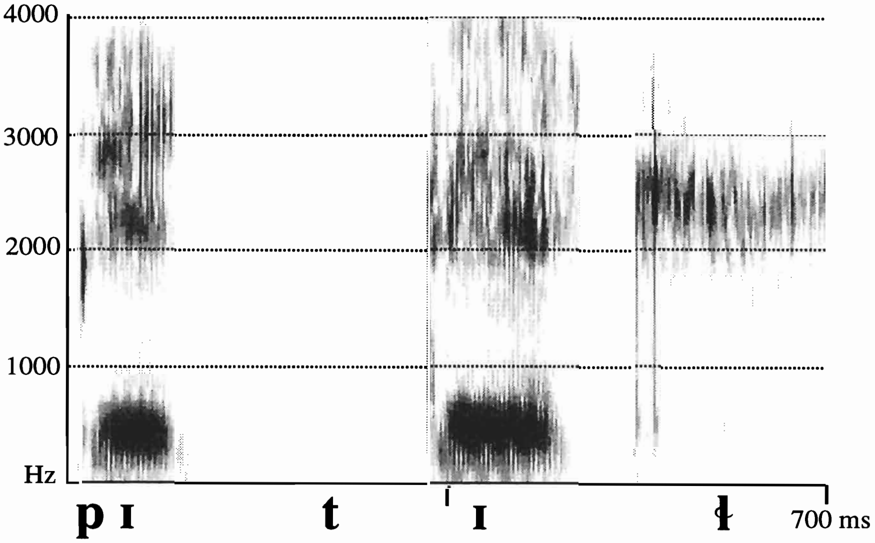


Figure 25. Voiceless unaspirated stop in the word **pɪtɪɬ** 'his blood' as produced by speaker F3.

The possible development of a contrastively voiced alveolar stop may be presently confined to those stops that vary dialectally between **[d]** / **[t]**, **[n]**, and **[nd]**, i.e. alveolar stops that are derived historically from **[n]**. It must be noted, however, that the word **[jaːjɪtah]** shows this same pattern with dialectal variants **[jaːjɪndah]** and **[jaːjɪnah]**, but was not pronounced with a voiced alveolar **[d]** by any of the recorded speakers.

3.3. Fricatives

Western Apache has four voiceless oral fricatives /ɬ/, /s/, /ʃ/, and /x/ (plus /h/, which will not be considerd further).The acoustic properties of these sounds were investigated in a set of words containing the fricatives in two intervocalic contexts, between a high front vowel and [a] and between two high front vowels. These words were digitized at a sampling rate of 22,050Hz using Kay MultiSpeech. The measured fricatives appear in boldface in the words in Table 6. (Note that one of the environments in which /x/ was measured was between /s/ and the high front vowel.) A 1,024 point window (approximately 46 ms) was centered around the middle of each fricative and an FFT spectrum of this window was calculated. Numerical spectra were then averaged together over the two tokens of each fricative appearing in the same environment for a given speaker. Because visual inspection revealed consistency across speakers in their spectral properties, the spectra for all speakers were averaged together for each fricative. These averaged spectra in the pre-[a] context appear in Figure 26 (female speakers) and Figure 27 (male speakers). (Spectra for the fricatives before the high front vowels were quite similar to those in the [a] context.)

Table 6. Words used to examine the spectral properties of Western Apache fricatives

Fricative	Before [ɪ]		Before [**a**]	
ɬ	**pɪɬɪt**	his smoke	**tʃ’ɪʔkotɪɬan̰**	centipede
s	**hɪsiː**	I will miss it	**ʃɪsanɛ**	my old lady
ʃ	**bɪʃɪʃ**	It stung him	**bɪʃaʃ**	his bear
x	**jɪzɪsxĩː**	he killed it	**bɪxaɬ**	his club

As Figures 26 and 27 indicate, the fricatives are for the most part well-differentiated in terms of their spectral characteristics. The strongest energy for /s/ occurs at frequencies above 4000Hz for the male speakers and above 6000Hz for the female speakers. The spectrum for /ʃ/ is characterized by a relatively sharp peak in energy at approximately 3500Hz for the males and 4000Hz for the females. /s/ and /ʃ/ are thus differentiated on the basis of the location of their strongest energy in the frequency domain: /s/ has more energy at higher frequencies than /ʃ/. /s/ also has a less sharp spectral peak than /ʃ/. /x/ is characterized by a very pronounced peak in energy below 2000Hz and then more diffuse peaks at about 4000 and 7000Hz. The spectral peak for /ɬ/ falls between that of /x/ and that of /ʃ/ at approximately 2500-3000Hz. In summary, the four oral fricatives of Western Apache can be distinguished from one another in terms of the location of their spectral peaks, and, to a less extent, in terms of the degree of sharpness of these spectral peaks. There is relatively little cross-linguistic data on the acoustic properties of fricatives. However, Jassem (1968) presents spectra from his speech of three of the four fricatives measured in Western Apache, /s/, /ʃ/, and /x/. The spectra of Jassem's fricatives bear close resemblance to those in Western Apache. /x/ has a sharp peak in energy below 2000Hz, /ʃ/ has an energy peak slightly higher at 3000-4000Hz, and /s/ has a peak at 4000Hz. Interestingly, Jassem also includes a spectrum of a dental fricative /s̪/ which actually bears closer resemblance to the Apache /s/ both in terms of location of spectral peak and degree of sharpness of the peak than Jassem's alveolar fricative /s/. Since Jassem's study is based on only a single speaker, however, we cannot infer with any certainty that the Western Apache /s/ is more dental than alveolar.

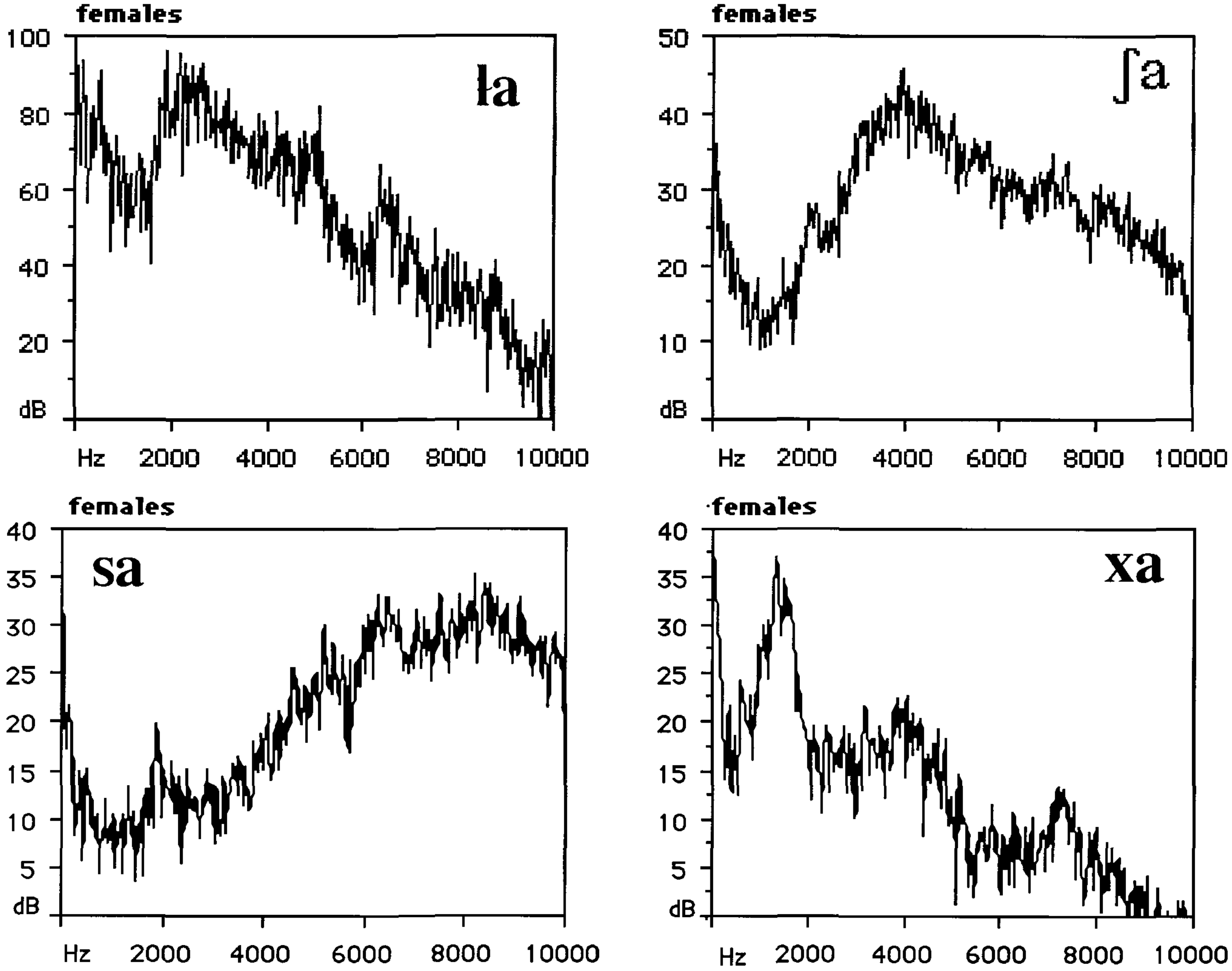


Figure 26. FFT power spectra of fricatives before [**a**] (female speakers)

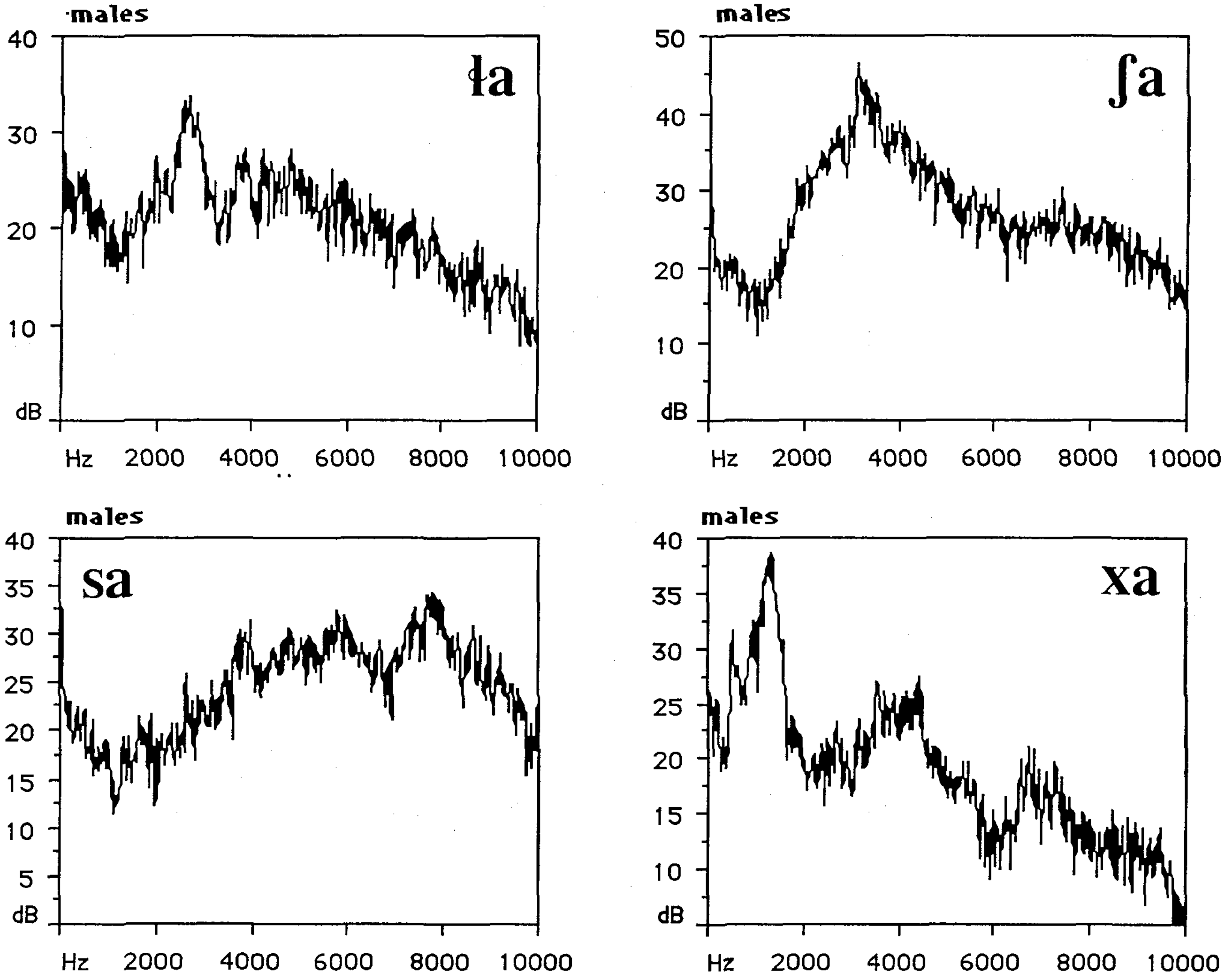


Figure 27. FFT power spectra of fricatives before [**a**] (male speakers)

4. Vowels

Like Navajo (cf. Hoijer 1945, McDonough & Austin-Garrison 1994), Western Apache has four phonemic vowel qualities: /**i**/, /**e**/, /**o**/, and /**a**/, each of which has two phonemic vowel lengths: short and long. Although the high back vowel [**u**] surfaces in several Western Apache words, Greenfeld (1972) demonstrates that [**u**] is an allophone of the phoneme /**o**/. /**o**/ is pronounced as [**u**] when preceding the vowel [**i**]. Long, high tone [**óː**] is pronounced as [**úː**] when preceded by [**t**^{h}], [**k**^{h}], [**s**], or [**j**] (de Reuse 1994). The Western Apache vowels appear in Table 7.

Table 7. The vowels of Western Apache.

Short			Long		
i			**iː**		
	e	**o**		**eː**	**oː**
	a			**aː**	

In addition, every vowel may be either oral or nasalized and may carry either high tone or low tone. (Current research by de Reuse and Tuttle (forthcoming) indicates the possible presence of a contrastive mid tone as well.) The contrastive use of tone is a common feature of Southern Athabaskan languages, including Navajo and Western Apache, as well as many Northern Athabaskan languages. Long vowels in Western Apache may also carry contour tones consisting of a high tone followed by a low tone or vice versa. Contour tones in Western Apache are fairly rare, however, and almost exclusively limited to morphologically complex words. The words analyzed for this paper appear in Table 8. Words illustrating the nasal vowel series were also recorded, but are not included in the present analysis.

Table 8. Words used to examine the quality, fundamental frequency, and duration of Western Apache vowels. Entries in parentheses were not pronounced as expected (see text).

Vowel	Short - Low Tone		Short - High Tone	
i	**tʃɪk**	blanket	**ʔɪɬk’ɪ́korɛstɔ̃**	to straighten it out
e	**k’ɛkowã**	home, ruins	**kotɛ’ɛ́ko**	finally
o	**bɛ́ːʃ tɪst’ɔk**	arrowhead	**(ʔɪk’án nást’óːk)**	dough
a	**jɪrɪts’ak**	he hears him	**pɪtsʰɪt’ákɛː**	top of his head
Vowel	Long - Low Tone		Long - High Tone	
i	**nákonɛtʃ’iːki**	locust	**na’ɪtɬ’iːki**	gopher
e	**taɬɛ́ɬt’ɛːko**	same way	**tijat’ɛ́ːkoʰ**	awful
o	**tʰú hajɪɬt’oːk**	he pumps water	**(nájɪɬt’oːk)**	he soaks it
a	**harɪtʃ’aːk**	it oozes	**hakon̩’áːkɛʰ**	corner

The goal of the list of words in Table 8 was to provide each oral vowel of the language in a minimally contrasting environment: between an ejective stop or glottal stop and the unaspirated velar stop. To the extent possible, words were chosen in which the vowel appeared as part of the root morpheme. The two entries in parentheses were not pronounced as expected from the list. The word selected to exhibit short, high tone /o/ was most typically pronounced with a long vowel, and the word selected to exhibit long, high tone /o/ was most typically pronounced with low tone. These items were excluded from the statistical analyses as appropriate. Note also that word final stops were unreleased, and stem final /k/ varied dialectally with /t/. This latter variation was present in the words for ‘blanket’, ‘he hears him’, ‘dough’, ‘locust’, ‘he pumps water’, ‘it oozes’, ‘gopher’, and ‘he soaks it’. Finally, short /e/ is rare in Western Apache and only five speakers recognized the word for ‘home, ruins’ and only three speakers recognized the word for ‘finally’.

4.1. Vowel duration

Vowel duration was measured from a waveform alongside a spectrogram. The duration of each vowel included the time from the onset of the first formant of the vowel to the offset of the first formant. Not surprisingly, phonemic long vowels were found to be significantly longer than phonemic short vowels ($p<.0001$). Long vowels averaged 187 ms collapsing all speakers and all vowel qualities, while short vowels averaged 89 ms. In terms of ratios, long vowels were thus 2.1 times longer than short vowels. This long-to-short ratio is smaller than the one found by McDonough and Austin-Garrison (1994) in their study of Navajo. They found that long vowels were 2.8 times as long as short vowels for monolingual speakers and 2.3 times as long as short vowels for bilingual speakers. (All the Western Apache speakers were bilingual.) Duration ratios between short and long vowels vary substantially across languages; the Western Apache ratio of approximately 2:1 falls squarely within the range of cross-linguistic variation (cf. Lehiste 1970 and Hubbard 1994 for representative values from other languages).

Figure 28 plots the duration of different vowel qualities, collapsing short and long vowels. The high vowels were found to be significantly shorter than all other vowel qualities, a familiar pattern from a number of languages (cf. Lehiste 1970 for a summary of the data). Neither of the mid vowels differed significantly from the low vowel in terms of duration, though /e/ was found to be significantly longer than /o/ (p=.0028).

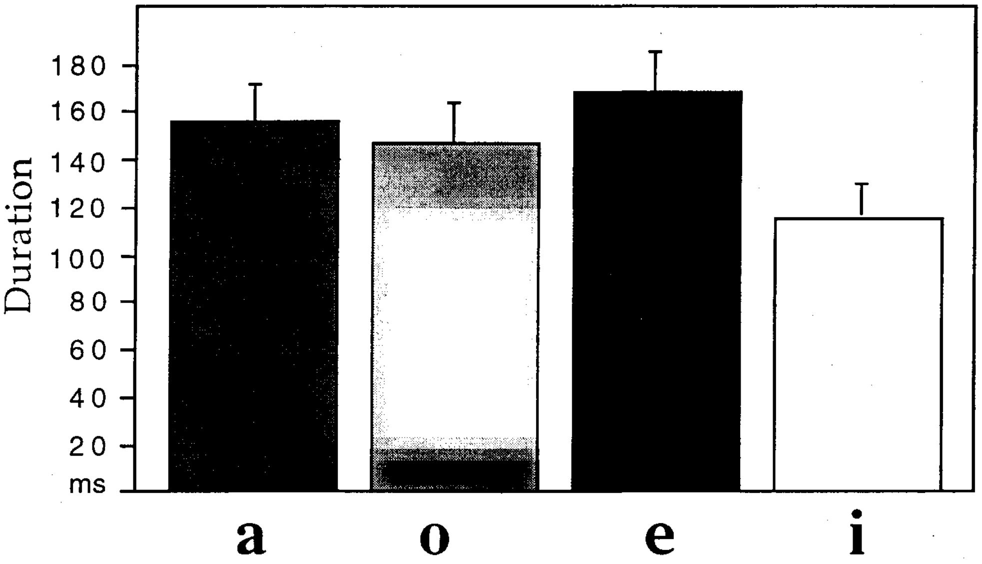


Figure 28. Duration of vowels of different qualities.

Vowels were also found to be significantly longer for the female speakers than for the male speakers (p=.0033). The significance of this result is due to the long vowels, which averaged 208 ms for the female speakers but only 174 ms for the male speakers. Short vowels were virtually identical in duration for the male and female speakers: 91 ms for the male speakers vs. 86 ms for the female speakers.

The confinement of the gender dependent length difference to the long vowels plausibly has similar motivations as the confinement of gender dependent length differences to aspirated and ejective stops. In both cases, it is the longer phonemic category, aspirated stops and phonemic long vowels, which displays duration differences as a function of gender. Manipulation of the length of the shorter category would potentially interfere with the perception of the phonemic length contrast, by making the shorter category more like the longer category. Lengthening the longer category, on the other hand, only enhances the percept of the phonemic length contrast.

4.2. Fundamental frequency

Fundamental frequency, the acoustic property defining tone, was measured using narrow band spectrograms. A measurement was taken from the tenth harmonic at a point with fairly level fundamental frequency and at approximately the mid point of each vowel. In cases where the tenth harmonic was not readily discernible, a measurement was taken from a lower harmonic. Measured values were divided by the number of the harmonic from which the measurement was taken, yielding a value for the first harmonic, i.e. the fundamental frequency. For male speakers, high tone vowels averaged 135 Hz, while low tone vowels averaged 118 Hz. For female speakers, high tone averaged 208 Hz, while low tone averaged 189 Hz.

Vowel quality did not have a statistically significant effect on fundamental frequency. /o/ had a slightly lower fundamental frequency (141 Hz) than /a/ (148 Hz) which in turn displayed a slightly lower mean value than /e/ (152 Hz) which in turn had a lower fundamental than /i/ (159 Hz). The only difference which reached statistical significance was the one between /o/ and /i/ (p=.0254).

Dividing the tokens according to whether they had high or low tone did not yield any additional significant results.

4.3. Vowel quality

The first three vowel formants were measured using an LPC analysis calculated over a 30 ms window using 12 coefficients. An FFT spectrum was also consulted to ensure that the measurements taken from the LPC were accurate. A spectrogram provided further corroboration of the accuracy of the LPC measures.

The first two formants from all tokens produced by the female speakers are plotted in Figure 29. Formant values for the male speakers are plotted in Figure 30. Ellipses encircle all points falling within two standard deviations of the mean, which is plotted as a large vowel in the middle of the ellipse. Formant values for each of the individual speakers appear in Table 8.

There is very little difference in the general form of the plots of the females' vowels and those of the males. The most salient difference between the two genders appears to be the location of long /**iː**/ relative to short /**i**/ in the vowel space. For the male speakers, long /**iː**/ is both much higher and much fronter than /**i**/. For the female speakers, the primary difference between /**i**/ and /**iː**/ lies in the front-back dimension, with /**iː**/ having a fronter articulation than /**i**/. There is a tendency for /**iː**/ to be slightly higher than /**i**/ for the female speakers, though this difference is much less striking than for the male speakers.

For the male speakers /**iː**/ is much higher than the long back vowels /**o**/ and /**oː**/. This difference is less apparent for the female speakers. For these speakers the difference between short /**i**/ and its short back counterpart /**o**/ is substantial, and in fact more salient than the corresponding difference for the male speakers. Whereas /**o**/ has first formant values approximately midway between those of /**i**/ and those of /**e**/ for the male speakers, /**o**/ has similar first formant values to /**e**/ for the female speakers.

In many languages with small vowel inventories, the high back vowel is not as high as the high front vowel. This is the case in Navajo (McDonough et al. 1993), which has a four vowel system /**i, e, a, o**/. Similarly Hupa (Golla 1970), another Athabaskan language, also lacks a high back vowel at least in its underlying phoneme inventory. In other language families the same situation is found. Banawá, an Arawakan language, spoken by about 75 speakers in northern Brazil (Ladefoged, Ladefoged and Everett 1997) has a four vowel system /**i, e, a, o**/, with qualities that are very similar to those of Apache. Wari' (Arawakan) has four vowels /**i, e, a, o**/ plus two front rounded vowels /**y,ø**/, but no high back vowel (MacEachern, Kern and Ladefoged 1997).

There are differences between the long and short vowels. The long back vowel /**oː**/ is backer and higher than its short counterpart /**o**/ for both female and male speakers, and the long low vowel /**aː**/ is slightly lower and backer than its short counterpart /**a**/. Comparison of short and long vowels thus indicates a general trend for the long vowels to be more peripheral in the vowel space than the short vowels. This is not the case, however, for the front mid vowel pair /**e**/ and /**eː**/. For most speakers, these vowels do not differ from one another in either front/back or height dimensions, as evidenced by the extensive overlap between their two ellipses and the close proximity of their means. Only two speakers (F3 and M2) do not display substantial overlap of short and long /**e**/ as shown in Table 9. For both of these speakers, short /**e**/ is backer than long /**eː**/, as reflected in the lower F2 values for /**e**/. First formant values for /**e**/ and /**eː**/ are virtually identical, however, suggesting that they do not differ in terms of tongue height.

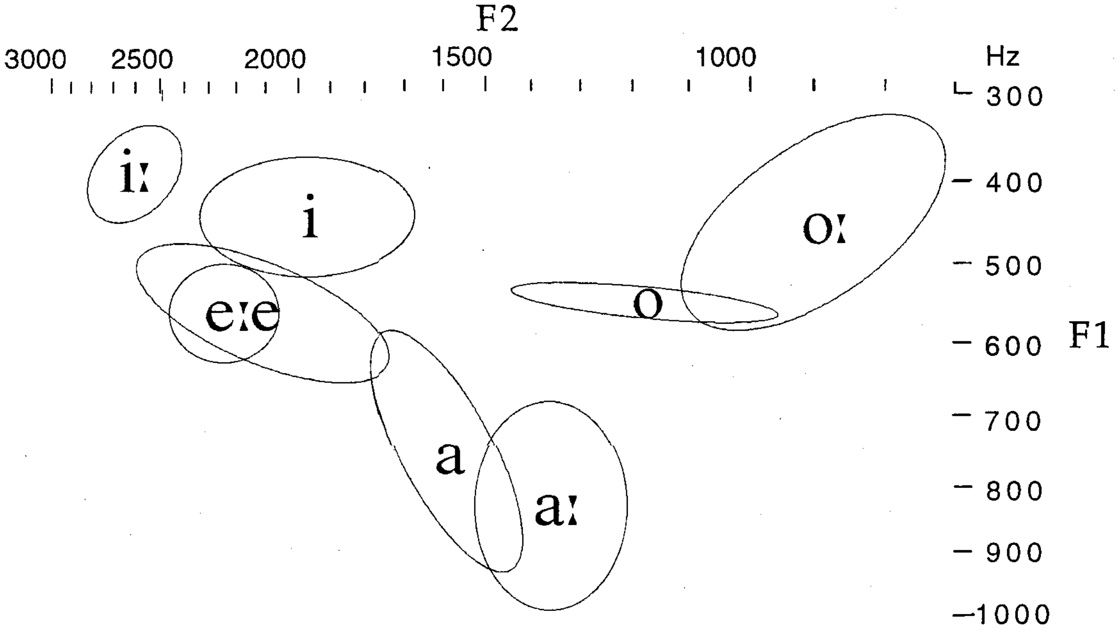


Figure 29. Plot of the first two formants for three female speakers.

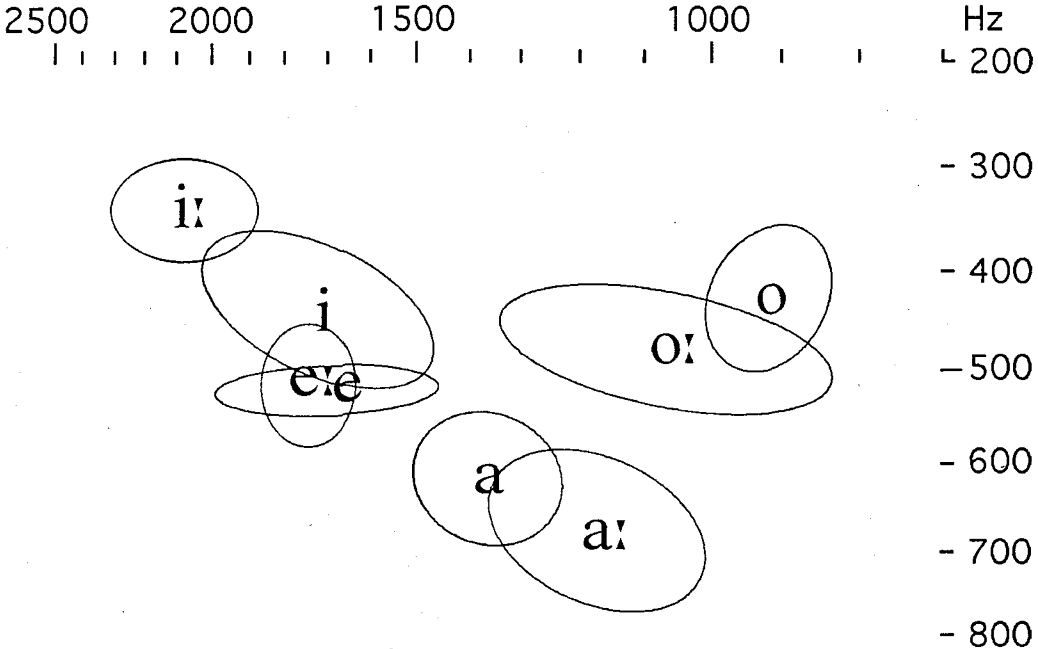


Figure 30. Plot of the first two formants for five male speakers.

Table 9. Mean values for the first three formants for five male and three female Western Apache speakers.

Spk	Short			Long			Spk	Short			Long		
	F1	F2	F3	F1	F2	F3		F1	F2	F3	F1	F2	F3
F1							**M2**						
i	463	1954	2812	413	2683	3367	**i**	468	1573	2412	362	2046	2440
e	536	2298	3183	551	2353	2853	**e**	518	1596	2114	532	1718	2281
a	858	1536	3041	913	1380	3078	**a**	582	1330	2091	674	1156	2302
o	556	1076	3168	436	821	3367	**o**	486	908	2201	468	940	2192
F2							**M3**						
i	426	1959	2788	408	2628	3087	**i**	376	1844	2284	316	1972	2289
e	----	----	----	559	2192	3013	**e**	----	----	----	486	1720	2133
a	669	1715	2926	819	1394	2807	**a**	614	1435	2729	624	1330	2367
o	541	1339	2532	422	908	2954	**o**	523	1009	2183	431	885	2174
F3							**M4**						
i	431	2078	2903	353	2550	3335	**i**	444	1779	2436	339	2092	2399
e	579	2003	2682	573	2247	2752	**e**	----	----	----	495	1711	2426
a	715	1583	2908	747	1371	3142	**a**	633	1348	2348	660	1105	2403
o	541	1220	2816	500	991	3027	**o**	495	1100	2238	385	945	2371
M1							**M5**						
i	458	1917	2550	339	2256	2518	**i**	445	1633	2417	339	2060	2578
e	523	1835	2385	532	1816	2473	**e**	523	1844	2366	514	1807	2376
a	601	1380	2509	679	1165	2619	**a**	651	1367	2477	711	1229	2573
o	431	1110	2431	413	935	2587	**o**	468	1257	2357	436	917	2509

6. Summary

In this paper, we have described several phonetic properties of Western Apache. The principal findings are as follows. For consonants, VOT was longer for aspirated stops than for unaspirated stops, whereas closure duration was longer for unaspirated stops than for aspirated stops. The ejectives had VOT and closure duration values intermediate between those of the aspirated and unaspirated stops. Place of articulation was also shown to affect both VOT and closure duration. VOT was longest for the velar stops, followed by the alveolars, followed in turn, by the bilabials. Conversely, closure duration was shortest for the velars, longer for the alveolars, and marginally still longer for the bilabials. There is some phonetic evidence for a four way manner contrast in the stop series between voiced, voiceless unaspirated, voiceless aspirated, and ejective stops, though voiced stops have a limited distribution relative to the other three categories. The four oral fricatives of Western Apache are well-differentiated in terms of the location of their spectral peaks in the frequency domain and their sharpness. /s/ has the greatest energy at frequencies above 4000Hz for the males and above 6000Hz for the females. /ʃ/ has a relatively sharp spectral peak at 3500Hz for the males 4000Hz for the females. /x/ has the most pronounced peak and the peak at the lowest frequency, about 1500Hz. /ɬ/ has a slightly flatter peak at frequencies intermediate between the peaks for /x/ and /ʃ/.

Concerning the vowels, long vowels tend to be slightly more peripheral than their short counterparts, except for the /e, eː/ pair which differed substantially in the first and second formant dimensions for only two speakers. As in other languages with small vowel inventories, the highest back vowel was more like [o] than [u]. Finally, female speakers differed from male speakers in having longer VOTs, closures, and vowel durations.

References

Broadwell, A. 1995. 1990 census figures for speakers of American Indian languages. *International Journal of American Linguistics*, 61.1, 145 - 149.

Cho, Taehong and Peter Ladefoged, in press. Variations and universals in VOT: Evidence from 18 endangered languages. *Journal of Phonetics.*

deJong, Ken and Joyce McDonough. 1994. Tone and tonogenesis in Navajo. *UCLA Working in Phonetics* 84, 165-182.

de Reuse, Willem. 1993. *Stylistic and Dialectal Variation in Western Apache Phonology.* Manuscript, Department of Anthropology, University of Arizona.

de Reuse, Willem. 1994. *An Introduction to Western Apache Linguistics* (3 vols.). Manuscript, Department of Anthropology, University of Arizona.

Durbin, Marshall. 1964. *A componential analysis of the San Carlos dialect of Western Apache: A study based on the analysis of the phonology, morphophonics, and morphemics.* Ph.D. Dissertation, State University of New York, Buffalo.

Golla, Victor. 1970. *Hupa Grammar.* Ph.D. Dissertation, University of California, Berkeley.

Gordon, Matthew. 1996. The phonetic structures of Hupa. *UCLA Working Papers in Phonetics* 93, 164-187.

Greenfeld, Philip. 1972. *The Phonological Hierarchy of the White Mountain Dialect of Western Apache.* Ph.D. Dissertation, University of Arizona, Tucson.

Hardy, Frank. 1979. *Navajo Aspectual Verb Stem Variation.* Ph.D. Dissertation, University of New Mexico, Albuquerque.

Hill, Faith. 1963. Some Comparisons between the San Carlos and White Mountain Dialects of Western Apache. In *Studies in the Athapaskan Languages*, edited by Harry Hoijer et al., 149-154. University of California Publications in Linguistics 29. University of California Press, Berkeley.

Hoijer, Harry. 1938. The Southern Athapaskan Languages. *American Anthropologist* 40. 75-87.

Hoijer, Harry. 1945. *Navaho Phonology.* University of New Mexico Publications in Anthropology. University of New Mexico Press, Albuquerque.

Hoijer, Harry. 1946. Chiricahua Apache. In *Linguistic Structures of Native America*, edited by Cornelius Osgood, 55-84. Viking Fund Publications in Anthropology, New York.

Hoijer, Harry. 1963. The Athapaskan Languages. In *Studies in the Athapaskan Languages*, edited by Harry Hoijer et al., 1-29. University of California Publications in Linguistics 29. University of California Press, Berkeley.

Hoijer, Harry. 1971. The Position of the Apachean Languages in the Athapaskan Stock. In *Apachean Cultures, History and Ethnology*, edited by Keith Basso and Morris Opler. Anthropological Papers of the University of Arizona, 2. University of Arizona Press, Tucson.

Hubbard, Kathleen. 1994. *Duration in Moraic Theory.* Ph.D. Dissertation, University of California, Berkeley.

Jassem, Wiktor. 1968. Acoustic description of voiceless fricatives in terms of spectral parameters.

In *Speech Analysis and Synthesis*, edited by W. Jassem, 189-206. Warsaw: Panstwowe Wydawnictwo Naukowe

Ladefoged, P., J. Ladefoged, and D. Everett. 1997. Phonetic structures of Banawá, an endangered language. *Phonetica* 54, 94-111.

Lehiste, Ilse. 1970. *Suprasegmentals.* MIT Press, Cambridge.

MacEachern, M., Barbara Kern, and Peter Ladefoged. 1997. Wari' phonetis structures. *Journal of Amazonian Languages.* 1.1, 3-28.

Maddieson, I., N. Bessel, and C. Smith. 1996. A preliminary phonetic report on the sounds of Tlingit. *UCLA Working Papers in Phonetics* 93, 125-148.

Maddieson, Ian. 1997. Phonetic universals. In *Handbook of Phonetic Sciences*, edited by William Hardcastle and John Laver, 619-39. Blackwells, Cambridge.

McDonough, Joyce and Martha Austin-Garrison. 1994. Vowel enhancement and dispersion in the vowel space of Western Navajo: A study of traditional Navajo speakers. *UCLA Working*

Papers in Phonetics 87, 93-104.

McDonough, Joyce and Peter Ladefoged. 1993. Navajo stops. *UCLA Working Papers in Phonetics* 84, 151-164.

McDonough, Joyce, Peter Ladefoged, and Helen George. 1993. Navajo vowels and universal tendencies. *UCLA Working Papers in Phonetics* 84, 143-150

Opler, Morris. 1983. The Apachean Culture Pattern and Its Origins. *In Handbook of North American Indians*, edited by William Sturtevant. Smithsonian Institution, DC.

Potter, Brian. 1997. *Wh / Indefinites and the Structure of the Clause in Western Apache*.Ph.D. Dissertation, University of California, Los Angeles.

White Mountain Apache Culture Center. 1972a. *Keys to reading and writing Apache*. White Mountain Apache Culture Center, Fort Apache.

White Mountain Apache Culture Center. 1972b. *Western Apache Dictionary*. White Mountain Apache Culture Center, Fort Apache.

Young, Robert & William Morgan, Sr. 1987. *The Navajo Language. A Grammar and Colloquial Dictionary*. University of New Mexico Press, Albuquerque.

Acknowledgements

We had a great deal of help from many Apache people in the course of collecting data for this paper. We are grateful to all the members of the San Carlos and White Mountain Apache tribes who were so good to us. May they and their language continue to flourish. We are also grateful to Siri Tuttle for helpful comments. Part of this work was completed while Brian Potter held a postdoctoral fellowship at the University of Calgary.

Unaspirated coronal stops in Jicarilla Apache

Siri G. Tuttle

stuttle@ucla.edu

1. Introduction

This paper looks at the acoustic phonetics of the consonants in Jicarilla Apache that are spelled in the practical orthography with **d**. In this Athabaskan language of the Eastern Apachean group, spoken in Northern New Mexico, sounds spelled **d** come from two sources: proto-Athabaskan *[n] and *[t]. This paper addresses the question: is there any phonetic difference between an alveolar stop that used to be a nasal, [n], and an alveolar stop that was always a stop?

In comparative Athabaskan analyses, the laryngeal distinctions for obstruents are held to include unaspirated vs. aspirated vs. glottalized; there is not normally a voiced member in the set. Thus, plain [t] and glottalized [t'] are the coronals expected in Jicarilla from the old stop series. The phone descended from aspirated [*tʰ] is no longer an alveolar, but a velar stop in the modern language (except in a few words in which [tʰ]varies with [tʃʰ] and the aspiration appears to act as an intensifier, as in Navajo cognates). In the velar place of articulation, plain [k], glottalized [k'] and aspirated [kʰ] are expected. There are no labial stops reconstructed for proto-Athabaskan. However, both labial and alveolar proto-sonorants may also occur as stops in Jicarilla.

Tables (1) and (2) show the proto-Athabaskan inventory of stops and sonorants at the labial, alveolar and velar places of articulation, and the Jicarilla stops and sonorants at the corresponding places. Throughout this paper, phonetic symbols are enclosed in square brackets: spelling symbols are presented in boldface. The two representations are paired in the charts for reference. The spelling symbols in the proto-Athabaskan chart follow Krauss and Leer (1981: 190).

Table 1. Labial, alveolar and velar stops and sonorants in proto-Athabaskan

	Labial		Alveolar		Velar	
Oral stops: unaspirated			[t]	**d**	[k]	**G**
Oral stops: aspirated			[tʰ]	**t**	[kʰ]	**q**
Oral stops: glottalized			[t']	**t'**	[k']	**q'**
Nasal sonorants			[n]	**n**		
Other sonorants	[w]	**w**				

Table 2. Labial, alveolar and velar stops and sonorants in Jicarilla Apache

	Labial		Alveolar		Velar	
Oral stops: unaspirated	[p]	**b**	[t] [ⁿd], [d]	**d** **d**	[k]	**g**
Oral stops: aspirated					[kʰ]	**k**
Oral stops: glottalized			[t']	**t'**	[k']	**k'**
Nasal sonorants	[m]	**m**	[n]	**n**		
Other sonorants	[w]	**w**				

The claim of this paper is that both [^{n}d] and [d], which show up in Table 2 as a second line in the "unaspirated oral stops" category, occur in modern Jicarilla as reflexes of *n in stem-initial position. Notice that not all the proto-Athabaskan *[n]s end up as Jicarilla [t],(**d**). This change occurs only in stem-initial position, and more particularly, in head-stem position.

Hoijer (1945) distinguishes between the [t] < [*t] and the [t]<*[n] in Jicarilla by writing [t]<*[t] as **d** and [t]<*[n] as **^{n}d** : thus he spells Jicarilla **-^{n}dá·** 'eye', in contrast with Navajo **-ná·'** and Kiowa-Apache **-dá·'̃**(see Table 3).

Table 3. Hoijer vs. the Spelling System

	Hoijer 1938, 1945	Pedagogical materials
'my eye'	shi^{n}dá·	shidáá
'my lips'	shida·	shidaa

It is not clear from the presentation in Hoijer (1945) whether he means for the Jicarilla **^{n}d** (shared by San Carlos, Chiricahua, Mescalero and Lipan in this analysis) to be taken as a phonetic descriptor. If it were to be so interpreted, a voiceless, unaspirated, prenasalized stop would be expected invariantly in these words. What seems more likely, however, is that Hoijer is notating **^{n}d**, [t]<*[n], as a phoneme which is phonetically similar to his **d** but non-identical, the phonetic realization of which may vary, but which may include a prenasalized, unaspirated alveolar stop. In short, Hoijer's **^{n}d** is a notation for a variable phone which he considered phonologically distinct from the phone he spelled **d**. Hoijer's representation, therefore, suggests that there was a phonetic distinction of some sort between the stops descended from *[n] and those descended from *[t] in the Jicarilla he heard in the late 1930s, but he does not make clear what this phonetic distinction is.

Pedagogical materials prepared for language and literacy education at Dulce, New Mexico (Phone et al.1981; Vicenti et al. 1981) use only **d** in their writing systems to represent both stops. These representations suggest that some further degree of merger may have occurred between the two phonemes Hoijer recognized; or at least, that the phonetic representation of **d** has converged to the point that educators do not find it necessary to write the historically different obstruents with different symbols. This representation is not inconsistent with Hoijer's notation, since there seem to be no words which must be distinguished from one another solely by a difference between his **^{n}d** and **d**; it simply collapses what may be a redundant distinction in the alphabet. It does suggest, however, that any phonetic difference between orthographic **d** and **d** is not worth listening for, and may not be even perceptible.

However, teachers of Jicarilla (Torivio, Sandoval, p.c.) report that they recognize a difference between two sets of words which are spelled with **d**. Torivio and Sandoval identify the difference as one of voicing. This fact suggests that the spelling system may have been developed under the influence of the Navajo writing system, which properly represents for Navajo only one unaspirated alveolar stop, with only one historical source.

So the phonetic status of unaspirated alveolar stops, both written with the symbol **d,** is not completely clear. Whatever differences exist may be quite subtle, if they still exist at all. However, it would be desirable to clarify the status of the two orthographic **d**s: if they were phonetically distinct half a century ago, are they still distinct? If they are not distinct, what is the nature of their acoustic expression?

2. Why is this question interesting?

In the Athabaskan language family, voicing in stops has always been considered nondistinctive: the salient distinctions are between unaspirated, aspirated and glottalized stops. But not all Athabaskan stops have the same history. The stops and affricates at the alveolar place of articulation in Apachean are historically stops and affricates. However, the (usually) single stop at the labial place of articulation is historically a sonorant, *w as reconstructed by Krauss and Leer (1981).

Throughout the Athabaskan language family, reflexes of *w and *n may vary according to their morphological position. In stem-initial position, recognized by Hoijer (1938, 76) and Krauss and Leer (1981:7) as the position which tends to cause sonorants to strengthen the most, *w frequently developed into unaspirated [p] (spelled **b** in the conventional orthography) unless it is followed by a nasal segment, in which case it may develop into [m]. In some languages (Krauss and Leer (1981:6) list Tanacross, Northern Tutchone (Selkirk dialect), Southern Tutchone, Tagish, Slavey, and Dogrib) the stem-initial strengthening process did not proceed so far, and a nasalized stop, written variously as **mb,** ^{m}**b** or **m**^{b} , is found; in others (Holikachuk, Lower Koyukon, Upper Kuskokwim, Northern Tutchone (Mayo dialect), Sarcee, Hupa, Tolowa/Tututni/Chasta Costa, Umpqua) the historic approximant only becomes a sonorant consonant, resulting in [m], **m**. Within the same set of languages, parallel development occurred in the alveolars, so **d** can represent a reflex of either *d or *n; but **n, nd,** ^{n}**d** or **n**^{d} may only represent a reflex of *[n]. (Note, however, that a sequence spelled **nd** may also represent a series of morphemes. The examples dealt with in this paper only involve single phones.)

While the process of "obstruentization" of stem-initial sonorants is parallel for labials and alveolars, the resulting consonant inventories are not. This is because sonorants and obstruents in the Athabaskan languages have different patterns of featural contrast. Sonorants may contrast in place of articulation and in nasality, but while they can occur in voiceless versions in some of the languages (Tuttle1998), their system of contrasts really has nothing to do with distinctions in voice onset time, which are the acoustic correlates of distinctions in voicing. Sonorants are generally expected to turn out voiced; voiceless sonorants are rare.

In Jicarilla obstruents a basic distinction in voice onset time exists between voiceless unaspirated and aspirated consonants. Like the sonorants, however, the obstruents are not usually distinguished by voicing per se, even though a phonetic voiced variant may be heard in certain contexts. Obstruents, unlike sonorants, will be expected to turn out voiceless except in very particular situations.

In other Athabaskan languages, stops descended from sonorants participate differently in the phonology from stops descended from stops. For example: in Slave, a northern Athabaskan language (Rice 1993), [t]<*[n] behaves differently from [t]<*[t], and [p]<*[w] behaves analogously to the alveolar ex-sonorant. Specifically, nasal forms of the stops occur in non-head positions in compounds: Rice cites the examples reproduced in (1):

(1) Slave examples of oral-nasal alternation (Rice 1993:323)

a) Alternating **d**

sɛ-dá 'my eye' (rough phonetics: [sɛ.tá]?)
1sPoss-'eye'

na-tú 'tears' (rough phonetics: [na.t^{h}ú])
'eye' + 'water'

b) Non-alternating **d**
sɛ-da 'my chin' (rough phonetics: [sɛ.ta])
1sPoss-'eye'

da-ɣá 'beard' (rough phonetics: [ta.ɣá])
'chin' + 'hair'

This alternation has proved difficult to describe without referring to the historical identity of phones when a word like Slave -dá, 'eye' is written with the conventional Athabaskanist **d**, and -da, 'chin' is written with the same letter. Rice (1993) proposes that the Slave reflex of *[n] is marked by a particular kind of voicing called "sonorant voice" (it is, thus, underlyingly voiced in contrast to [t]<*[t]). In effect she adds this phone to the inventory, as Hoijer did with his Jicarilla ^{n}**d.**

Jicarilla shows the same pattern in compounds that Slave does, as shown in (2):

(2) Jicarilla examples of oral-nasal alternation

a) Alternating **d**
shi-dáá 'my eye' ([ʃɪ.dáː])
1sPoss-'eye'

na-kóh 'tears' ([na-k^{h}óh])
'eye' + 'water'

b) Non-alternating **d**
shi-daa 'my lips' ([ʃɪ.taː])
1sPoss-'lips'

-da-ghaa 'beard' ([ta.ɣaː])
'lips' + 'hair'

This paper will show that modern Jicarilla has plenty of evidence for a distinction between stops derived from sonorants and other stops, and that for the data set analyzed here, voicing is precisely the phonetic marker most saliently dividing the two groups. This means that it is not necessary to refer to the historical status of these sounds in discussing modern phonology: they always were distinct, and have not lost their distinctiveness, though it has shifted from a sonorant-obstruent distinction to a distinction signaled by voice onset time.

3. Methods

1. Four speakers: Jonathon Wells (70), Jackson Velarde (60), Patricia Torivio (60) and Merton Sandoval (45) wished to be identified as the speakers of this endangered language who were recorded for this investigation of alveolar stops.
2. A word list was created. Table 4 below shows a transcription of the prevalent version or versions of the word as heard during the recording session, gloss, historical form of the consonant of interest, and morphological/phonological position for the consonant of interest. The third person singular possessive is used in many cases to provide an intervocalic environment for the consonant of interest; this morpheme varies in realization from [pɪ] to [^{m}bɪ] [bɪ] to [mɪ], and is arbitrarily transcribed with the nasal version in the appendix to identify the morpheme.
3. The four * [n]-words are contrasted here with six *[t] examples, seven examples of *[t^h], six examples of *[k^h], six examples of stem-initial *[w], and three examples of intensifier-

aspiration. Because of the particular environment in which the obstruentized [n] has developed, it is not available in word-initial position, so all examples are intervocalic. "SIM" in the appendix stands for "stem-initial, medial".

4. Consonants were analyzed for VOT, closure and presence of nasality during closure.

Table 4. Wordlist for Jicarilla coronal stops (SIM = "Stem-initial, phonologically medial")

Case	Spelling	Transcription	Gloss	Original Consonant	Position
1	midee	mɪtɛ:	its horn	[t]	SIM
2	midiɬ	mɪtɪɬ	his, her, its blood	[t]	SIM
3	miidáan	mɪtâ:n	his, her food	[t]	SIM
4	sidá	sɪtáh	he, she sits	[t]	SIM
5	midibé	mɪtipɛʔɛ́:	his, her sheep	[t]	SIM
6	midaa	mɪta:	his, her lips/beak	[t]	SIM
7	diką́	tɪk^{h}ã́h	it's thick	[t^h]	SIM
8	miká	mɪk^{h}á:ʔ	his, her forehead	[t^h]	SIM
9	mika'éé	mɪk^{h}áʔɛ́:	his, her father	[t^h]	SIM
10	mikóh'éé	mɪk^{h}óʔɛ́:	his, her,its water	[t^h]	SIM
11	nkeeɬ	nk^{h}ɛ:ɬ	it's wide	[t^h]	SIM
12	sikį́	sɪk^{h}ĩ́	person is lying down	[t^h]	SIM
13	siką́	sɪk^{h}ã́	sticklike object is there	[t^h]	SIM
14	midádéé	mɪdádɛ́:	his, her older sister	[n]	SIM, SFM
15	hidaa	hɪda: hɪ^{n}ta:	alive	[n]	SIM
16	midáá	mɪdá: mɪ^{n}tá:	his, her eyes	[n]	SIM
17	mii'idaa	mɪʔɪda: mida: mɪʔɪ^{n}ta:	his, her enemy	[n]	SIM
18	didé	tɪdɛ́h tɪ^{n}tɛ́h	man, person	[n]	SIM
19	miké	mɪk^{h}ɛ́	his, her shoes	[k^h]	SIM
20	mikįh	mɪk^{h}ĩh	his, her house	[k^h]	SIM
21	mikǫ	mɪk^{h}õh	his, her fire	[k^h]	SIM
22	dasiikaa	tasik^{h}a:	they are spread out, e.g. bushes	[k^h]	SIM
23	siké	sɪk^{h}ɛ́	they sit, two people	[k^h]	SIM
24	miińké'	mi:ńké'	its trail, tracks	[k^h]	SIM

25	miibaas	mipaːsi mipaːzi	his, her wagon	[w]	SIM
26	dasibaał	tasipaːɬ	awning or curtains hang	[w]	SIM
27	miibezhii	mipɛ́ʒi	his, her knife	[w]	SIM
28	naabaas	naːpãːs	it is rolling around (a car)	[w]	SIM
29	sibésh	ʃɪpɛ́ʃ	it is boiled	[w]	SIM
30	naabéh	naːbɛ́h	he or she is bathing	[w]	SIM
31	ntoh	nt^{h}õh ntʃ^{h}õh	he or she is pouting	[t] + INT	SIM
32	ntoo'é	nt^{h}õːʔɛː ntʃ^{h}õːʔɛ́ː	it's no good, a bad one	[t] + INT	SIM
33	nłtó	nɬt^{h}ṍ nɬtʃ^{h}ṍ	it stinks	[t] + INT	SIM

As can be seen from this list, the elusive phone descended from stem-initial *n is hard to capture even in a wordlist: certain verbs use [n] even in stem-initial position, and so had to be discarded. The four solid [n]-words are contrasted here with six *[t] examples, seven examples of *[t^{h}], six examples of *[k^{h}], six examples of stem-initial *[w], and three examples of intensifier-aspiration. Because of the particular environment in which the obstruentized [n] has developed, it is not available in word-initial position, so all examples are intervocalic.

4. Results

The results of this study show two things. Firstly, it is very clear from acoustic measurement, as from observation of consonant shapes, that the reflex of *[n] in Jicarilla Apache is clearly distinguishable and distinguished from the reflex of *[t]. Secondly, there appear to be differences in realization of the *[n] phone both within and between speakers.

4.1 Qualitative comparison: No merger here

Figure 1 shows a spectrogram of a word containing the prevalent realization of *[n], which is a voiced [d] (as reported, it should be remembered, by the language teachers above) , and Figure 2 shows a comparable example of a word containing [t].

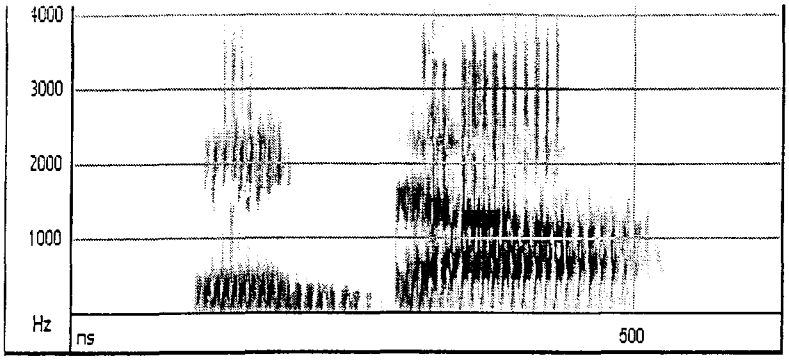


Figure 1. [bɪdáː] 'His or her eyes' Speaker JW

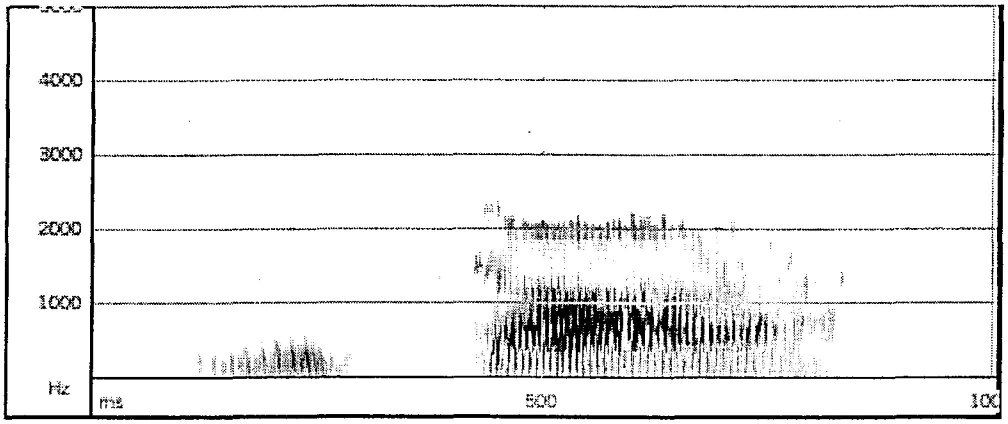


Figure 2. [bɪtaː], 'his or her lips'

Notice in Figure 2 that the closure of the [t] is completely voiceless, as opposed to the voiced closure of the [d] seen in Figure 1. Following the burst of the [t], there is also a very tiny amount of aspiration – less than 10 milliseconds.

Figure 3 provides a comparison with a truly aspirated consonant, [k^h], in the same phonological environment:

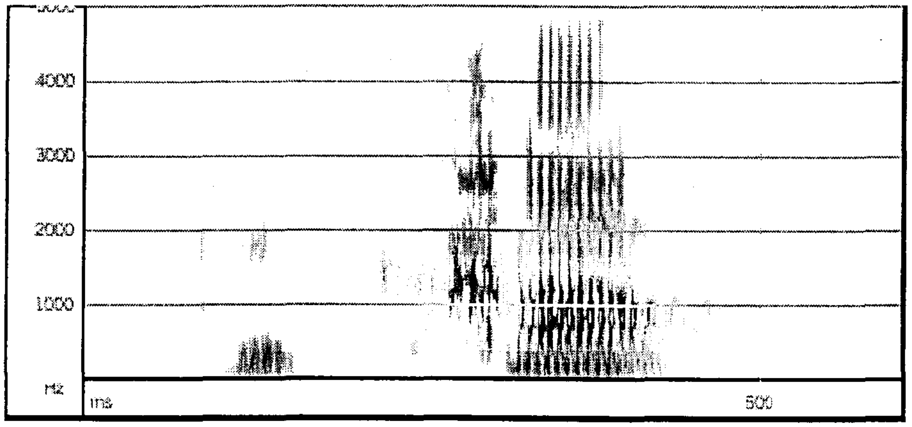


Figure 3. [tɪk^{h}aː], 'it's thick'

Comparison of Figure 2 with Figure 3 shows that the [t] phoneme cannot be considered aspirated by Apachean standards (see McDonough and Ladefoged 1993 for comparison with Navajo stops, and Potter et al (2000) for comparison with Western Apache). However, [t] is different again in voice onset time from stem-initial [p], as shown in Figure 4:

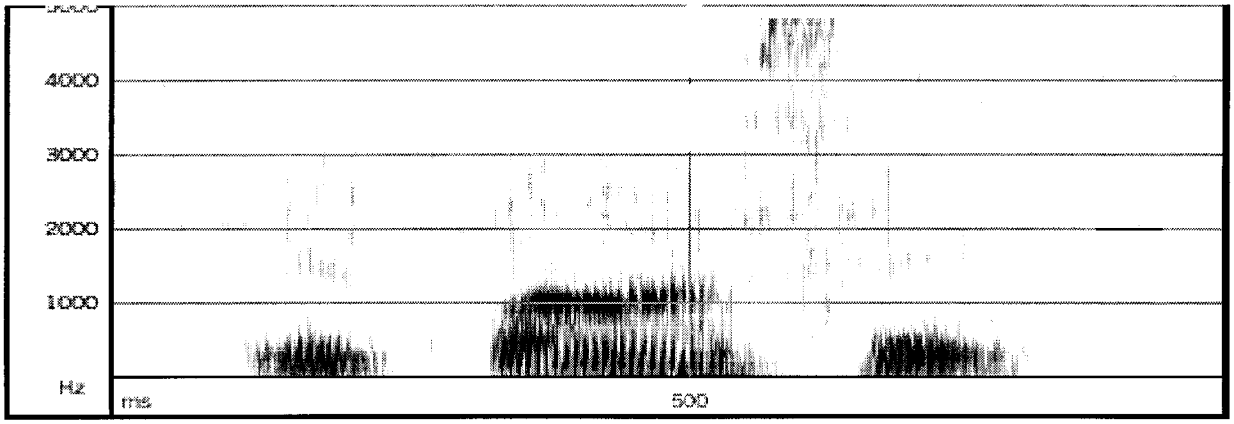


Figure 4. [bɪpaːsi], ’his or her wagon’

The labial in stem-initial position is a completely unaspirated voiceless stop, the realization of which is distinct from the voiced and varying labial in the prefixes seen in these four examples.

Thus, [d] and [t] are quite distinct from one another in the prevalent pattern in this data set, with [d] being voiced and [t] voiceless and slightly aspirated. Their patterns of voice onset time are also distinct from those of phones at other places of articulation; stem-initial [t] is more aspirated than stem-initial [p], but nowhere close to the truly aspirated $[k^h]$.

4.2 Quantitative analysis: No merger here either

An analysis of variance, taking all tokens from all speakers together, shows a significant difference in voice onset time between the phone descended from *[n] and that descended from *[t], with the first showing voiced closure overall, and the second showing no voicing, but a tiny amount of aspiration instead.

Figure 5 shows the quantitative effect of historical identity:

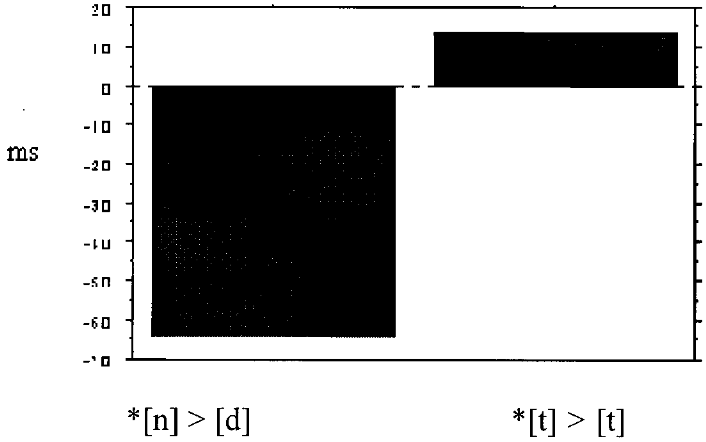


Figure 5. Effect of historical identity, *[n] (now [d]) vs. *[t] (always [t])

Based on this analysis, it would seem to make most sense to designate the Jicarilla reflex of stem-initial *[n] as voiced [d], in contrast to the unaspirated to slightly aspirated [t]. The difference is significant at $p < .0001$.

4.3 Intra- and interspeaker variation

The overall significance of the VOT distinction cannot be allowed to obscure another very interesting finding. Both within the production of individual speakers, and across the speakers, there is variation in the production of the segment we are now calling /d/.

The most prevalent variant, as discussed in section 4.2, is the voiced stop. However, there are also two different prenasalized stop variants, one voiceless and one voiced. Figures 6 and 7 show a comparison of waveforms of two performances of the same word by the same speaker: JV gives one token of the word for 'his or her eyes' with a voiceless prenasalized stop, and one with a voiced prenasalized stop.

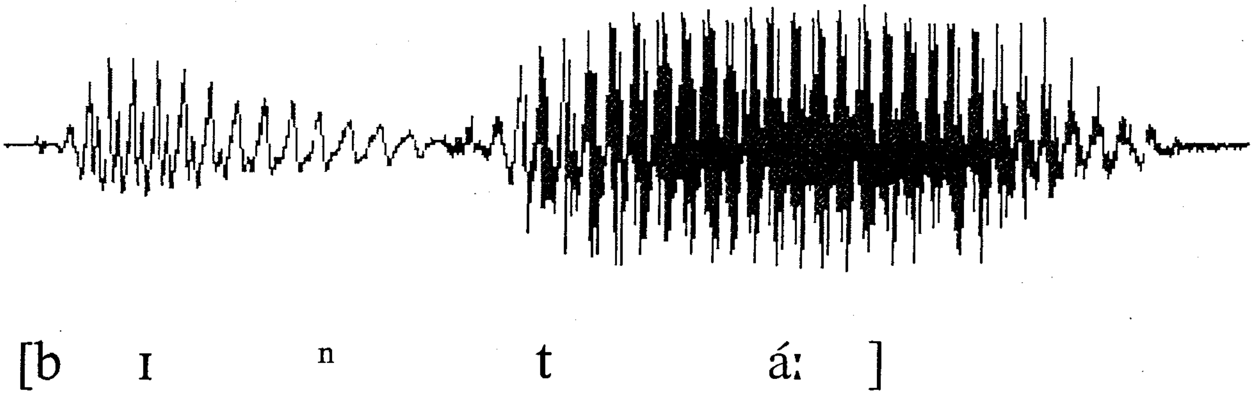


Figure 6. ’His or her eyes’ (Token 1, JV)

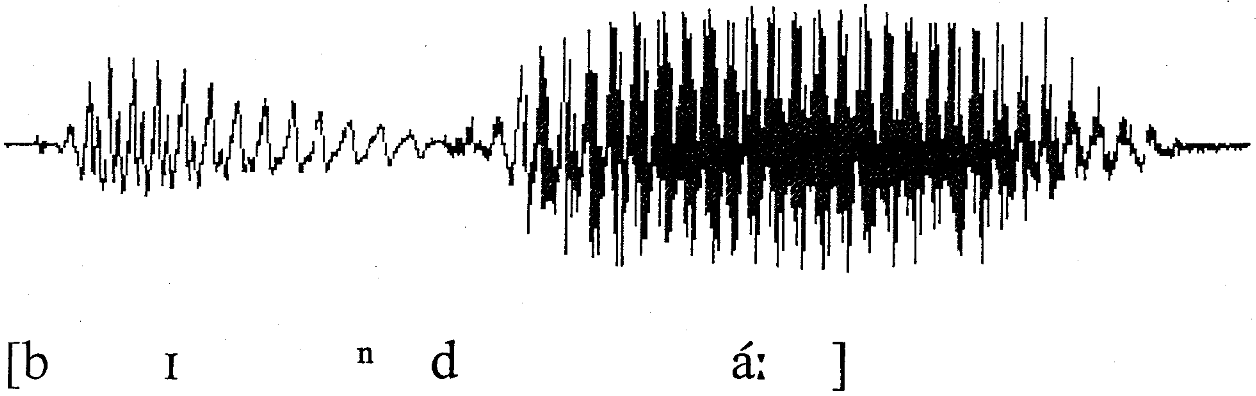


Figure 7. [’His or her eyes’ (Token 2, JV)

The difference between these two performances is that the first has a very short period of voiceless closure, while in the second, the nasal stretch following the first vowel slides right into the stem-initial stop and serves as its closure.

Compare these tokens with a voiced stop in MS’s production, in which the same word shows up with a voiced consonant with no trace of nasality, as shown in Figure 8.

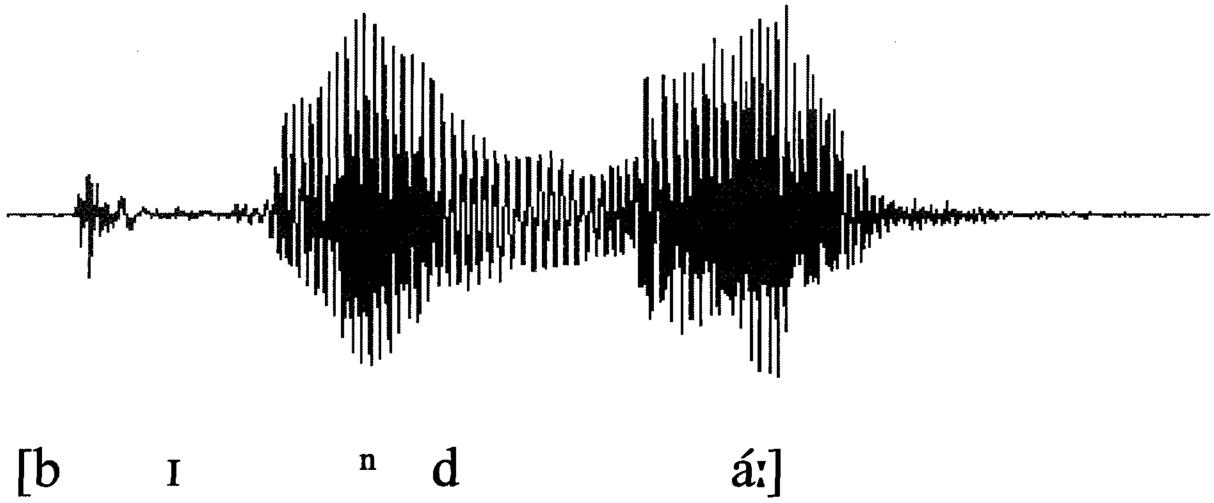


Figure 8. The same word as in figure 6 and 7, produced by speaker PT)

This difference is hard to see in the waveform, and in fact it is not easy to hear, either. The nasal portions of JV's prenasal stops are very quiet, and the voicing in PT's /d/ is very loud for voicing (perhaps this loud voicing is a step between a prenasalized and a more ordinary voiced stop). However, the difference between any of these productions and the voiceless, non-nasal [t] is easy both to see and hear. For comparison, look at MS's plain stop in 'lips' (Figure 9):

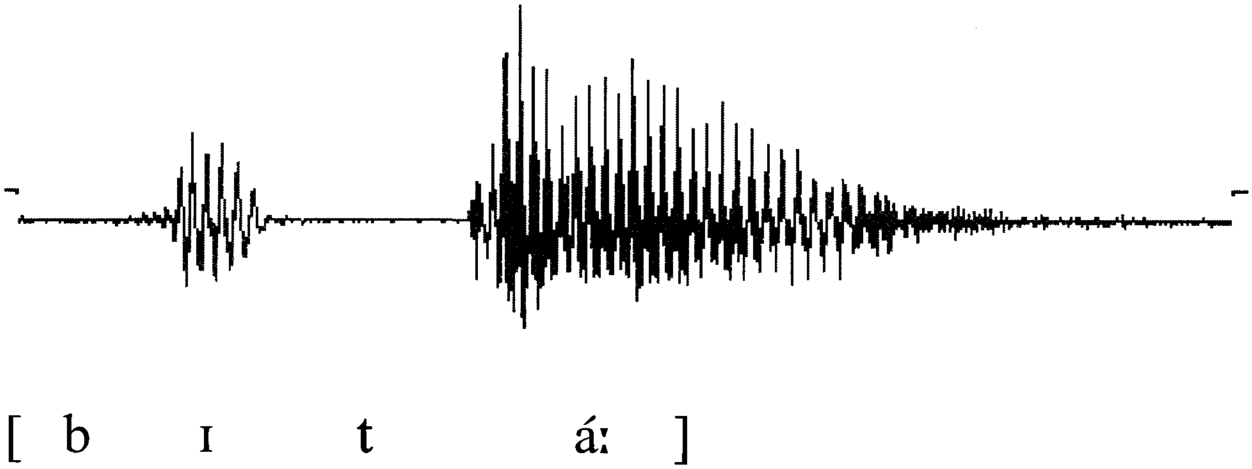


Figure 9. 'his, her lips' (MS)

Overall, about 25 percent of the performances of *n-words have a prenasalized stop in the output. This is shown in Figure 10.

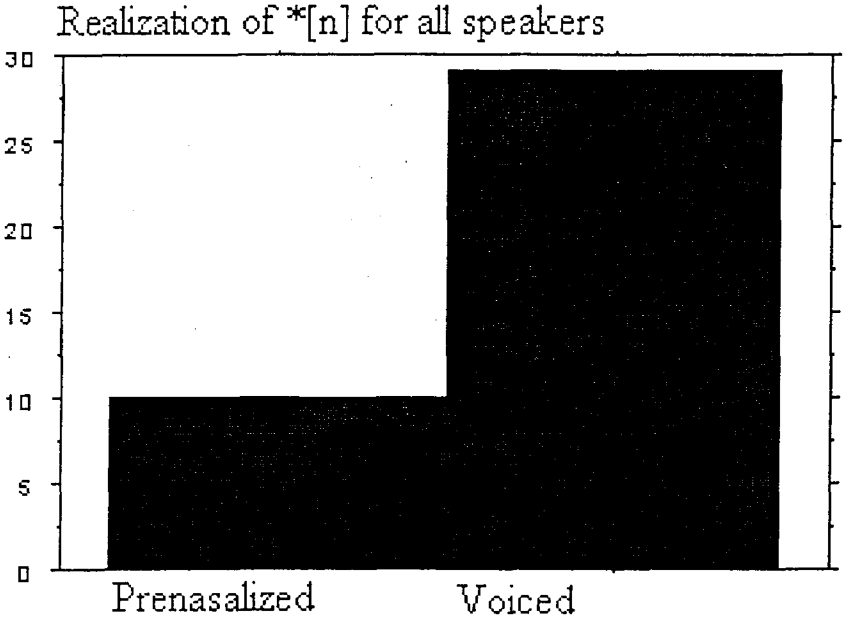


Figure 10. Realization of *[n] reflex for all speakers

The distribution of prenasalized vs. voiced tokens for *[n] is different for each speaker. Figure 11 shows the different percentages of prenasalized stops for each speaker: JV having the largest percentage, and PT, with 0 percent, having the smallest percentage of prenasalized stops.

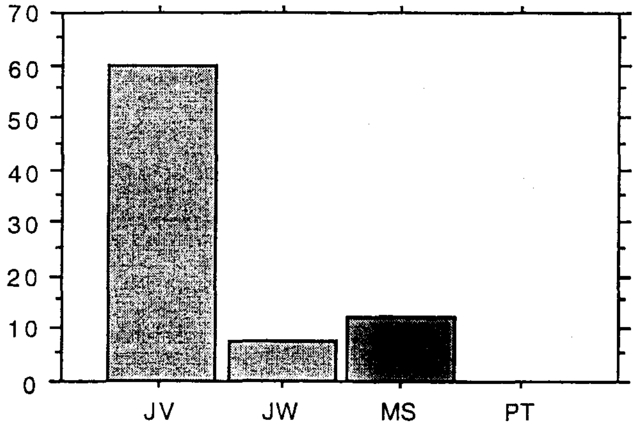


Figure 11. Percentage of prenasalized tokens of *[n], by speaker

This variety in production does not seem to correlate either with age or with dialect in the present data set, as shown in Table 5:

Table 5: Speaker information and percentage of prenasalized stops

Speaker	Age	Dialect reported	Percentage of prenasalized stops
JW	70	Lajara / Cordoba Canyon	10%
JV	60	Alawona / Horse Lake	90%
PT	60	Burns Canyon	0%
MS	45	Alawona / Horse Lake	10%

The youngest and oldest speakers have the same percentage of prenasalized stops in this data set; the youngest speaker shares a dialect with JV, who does the most prenasalizing; and JV shares an age with PT, who does no prenasalizing at all. So it is not clear from this data set whether or not there is a pattern to the variation in speakers' production of the reflex of *[n]. While further research may discover such a pattern, it may also be the case that these realizations are in free variation.

4.4 Summary

To summarize: the production data gathered in this data set show clearly that *[n] and *[t] are different in modern Jicarilla Apache, confirming the impressions of Hoijer (1938) and of the Jicarilla language teachers. Predominantly, voicing makes the difference between the two phones, with an analysis of variance significant at $p<.0001$.

However, both speakers and performances differ in their realization of the *[n] phone, with a prenasalized variant showing up about one quarter of the time in this data set. One speaker is responsible for most of these tokens. However, it is not clear that age or dialect can be shown to be responsible for this speaker-to-speaker variation.

5. Coda 1: Perception favors VOT

As part of this investigation of *[n] and *[t], a small perception experiment was attempted to pair with the production data. Because of field conditions and speaker preferences, this experiment was not carried out in a very controlled fashion, and for this reason, the results will be reported only informally. However, it may be of interest to know how this experiment was structured.

One of the difficulties in creating data sets for Athabaskan languages, generally, is the nearly complete absence of minimal pairs. The acute reader will have observed above that while there is a pair [bɪtaː] 'lips' vs. [bɪdáː] 'eyes', it is not a minimal pair because the two words differ in tone as well as in their stem-initial consonant. Other near-pairs turn up as well, but end up not being usable because of vowel length or other minor differences.

For this reason, a set of manipulated recordings was created for this investigation based on a single naturally recorded word, a production of [bɪtaː] 'lips' . In two tokens, this recording was presented as it was: in two, the closure was replaced by voiced closure; in two, the closure was doubled in length; in two, 40ms of aspirated noise was inserted; and in two, the consonant was replaced with an aspirated [k^h]. Speakers listened to these tokens on a laptop computer and gave judgments as to whether they thought the word sounded more like 'lips' (with a [t]) or 'eyes' (with a d or ^{n}t/d).

As noted above, the conditions were not right on this first attempt for a controlled measurement of judgments, and in fact most of the responses were all over the map. However, for all speakers and for all tokens, the tokens with *voiced* closure were identified as having the right consonant for 'eyes' (a correct response). This suggests that the voicing difference which the teachers first reported, and which shows up so clearly in overall statistics, is salient to speakers in this distinction.

It is also worth noting that these speakers were quite willing to attempt the listening task, and refinements of such materials would probably product more interesting results.

6. Coda 2: Why are these two sounds spelled the same?

Given the findings reported above, it might occur to the reader to wonder why the designers of the Jicarilla spelling system, or anyone else, would confuse the two phones we are writing here as [d] and [t].

It is not actually that hard to do. In the initial stages of research for this paper, a set of *[n]-*[t] contrasts were recorded using a head-mounted, noise-canceling microphone, and analyzed using PCQuirer software. In this set, voicing is not visible in spectrograms and is barely observable in waveforms. The more complete data set on which the present findings are based was recorded using a lapel microphone and with the input level set higher, and in this set, voicing is clearly visible in the same words. It is not obvious whether the difference is due to recording conditions or to speaker variation; either way, it is clear that it is possible to hear and record a [d] which is extremely weakly voiced. Figures 15 (the preliminary data) and 16 (the best data so far) contrast waveforms of the same word, spoken by the same speaker (PT) in the two different sessions.

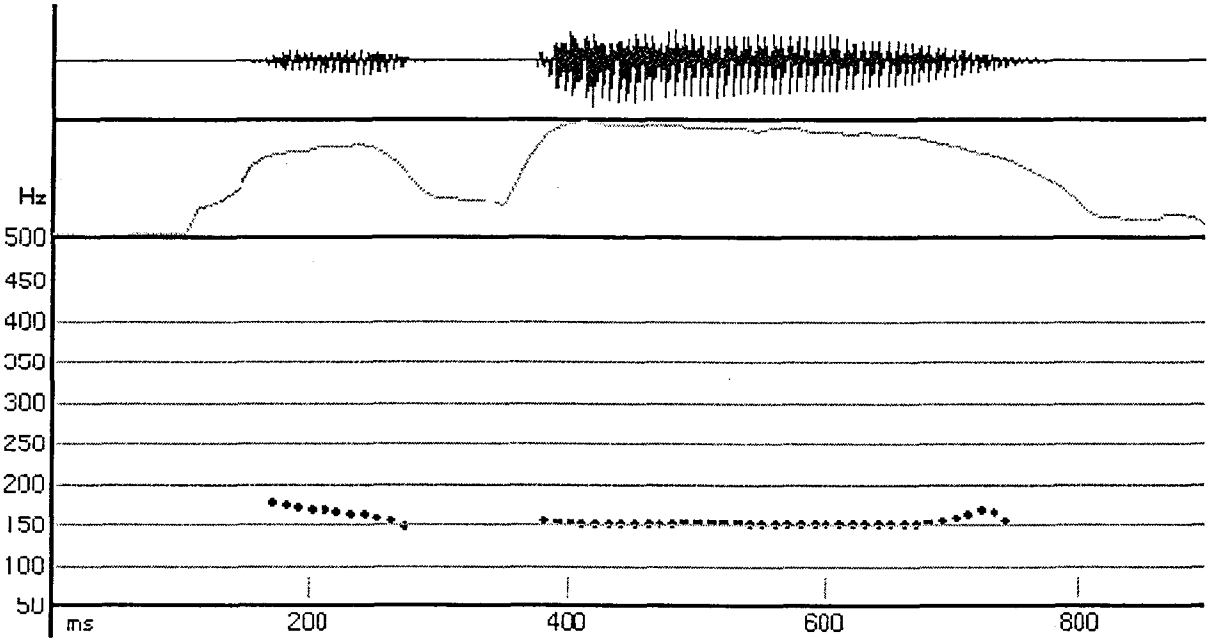


Figure 12. PT's first–session recording of [ʔida:] 'enemy' (a [d]-word)

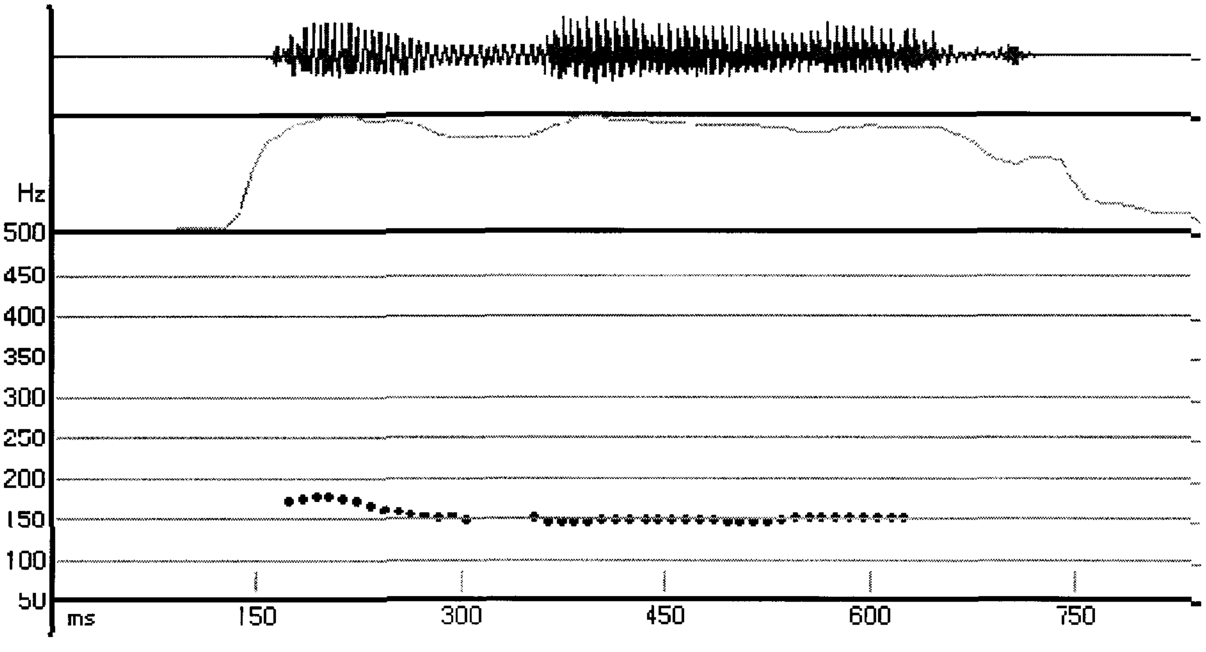


Figure 13. PT's second-session recording of [ʔidaː] 'enemy'

The data from the first session showed significant differences between the [t]-words and the [d]-words, but these differences were limited to the tiny distinction in aspiration (significant at $p<.01$) and a significant difference in closure duration ($p< .0001$). Because most of the tokens of [d] were not clearly voiced compared to tokens of [t], this session's output looks much more like a near-merger than the output of the second session. Given the intraspeaker and interspeaker variation observed above, and this example of within speaker variation under very similar recording conditions, it seems quite likely that non-Jicarilla listeners not using high-powered equipment could conclude that a merger of [t] and [d] is in the process of taking place.

However, because of the greater completeness (more words) and improved recording conditions (more amplitude) of the second session, we have chosen to report the salient voicing distinction in the present paper.

7. Conclusion

To conclude: the results of this investigation into the reflexes of *[n] and *[t] are clear: there is still a phonetic distinction between these phones, which reflects a phonological difference recognized by speakers. This is in spite of the fact that there are no true minimal pairs distinguished only by this difference.

The phonetic difference is predominantly one of voicing: *[n] shows up as [d] most of the time. However, there is a prenasalized variant of [d], which may or may not have voiced closure, and which three of the four speakers produced even in this short data set. The distribution of prenasalized [^{n}t] ~ [^{n}d] does not appear to correlate nearly with age of speakers or with dialect, but with a larger sample perhaps some pattern could be discovered.

References

Cho, Taehong and Peter Ladefoged (1999) Variations and universals in VOT: evidence from 18 languages. Journal of Phonetics 27:2, pp 207-229.

Hoijer, Harry (1938) The southern Athapaskan languages. American Anthropologist., n.s., v. 40, pp. 75-87. Menasha, Wisconsin.

Hoijer, Harry (1946) The Apachean verb, part II: the prefixes for mode and tense. International Journal of American Linguistics, 12:1-13.

Krauss, Michael and Jeff Leer (1981) Athabaskan, Eyak, and Tlingit Sonorants. Alaska Native Language Center Research Papers #5. Fairbanks: Alaska Native Language Center.

McDonough, Joyce and Peter Ladefoged (1993) Navajo Stops. UCLA Working Papers in Phonetics, 84.

Phone, Wilma, Torivio, Patricia, et al (1981) *Jicarilla mezaa medaoółkai: áłáéé*. Albuquerque : Native American Materials Development Center.

Potter, Brian, Matthew Gordon, John Dawson, Willem de Reuse and Peter Ladefoged (2000) Phonetic structures of Western Apache. UCLA Working Papers in Phonetics, 98, 1-30.

Rice, Keren D. (1993) A Reexamination of the Feature (Sonorant): The Status of 'Sonorant Obstruents' Language: Journal of the Linguistic Society of America (Language), Washington, DC, ISSUE: vol. 69 no. 2 PAGES: 308-44

Tuttle, Siri G. (1998) *Metrical and tonal structures in Tanana Athabaskan.* Ph.D. dissertation, University of Washington.

Vicenti, Carson (1981) *Jicarilla Apache Dictionary.* Albuquerque: Native American Materials Development Center.

Young, Robert, and William Morgan (1987) *The Navajo Language: A grammar and colloquial dictionary.* Albuquerque: University of New Mexico Press.

An Acoustic and Aerodynamic Study of Consonants in Cheju[*]

Taehong Cho, Sun-Ah Jun & Peter Ladefoged

Acoustic and aerodynamic characteristics of Cheju consonants were examined with the focus on the well-known three-way distinction among stops (i.e., lenis, fortis, aspirated) and the two-way distinction between s and s*. Acoustic parameters examined for the stops included VOT, relative stop burst energy, f0 at the vowel onset, H1-H2, and H1-F2 at the vowel onset. For the fricatives s and s*, acoustic parameters were fricative duration, f0, centroid of the fricative noise, RMS energy of the frication, H1-H2 and H1-F2 at the onset of the following vowel. In investigating aerodynamics, intraoral pressure and oral flow were included for the bilabial stops. Results indicate that, although Cheju and Korean are not mutually intelligible, acoustic and aerodynamic properties of Cheju consonants are very similar in every respect to those of the standard Korean. Among other findings there are three crucial points worth recapitulating. First, stops are systematically differentiated by the voice quality of the following vowel. Second, stops are also differentiated by aerodynamic mechanisms. The aspirated and fortis stops are similar in supralaryngeal articulation, but employ a different relation between intraoral pressure and flow. Finally, our study suggests that the fricative s is better categorized as 'lenis' than as 'aspirated' in terms of its phonetic realization.

1. Introduction

Cheju is spoken on Cheju Island, which is located about 100 km south of the Korean Peninsula, and 160 km west of Japan. The island is 73 km from east to west, 41 km from north to south in the space of an oval, with a total area of 1,845 km^2. Politically, Cheju Island is an integral part of South Korea, one of the eight provinces of the country. It differs from the other provinces in its tropical weather and plants, which attract many tourists. The population of Cheju has been increased rapidly in past decades due to the migration from the mainland Korea. As of 1997, there were approximately 540,000 people in Cheju Island. Due to the high level of education and the influence of mass media, it is not easy to find a pure native speaker of Cheju. Cheju is spoken more commonly in the inland, rural, areas.

Cheju can be considered to be simply a dialect of Korean. It shares the same morphosyntactic structure and the same writing system as other dialects of Korean. It is, however, not mutually intelligible with the rest of Korean, and might be considered to be a separate language. However, in this paper, we will take no stand on this issue. Instead, we will simply compare the consonants of Cheju, whenever necessary, with those of the standard Korean spoken in Seoul, which we will refer to as Korean in this paper.

This paper is based on recordings of eight male speakers of Cheju. Three speakers (S1-S3) were recorded in Cheju city, aged 55, 61 and 68, and the other five speakers (S4-S8) were recorded in the more mountainous area of southern Cheju (Namcheju-kun), in two villages: Shinrye-ri and Uikwi-ri. The five men in this region were between 66 and 74 years old. All the speakers understood Korean, were literate, and above average in their socioeconomic status. We

---

[*] This paper is to appear in *Korean Journal of Speech Science,* vol. 7.

also recorded eight female speakers, three in Cheju city (aged 62, 66 and 75) and five in the rural area (aged between 68 and 78). Data from these speakers will be considered in a later paper.

Each of the speakers was recorded using a close-talking, noise-canceling Shure microphone and a Sony DAT recorder. The recordings were made in the home of one of the speaker's, or in a quiet part of the village community center. The word list consisted of 80 Cheju words, each of which was written on an index card in Korean orthography reflecting the sounds of Cheju as close as possible. To help speakers produce each word more naturally, a word triggering the context was written next to the target word on the index card. For example, a word for 'baby' was written for the target word 'to give a birth'; a word for 'ground' was written for the target word 'to dig'. Speakers produced each word twice after being prompted by one of the authors, S-A. Jun, or by a Cheju-native linguist (see Acknowledgements). The word list was rehearsed with each speaker before the recording was made.

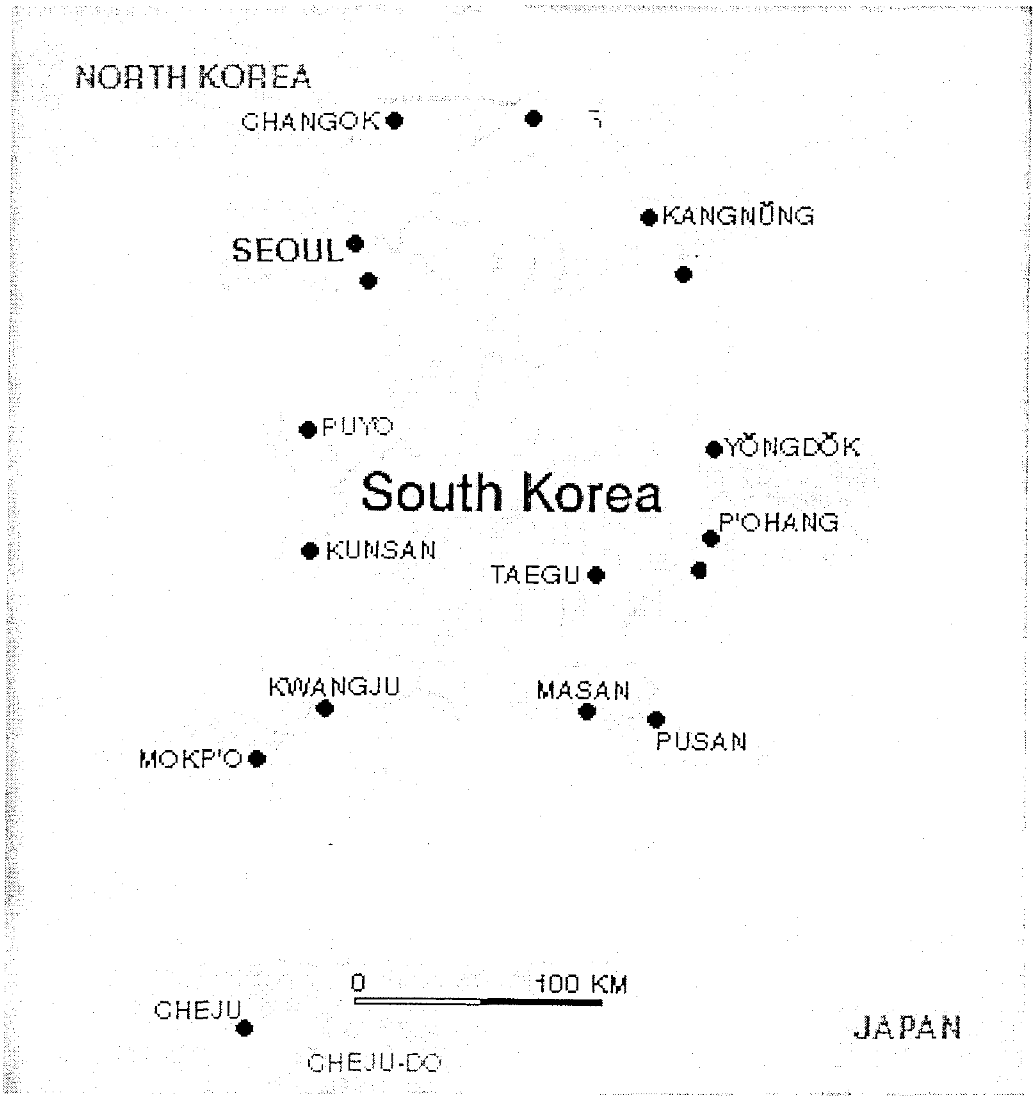


Figure 1. The location of Cheju Island.

2. Cheju obstruents

A great amount of work has examined the acoustic and articulatory properties of Korean consonants, especially the typologically unusual system of Korean stops. Cheju is like Korean with respect to its stop system. Both Cheju and Korean stops occur at three places of articulation, bilabial, denti-alveolar, and velar. At each place of articulation, stops fall into three different categories, often called lenis, fortis and aspirated. The lenis stops have been described as lax and slightly aspirated, the fortis stops as tense and unaspirated, and the aspirated stops as being strongly aspirated. Along with the stops, both Cheju and Korean distinguish post-alveolar affricates with a similar three-way distinction, lenis, fortis and aspirated. In addition, there is a two-way distinction in denti-alveolar fricatives: plain vs. fortis. The plain **s** is sometimes categorized as lenis, and sometimes as aspirated. In this paper, we will refer to this **s** as a 'plain' **s** in order to avoid any unnecessary confusion brought about by the terminology, lenis vs. aspirated. Table 1 shows all the Cheju and Korean obstruents. In this table, s is categorized as lenis for the sake of simplicity. The diacritic '*' is used to mark fortis obstruents in this paper. Table 2 gives minimal triplets (or, in the case of the denti-alveolar fricatives, a minimal pair) for the Korean sounds. We were not able to find equally minimal contrasts for Cheju, but all these oppositions exist, as will be demonstrated below.

Table 1. Obstruents in Cheju and Korean: lenis, fortis, and/or aspirated distinctions

lenis series	p	t	tʃ	k	s
fortis series	p*	t*	tʃ*	k*	s*
aspirated series	p^h	t^h	$tʃ^h$	k^h	

Table 2. Minimal contrasts for Korean obstruents in word-initial position

Lenis		Fortis		Aspirated	
paŋ	'room'	p*aŋ	'bread'	p^haŋ	'bang'
tal	'moon'	t*al	'daughter'	t^hal	'mask'
tʃata	'to sleep'	tʃ*ata	'to squeeze'	$tʃ^h$ata	'to kick'
kæta	'to fold up'	k*æta	'to break'	k^hæta	'to dig up'
sata	'to buy'	s*ata	'to wrap'	—	—

3. Previous investigations of Korean obstruents

In this section, we will review acoustic and articulartory properties of Korean obstruents, focusing on stops and fricatives. Acoustic and articulatory properties of the three way manner distinctions in the affricates are similar to those of stops, except that affricates have a frication component after stop closure.

3.1 *Stops*

In most of the world's languages, the voicing feature for syllable-initial stops can be specified quite well in terms of the three categories, voiced, voiceless unaspirated and aspirated (e.g., Lisker and Abramson 1964; Klatt 1975). However, Korean does not have a voicing distinction among stops phonologically, though it does phonetically in a word-medial position. They are all voiceless. Nevertheless, there are consistent three-way differences in VOT among the three stop categories. The fortis stops are unaspirated, the lenis stops moderately aspirated, and the aspirated stops are strongly aspirated (Abberton 1972; Lisker and Abramson 1964, *inter alia*). However, other acoustic and physiological studies have suggested that VOT alone does not fully account for the observed three-way phonemic distinction of Korean stops (C. Kim 1965, 1970, Han and Weitzman 1970, Hardcastle 1973, Hirose et al. 1974; M. Kim 1994, Y. Kim 1995, Han 1996, Cho 1996).

As VOT is not the only distinguishing feature of the manner distinction of Korean stops, some phonologists and phoneticians have proposed different ways of identifying these stops in initial position. Han and Weiztman (1970) noted that, in addition to the differences observed in VOT among these stops, different acoustic features can be observed in the onset phase of voicing following the stop release. They reported that the onset value of fundamental frequency and the intensity characteristics of the initial phase of voicing also contribute to the manner distinction. In general, f0 after aspirated or fortis stops is relatively higher than after lenis stops. Similar results were found by Hardcastle (1973), Kagaya (1974), M. Kim (1994) and Cho (1996). These studies indicate that f0 contrasts serve as a suplementary cue to distinguish lenis stops from fortis and aspirated stops, but not necessarily to distinguish aspirated and fortis stops. Furthermore, Jun (1993, 1996) found that the f0 difference is not simply due to the phonetic perturbation that has been noticed in other languages (Lehiste and Peterson1961, Lieberman 1963, Ladefoged 1964, Hombert 1978, Hombert, Ohala, William 1979, *inter alia,*), but is phonologically encoded in the intonation system in most dialects of Korean (including Seoul dialect), so that a phrase-initial syllable beginning with a fortis or an aspirated obstruent is realized with a high (H) tone.

Han and Weitzman (1970) also reported that the harmonic components are weaker for lenis stops, intermediate for aspirated stops, and stronger for fortis stops. They further argued that these observations are indicative of a difference in the intensity build-up following the voice onset associated with each stop: relatively more time is needed for glottal intensity to build up following a lenis stop or an aspirated stop than that following a fortis stop.

The duration of the stop closure has been considered as another acoustically distinctive feature associated with Korean stops — in general, the stop closure is shortest for lenis stops, intermediate for aspirated stops, and longest for fortis stops (Silva 1993; M. Kim 1994; Han 1996). But an electropalatographic (EPG) study by Cho & Keating (1999) showed that there is no significant difference between aspirated and fortis stop closure durations, all else being equal. Cho & Keating (1999) also examined linguopalatal contact (the contact between the tongue and the roof of the mouth) for different stops and found that lenis stops have less linguopalatal contact than aspirated or fortis ones. The longer duration and the greater linguopalatal contact associated with both the aspirated and the fortis stops indicate that they are articulatorily 'strong' stops, compared to the lenis stop.

The nature of the aspirated and the fortis stops has been a long term research topic. One of the earlier studies, C. Kim (1965), characterized the stops in terms of two features: 'tension' of the articulation, which served to distinguish aspirated and fortis stops from lenis stops; and 'aspiration', which served to distinguish aspirated stops from fortis stops (and lenis stops). Some of the supportive findings for 'tension' associated with fortis and aspirated stops were a faster rate of vocal fold vibration after release, greater amplitude of pressure, longer duration of increased pressure, faster pressure build-up, greater linguopalatal contact, greater lip muscle activity (for bilabial stops). In line with C. Kim, on the basis of results of VOT, frequency of glottal cycles at vowel onset and air-flow rate, Hardcastle (1973:271) suggested that "a feature 'tensity', defined in terms of isometric muscular tension in the vocal cords and pharynx, can usefully be employed to explain some of these properties".

As techniques developed for examination of the glottal area, researchers started investigating vocal fold configurations that may contribute to the production of the contrasting stops. C. Kim (1970), using cineradiographic evidence, reported that the glottal opening is larger for aspirated stops, intermediate for lenis stops, and narrower for fortis stops, arguing that the degree of aspiration is proportionally correlated with the degree of glottal opening at the time of release of the oral closure. Kagaya (1974), in a fiberscopic study, found that fortis stops have approximated vocal folds well before the articulatory release, while the glottis is quite open for the lenis stop but not as open as for the aspirated stop at the time of the release. Kagaya suggested that both the fortis and aspirated stops are characterized by some intrinsic laryngeal gestures. In his view, fortis stops can be characterized by (1) a completely adducted state of the vocal folds before the articulatory release, (2) stiffening of the vocal folds, (3) abrupt closure of the vibrating vocal folds near the voice onset, (4) increasing subglottal pressure, and (5) lowering of the glottis immediately before the release. Aspirated stops are associated with positive abduction of the vocal folds and heightened subglottal pressure. On the other hand, none of these positive laryngeal gestures are observed for lenis stops. More recently, Jun, Beckman, & Lee (1998), in their fiberscopic study of vowel devoicing, found a similar result in terms of the timing and size of the glottal opening for the three stop types. They also found that the glottal opening area is greater when the stop is in phrase (i.e., Accentual Phrase) initial than in phrase medial position, suggesting that there is domain-initial strengthening associated with the glottal gesture. (See Cho, 1998; Cho & Keating, 1999; Keating, Cho, Fougeron & Hsu, 1999 for further discussion on domain-initial strengthening.)

The glottal state in these stops was further examined by the Electromyography (EMG), which allows us to investigate the activities of the intrinsic laryngeal muscles. Hirose, Lee, and Ushijima (1974) in an EMG study, reported that fortis stops are characterized by a sharp increase in thyroarytenoid activity before the stop release, which presumably resulted in increased tension of the vocal folds and constriction of the glottis during or immediately after the stop closure. In aspirated stops, all activity of the adductor muscles of the larynx was suppressed immediately after the articulatory release. A steep increase in activity of the adductor muscles always followed this suppression, presumably due to the movement into the position for voicing. In lenis stops, the suppression of adductor muscle activity is not significantly involved as compared with aspirated stops, and there is no transient increase in thyroarytenoid activity before the articulatory closure. These results suggest that a simple dimension of adduction-abduction of the vocal folds in characterizing Korean stops, implied in the studies of C. Kim (1970) and Kagaya (1974), is not sufficient, and another dimension is required.

Dart (1987) investigated the different aerodynamic properties of fortis and lenis stops in Korean. She measured intraoral air pressure and oral flow associated with the fortis and lenis stops in prevocalic position. One of the main results in her study is that the production of fortis stops is characterized by "a higher intraoral pressure before release, yet a lower oral flow after release," which is counterintuitive since in general higher intraoral pressure is associated with greater oral flow. Dart's aerodynamic modeling accounted for the pressure-flow relations by modeling fortis stops with tightly adducted vocal folds before the articulatory release and greater vocal tract wall tension.

As it became clear that the different types of stops are associated with different glottal configurations in the production of not only the stops themselves but also the onset of the following vowel, researchers began to raise the question of whether the voice quality of the vowel is influenced by the preceding consonant. They hypothesized that the voice quality of the vowel is similar to a breathy voice after the lenis stop and to a laryngealized or 'pressed' voice after the fortis stop. A laryngographic study by Abberton (1972) showed that the onset of vowels after fortis stops has some of the characteristics of creaky voice with a long closed phase and a slow opening phase. Han (1998) reported that vowels after lenis stops have a breathy voice as indicated by positive H1-H2 values (the difference in amplitude between the first and the second harmonics). On the other hand, vowels after fortis stops do not always have negative H1-H2 values, which, if present, would have indicated a laryngealized or pressed voice quality.

3. 2. *Fricatives*

There are three fricatives in Korean, **s**, **s***, **h**. As in the stops, there is no voicing contrast among fricatives, but unlike stops, there is only a two way contrast between denti-alveolar fricatives, **s** and **s***. **s*** is a tense or fortis fricative, but the categorization of **s** has been controversial. Korean orthography regards it as lenis and it behaves that way in phonological processes. But its behavior in phonetic processes and its phonetic realizations are generally believed to be similar to stops in the aspirated category. For example, **s** becomes tense after a lenis stop (e.g., /pak**s**a/ → [paks*a] 'Ph.D') as does the lenis stop (e.g., /pak**t**a/ → [pak**t***a] 'to pin down'). However, **s** is not likely to become phonetically voiced between voiced segments as lenis stops do, and it generally triggers a high tone in the beginning of an Accentual Phrase as does the aspirated stop (Jun 1993).

In its phonetic realization, **s** consists of two parts, frication and aspiration, as has been shown in spectrograpic studies (Yoon 1998, Park 1999). Fiberscopic data in Kagaya (1974) and Jun et al. (1998) showed that **s** has a glottal opening configuration similar to aspirated stops, and Jun et al. further showed that **s** has a larger glottal opening than **s***. Park (1999) reported that plain **s** has greater H1-H2 values than fortis **s***, suggesting that the vowel onset after plain **s** is breathier than after fortis **s***. However, the relative breathiness associated with plain **s** cannot be a direct metric for whether plain **s** has the characteristics of an aspirated segment, since the lenis segment, as noted earlier, can also be associated with breathy voicing in the following vowel onset.

4. Cheju data

In this section, we will describe the procedures employed to investigate the acoustic and aerodynamic properties of Cheju consonants, and the measurement criteria for both stops and fricatives. For stops, we will examine VOT, burst energy, f0, H1-H2, H1-F2 (the second formant), and oral pressure and flow data. (Affricates were included only for VOT measurement.) For fricatives, we will examine the duration, centroid of fricative noise, RMS energy, f0, H1-H2, and H1-F2, with particular reference to the contrasting nature of **s** and **s***.

All nine stops and three affricates were examined in the acoustic study. Each consonant was placed in initial and medial positions in the word, and was followed (and preceded, for word medial tokens,) by the open vowel **a.** In this paper, we will discuss word initial tokens only. Table 3 shows the word list in a phonemic transcription. Note that for the lenis **t**, two different words were used because some speakers were more familiar with one than the other.

Table 3. Stops and affricates recorded for acoustic measurements

p^{h}	**p^{h}amtʃə**	to dig
p*	**p*amtʃə**	to squeeze oil
p	**pataŋ**	sea
t^{h}	**t^{h}amtʃə**	to get (shy)
t*	**t*api**	ground
t	**talkwatʃəmtʃə**	to heat iron
	tɐlanamtʃə	to run away
tʃ^{h}	**tʃ^{h}amtʃə**	to kick
tʃ*	**tʃ*amtʃə**	to conspire
tʃ	**tʃamtʃə**	to sleep
k^{h}	**k^{h}amtʃə**	to get burned
k*	**k*amtʃə**	to peel
k	**kamtʃə**	to go

The fricatives **s**, **s***, **h** were recorded in initial and medial position, in the words shown in Table 4. As will be discussed in section 7, we also recorded all the other consonants of Cheju. Here we will simply note that the approximant **l** in word initial position (i.e., /latio/ 'radio') was used to provide a base line when comparing the voice qualities of vowels conditioned by the type of preceding consonant.

Table 4. Fricatives recorded for acoustic measurements

Word initial			Word medial		
s	**salamtʃə**	to live	s	**asakatʃə**	To take away
s*	**s*amtʃə**	to wrap	s*	**p^{h}as*ak**	Breaking noise
h	**hata**	many/much	**h**	**nahan**	(from) 'opek nahan'

For the aerodynamic study, a corpus with only bilabial stops was designed as shown in Table 5. Four sets of data were constructed to make the target consonants (**p**, **p**^{h}, **p***) vary in four prosodic positions -- word-initial vs. word-medial in isolation, and word-initial (i.e., phrase initial) vs. word-medial within a sentence.

Table 5. Word list for aerodynamic study

Word in isolation, word initial			Word in isolation, word medial	
p	**pe**	ship	**tʃapamtʃə**	to catch
p^h	**p^he**	card	**ap^hamtʃə**	sick
p*	**p*e**	bone	**pap*amtʃə**	busy
Sentence medial, word initial			Sentence medial, word medial	
p	**ikəsi petʃu**	This is a ship	**kai tʃapamtʃə**	He catches (object)
p^h	**ikəsi p^hetʃu**	This is a card	**kai ap^hamtʃə**	He is sick
p*	**ikəsi p*etʃu**	This is a bone	**kai pap*amtʃə**	He is busy

For the acoustic study, each word in Tables 3 and 4 was read twice by each speaker. For the aerodynamic study, the items in Table 5 were read four times each by each speaker (except for one male speaker in the rural area). Recorded materials were digitized at a sampling rate of 22,050 Hz and analyzed using the Kay Elemetrics's Computerized Speech Lab (CSL).

4.1. *Measurements for stops.*

•*Voice Onset Time:* VOTs for all stops and affricates were taken from the point of the stop release to the voice onset of F2 and higher formants in the following vowel, as seen in spectrograms. Thus for the lenis stops, any breathy voicing with only low-frequency harmonics was included in the VOT, and for the affricates, frication after stop release was included in the VOT because there was no clear cut between frication and aspiration.

• *Relative Burst Energy:* The acoustic energy at the burst and in the middle of the vowel were measured from an acoustic energy profile, using a 10 ms window. The percentage value of the burst energy relative to the energy at the midpoint of the vowel was employed to examine the characteristics of the stop release for the different types of stops. Greater burst energy for a stop can be expected in two cases — when a consonant (i.e., /t/) has a relatively small amount of linguopalatal contact, resulting in a fast release (Stevens, Keyser, & Kawasaki, 1986) and when the air flow is greater at the release (which is presumably due to a greater air pressure behind the constriction immediately before the release).

• *Fundamental Frequency (f0):* f0 was measured at the onset and the midpoint of the vowel, using the pitch track along with the first harmonic values from an FFT with a 25 ms window as supplementary checks. From the f0 differences, we would infer some physical information about the vocal folds (e.g., tension or stiffness) associated with different consonant types.

• *H1-H2 and H1-F2:* Energy values (dB) for the first (H1) and second (H2) harmonics, and the peak harmonic forming the second formant (F2) were taken at the onset of the vowel, using FFT spectra with a 25 ms window. All nine stops were included in this measure. In addition, harmonic values were also measured at the midpoint of the vowel after the liquid **l**, which was used as a control data representing modal voicing. The difference in amplitude between H1 and H2 has been frequently used to distinguish between breathy and modal voicings (e.g., Bickely 1982, Ladefoged 1983, Huffman 1987, Klatt & Klatt 1990, Blankenship 1997). Breathy voicing is produced with a relatively larger open quotient with the vocal folds remaining closed for a shorter time (e.g., open for 80-100% of the cycle for breathy phonation, vs. 65-70% for modal phonation (Childers & Lee, 1991)). As a result, the spectrum is dominated by the energy at the fundamental frequency, resulting in the amplitude of H1 being markedly higher than that of other harmonics. On the other hand, laryngealized vowels with creaky voicing (or 'pressed voicing' in Stevens' (1999) term) would have the opposite result, as they are produced with a smaller open quotient with the vocal folds remaining closed for a longer time (Blankenship, 1997). Thus, a greater H1-H2 would indicate breathiness of the vowel and a smaller or negative H1-H2, would indicate creakiness.

The spectral slope, as obtained from H1-F2, is an indicator of the abruptness of vocal fold closure. Blankenship (1997:17) explains that a gradual adduction of the vocal folds excites mainly the lower resonances of the vocal tract and, as a result, the sound wave is nearly sinusoidal with most energy near f0. The resulting spectrum has a steep downward slope from which a greater H1-F2 is expected. We expect that any breathy voicing should be associated with a gradual adduction with a greater H1-F2. On the other hand, when there is an abrupt adduction with the vocal folds coming together all at once, which is one of the characteristics of pressed or creaky voicing, the abruptness of the closure excites a wider range of frequencies. As a result, the sound wave has a spectrum with energy spread across a higher range of frequencies. Schematized spectra for different types of voicing are shown in Figure 2, based on Stevens (1999:86, 90).

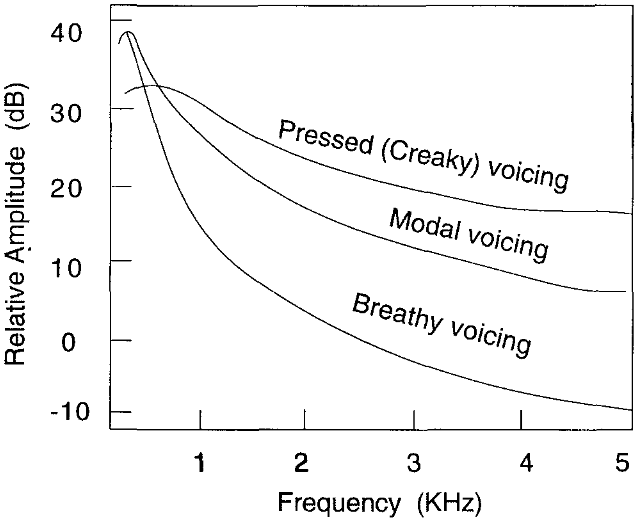


Figure 2. Schematized spectra for different phonation types, based on Stevens (1999:86, 90).

- *Oral Pressure and Flow:* Oral airflow and pressure were recorded using the Macquirer X16 system. Speakers held a face mask against the lower part of the face, below the nose, capturing all the oral airflow. They also held a tube (internal diameter 2 mm) between their lips to record the pressure of the air in the mouth. A microphone within the face mask recorded the audio signal. The flow and pressure signals were sampled at a rate of 2 kHz and the audio signal was sampled at 10 kHz. Speakers found little difficulty in producing the required phrases in these conditions, and the audio signal indicated that the utterances sounded reasonably natural. Only bilabial stops, as shown in Table 5, can be investigated in this way. (The speakers were too elderly and unaccustomed to phonetic experimentation to ask them to pass a tube through the nose so that pressures behind dental and velar closures could be recorded.) The maximum flow after the release of the closure and the peak oral pressure during the closure were measured, as indicated by the arrows (a) and (b) in Figure 3, respectively.

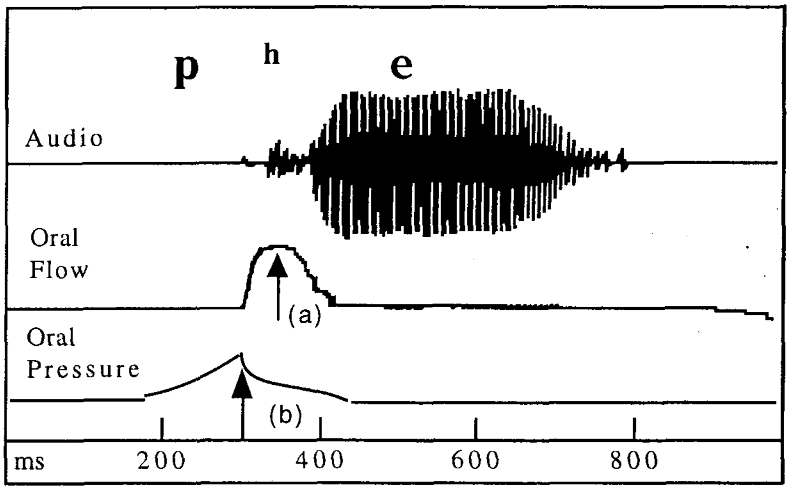


Figure 3. Waveforms of audio, oral airflow and oral pressure. The points marked by arrows indicate peak oral flow (a) and oral pressure (b).

Oral pressure and flow provide information about the degree of glottal constriction (Ladefoged, Maddieson & Jackson 1987). Other things being equal, the ratio of flow to pressure in a voiced sound should be constant. The ratio of flow to pressure will be higher if the vocal folds are held together more loosely, as in a breathy voiced vowel, and lower when the vocal folds are more tightly together as in a pressed or creaky vowel. As we have noted, Dart (1987) and others using these measures have found indications of differences in phonation type in the vowel onsets after the different Korean stops. We wanted to determine whether similar variations occurred in Cheju.

4.2. *Measurements for fricatives*

- *Fricative duration:* Duration measurements were made from the spectrograms of each token for s and s* both in word-initial and word-medial positions. For s, we also measured the duration of both frication and aspiration, separately and combined.
- *Centroid of the fricative noise:* The centroid is the center of gravity of a defined part of the spectrum, each frequency being weighted according to its amplitude. Centroid values were taken from FFT spectra in the frequency range from 500 Hz to 10,000 Hz, using a 25 ms window centered around the midpoint of the fricative portion of s*, **h**, and the midpoint of the fricative and the following aspiration for the plain s. Due to the low-frequency characteristics of **h** (Stevens 1999), we expect that **h** would be associated with a lower centroid. On the other hand, for denti-alveolar fricatives, s and s*, a higher centroid is expected as the source is filtered by the front cavity resonance, resulting in a spectrum peak in the vicinity of F4 or F5 (Stevens 1999). We were particularly interested in determining whether there was a difference in the centroid frequency of s and s*.
- *RMS energy:* The acoustic energy at the center of the frication was measured from an FFT spectrum giving the RMS value over all frequencies. A 25 ms window was centered around the midpoint of the frication for s and s*.
- *H1-H2 and H1-F2:* As measured for stops, H1-H2 and H1-F2 were measured at the onset of the vowels that follow s and s*.
- *Fundamental Frequency (f0):* As measured for stops, f0 was taken at the onset and the midpoint of the following vowel.

Along with these parameters that were analyzed quantitatively, qualitative observations of the spectrographic characteristics were made for each token.

5. Results and discussion for stops

5.1 *VOT*

We found a significant effect of consonant type on VOT (F(2, 180) = 280.571, $p < .0001$). Pairwise Fisher's PLSD post hoc comparison showed that VOT is shortest for the fortis stop, intermediate for the lenis stop, and longest for the aspirated stop ($p < .0001$ for each comparison). This finding is true across all places of articulation, as summarized in Figure 4. It remains true for stops at each place of articulation taken separately, except that there is no significant difference in VOT between the lenis and the fortis post-alveolar affricates, presumably because there was a trade off between the aspiration and the frication periods (Note that the frication component of an affrricate was included as VOT).

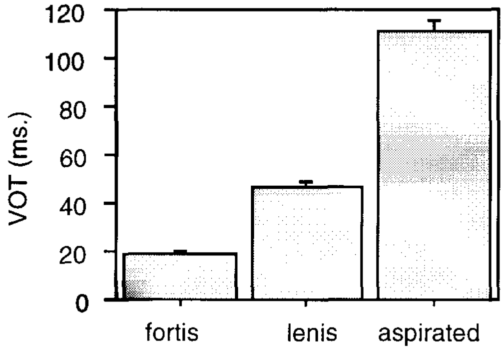


Figure 4. VOT for different types of stops and affricates. Data are pooled across stops/affricates, speakers and place of articulation. Error bars refer to standard errors.

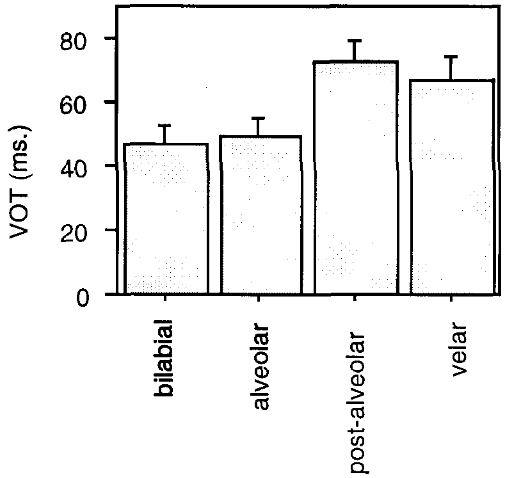


Figure 5. Variation in VOT depending on place of articulation.

Figure 5 shows the variation in VOT due to place of articulation. There is a significant effect of place of articulation on VOT (F(3, 180) = 15.253, p < .0001). Pair-wise Fisher's PLSD post hoc comparison showed that VOT of the velar stop is significantly longer than that of either the bilabial or the denti-alveolar stops. There was no significant difference between the bilabial and the denti-alveolar or between the post-alveolar and the velar VOT's. This is in general agreement with the tendencies noted in other languages, notably that the further back the closure, the longer the VOT (Fischer-Jørgensen 1954; see also Cho & Ladefoged, 1999, for an extended discussion on VOT).

5.2 *Relative burst energy*

We found a significant effect of the consonant type on the relative burst energy (F(2, 135) = 11.091, p < .0001). Fisher's PLSD pairwise post hoc comparisons revealed that the aspirated stop has significantly greater burst energy than the other two types of stops while there is no difference between the fortis and the lenis stops, though there is a trend for the lenis stop to have a greater burst energy than the fortis stop, as shown in Figure 6.

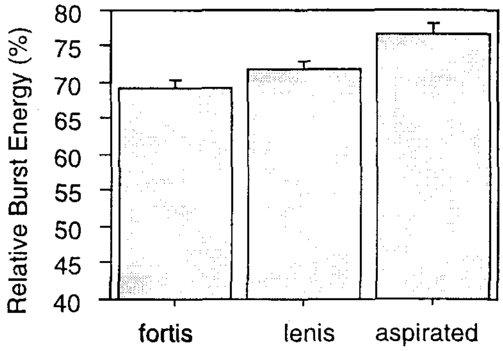


Figure 6. Relative burst energy (%) for different types of stops.

As mentioned earlier, a greater burst energy is due either to a smaller linguopalatal contact (Stevens et al. 1986) (and, therefore, a greater speed of movement) or a greater air flow at the release. Cho and Keating (1999) found that both fortis and aspirated Korean stops have greater linguopalatal contact than the lenis stops while there is no difference between the fortis and aspirated stops. If the greater contact induces slower release, which is in turn associated with lesser burst energy, then both fortis and aspirated stops should have lesser burst energy than lenis stops, all else being equal. Our data, however, show that only fortis stops have lesser burst energy than lenis stops. This indicates that the amount of the contact itself is not the primary source for the greater burst energy.

Dart (1987) found that the lenis stop has a greater air flow at the release than does the fortis one, though the fortis stop has a greater air pressure before the release. (A similar result was found in our own aerodynamic study. See below.) This explains why the lenis stop tends to have a greater burst energy than the fortis stop in our data. The lenis stop has a greater air flow at the release which is responsible for the greater burst energy. We can also expect that the aspirated stop is produced with a greater oral pressure and a greater flow than the lenis stop at the release, which would lead to a greater burst energy. In fact, our aerodynamic data show that the aspirated stop is produced with the greatest air flow at the release (see Fig. 17).

This aerodynamic explanation of the burst energy is supported by the observation that there is a significant effect of place of articulation (F(2, 141)=6.787, p =.0015 across speakers). The burst energy for the velar stop is significantly greater than that for both the bilabial and denti-alveolar stops (p < .001, Fisher's PLSD posthoc comparisons) between which there is no difference, as shown in Figure 7. The velar stop, which is usually produced with a relatively greater linguopalatal contact, might be expected to have smaller burst energy. This would be the case if the amount of linguopalatal contact and its consequent slow speed of the release movement is the primary factor responsible for the burst energy difference. But the cavity behind the constriction of the velar closure is relatively smaller than those of other stops, and, as a result, will have relatively greater air pressure buildup before the release. The greater air pressure would give higher air flow volume velocity, which appears to be responsible for the greater burst energy for the velar stop. However, we should not ignore the possibility that the difference may be due to the shape of the constriction just after the release.

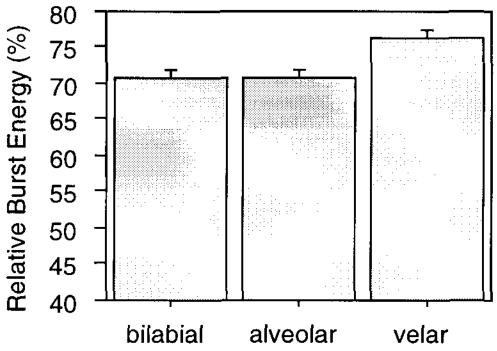


Figure 7. The relative burst energy (%) at different places of articulation.

5.3 *f0 differences*

We found that f0 in both the onset and the midpoint of the vowel is significantly influenced by the type of preceding consonants (F(2, 135) = 27.999, p < .0001 for the onset; F(2, 116) = 13.461, p < .0001 for the midpoint). Each consonant type is associated with significantly distinctive f0 in the onset of the following vowel at least at the p < .05 level while, in the midpoint of the following vowel, a significant difference is maintained between the lenis and the other two stops, but not between the fortis and the aspirated stops. As shown in Figure 8, for the vowel onset, lenis stops have smaller f0 values than fortis stops which in turn have smaller f0 values than aspirated stops. Note, however, that the difference between the fortis and the aspirated stops is smaller than that between the lenis and the other stops. For the vowel midpoint, similar patterns are observed except that the difference between the fortis and the aspirated stops is not significant.

The substantial difference in f0 between the lenis and other stops cannot be understood in terms of the phonetic pitch perturbation caused by the voicing of the preceding consonant (i.e., microprosody) as found in many other languages (Hombert 1978, Hombert et al. 1979). The f0 difference between the groups shown in Figure 8 is substantially greater than that found in other languages, and the f0 difference is still present at the mid point of the syllable. Furthermore, all three stops are voiceless.

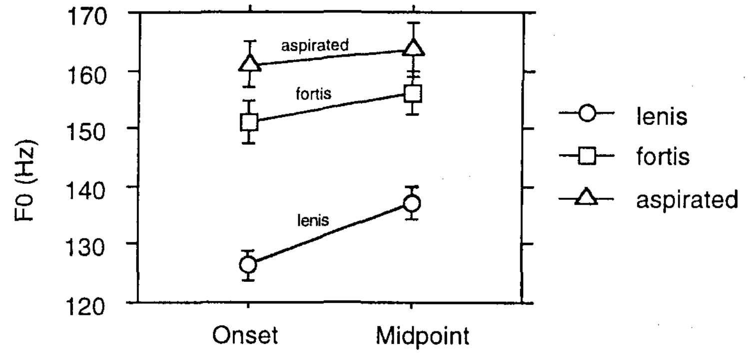


Figure 8. f0 differences in the onset and the midpoint of the following vowels.

The same f0 pattern was found in Seoul Korean (Jun 1996a). Jun (1996b) claims that this segmentally triggered f0 distinction in Seoul Korean is phonologized in the intonation system of Seoul Korean as described by Jun (1993, 1996a, 1998). In this intonation system, the Accentual Phrase is realized in two distinct tonal patterns, LHLH and HHLH, where the phrase initial tone (L vs. H) is determined by the type of the consonant occurring in the onset of the phrase initial syllable. When there is a fortis or aspirated obstruent in the onset, the Accentual Phrase starts with H, and otherwise (i.e., when there is a lenis obstruent, a sonorant consonant, or a vowel), it starts with L. This phonologization of microprosody seems to occur in Cheju, though we do not yet know the detailed intonation system of Cheju.

5.4 *H1-H2 and H1-F2*

There is a significant effect of consonant type on H1-H2 (F(2, 135) = 27.419, p < .0001). Pairwise *posthoc* comparisons by Fishers PLSD showed that stops differ significantly from each other at p < .05. H1-H2 is greater (positive) for lenis stops, intermediate for aspirated stops, and smaller (negative) for fortis stops, as shown in Figure 9. In other words, the onset of the vowel (the first 30 ms. period) has a more breathy voicing with a larger open quotient of the vocal folds immediately after the lenis stop, and more pressed (creaky) voicing with a smaller open quotient after the fortis stop. The voice quality of the vowel after the aspirated stop is close to the modal voice. Its H1-H2 value is about the same as the control (modal) value taken from the midpoint of the vowel after l which has no known effect on the phonation type of the following vowel. What is especially interesting about this finding is that it indicates that both the fortis and the lenis stops are associated with non-modal phonation.

Now let us consider the spectral slope (H1-F2), a hypothesized indicator of the abruptness of vocal fold closure. Results show that there is a significant effect of consonant type on H1-F2 (F(2, 135) = 9.939, p < .0001). The fortis stop has negative H1-F2 while the lenis and aspirated stops have positive H1-F2, as shown in Figure 10. This indicates that the closing gesture of the vocal folds during the voicing is more abrupt for the fortis stop than for the other stops. The folds come together more rapidly after the fortis stop, providing more energy in the second formant region, presumably because they are under greater tension. Interestingly, however, the expected higher H1-F2 associated with the breathy voicing for the lenis stop was not observed as compared to H1-F2 for the aspirated stop. This indicates that the H1-H2 (an indicator of the open quotient) and the H1-F2 (an indicator of the abruptness) do not necessarily go hand in hand. In fact, there is only modest correlation between the two parameters (R = .466).

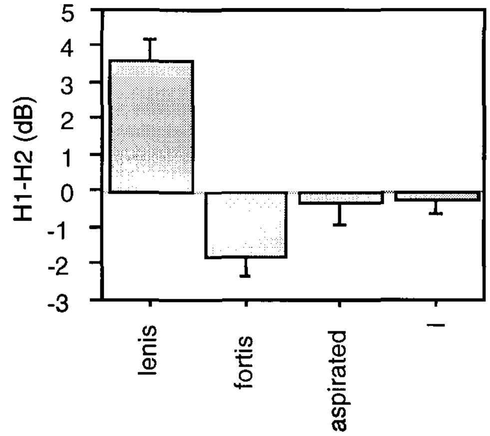


Figure 9. Difference between the amplitudes (dB) of the first harmonic (H1) and the second harmonic (H2) for different stop types and /l/. Error bars indicate standard errors.

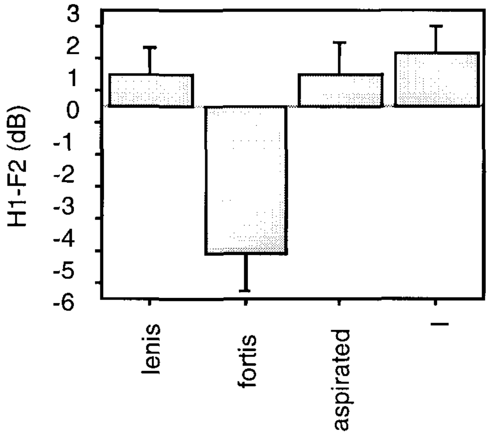


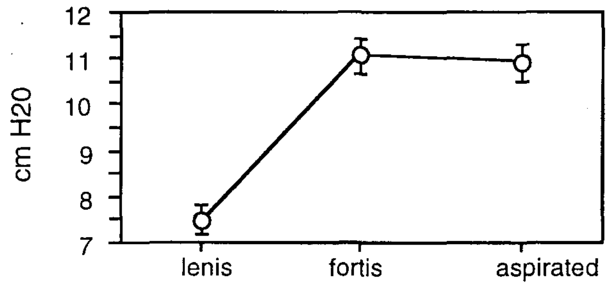
Figure 10. Difference between the amplitudes (dB) of the first harmonic (H1) and the second formant (F2). Error bars indicate standard errors.

5.5 *Oral Pressure and Flow*

Both oral pressure and flow were found to be significantly influenced by the consonant type (F(2, 317) = 28.985, p < .0001 for oral pressure; F(2, 317) = 87.257, p < .0001 for air flow). As shown in Figure 11a, the peak oral pressure during the closure is smaller for the lenis stop than for the fortis and aspirated stops, while there is no difference between the latter two. On the other hand, the lenis stop is produced with a slightly greater flow than the fortis stop, as found by Dart (1987), but with a much smaller flow than the aspirated stop, as shown in Figure 11b.

The fact that the fortis stop has greater oral pressure but little flow can be accounted for by several factors such as the glottal impedance (which depends largely on the glottal area), the vocal tract wall tension, and a possible increase in the subglottal pressure, as discussed in Dart (1987). The smaller glottal opening area (Kagaya 1974, Jun, Beckman, and Lee 1998) would have the effect of decreasing flow through the glottis. A greater degree of vocal tract wall tension (C. Kim, 1965; Hardcastle, 1973) also contributes to an increase in oral pressure. Dart posited that the pressure increase is due to stiffening the wall, which presumably reduces passive vocal tract expansion. The suppression of the vocal tract expansion is effectively the same as reducing the supraglottal cavity volume which otherwise would be expanded under normal circumstances. The stiffened vocal tract wall can also result in a decrease in the volume velocity of air flow since the reduction in the elasticity of the walls would also reduce the amount of elastic recoil of the walls, which would in turn cut down the initial flow volume velocity at release. Finally, the pressure increase can be understood as the result of heightened subglottal pressure, as suggested by C. Kim (1965) and Kagaya (1974). Dart's model suggested that the heightened subglottal pressure is primarily due to a larynx raising caused by a more rapid increase in respiratory muscle force, although it is also possible that any subglottal pressure increase can be a result of larynx lowering, caused by expanding the supraglottal cavity.

(a) Oral Pressure (cm H20)



(b) Air Flow (1/msec.)

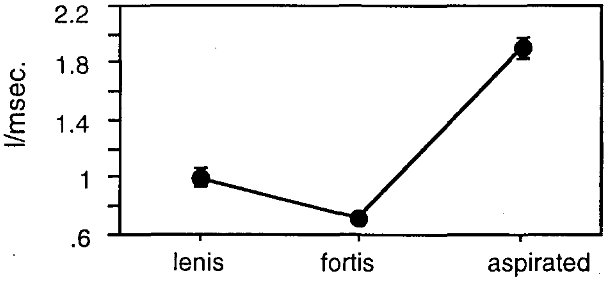


Figure 11. (a) Oral pressure (cm H_20) and (b) air flow (l/msec) with data pooled across prosodic positions and speakers. Error bars indicate standard errors.

Now let us look at aerodynamic characteristics of the aspirated stops, which were not included in Dart's study. Aspirated stops are characterized by an increase in both oral pressure and flow, as shown in Figure 11. The greater oral pressure and flow can be accounted for by a larger vocal fold opening (Kagaya 1973, Jun et al. 1998) and a greater subglottal pressure due to an increase in the respiratory muscular force. As C. Kim (1965) noted, there may be a greater vocal tract wall tension during the closure of the aspirated stop as in the case of the fortis stop, which will contribute to an increase in the oral pressure. However, as discussed above, the greater wall tension would counteract the increase in flow velocity immediately after the stop release. It is conceivable that the effect of glottal area function and the respiratory muscular force is large enough to override that of the vocal tract wall tension, if there is any.

To examine the position effect on aerodynamic parameters, the data were submitted to three-way ANOVAs with factors, Consonant Type, Position-in-Word (initial vs. medial), Position-in-Sentence (word in isolation vs. word inside a sentence). First, the results show that there is a significant effect of Position-in-Word on flow ($F(1, 308) = 33.535$, $p < .0001$) — air flow was significantly greater for word-initial stops than for word-medial stops. But, there is no significant effect of Position-in-Word on oral pressure. Second, there is no effect at all of Position-in-Sentence on either oral pressure or flow — i.e., no difference between the word in isolation and the word inside a sentence. In addition, there is no significant interaction between Position-in-Word and Position-in-Sentence. Third, a significant interaction between Consonant Type and Position-in-Word is found for both oral pressure (F(2, 308) = 12.568, p < .0001) and flow (F(2, 308) = 18.788, p < .0001). The effect of Position-in-Word and its interaction with Consonant Type is illustrated in Figure 12. Note that since there is no effect of Position-in-Sentence (word in isolation vs. word inside a sentence), we present data for the word-initial stops pooled across Position-in-Sentence.

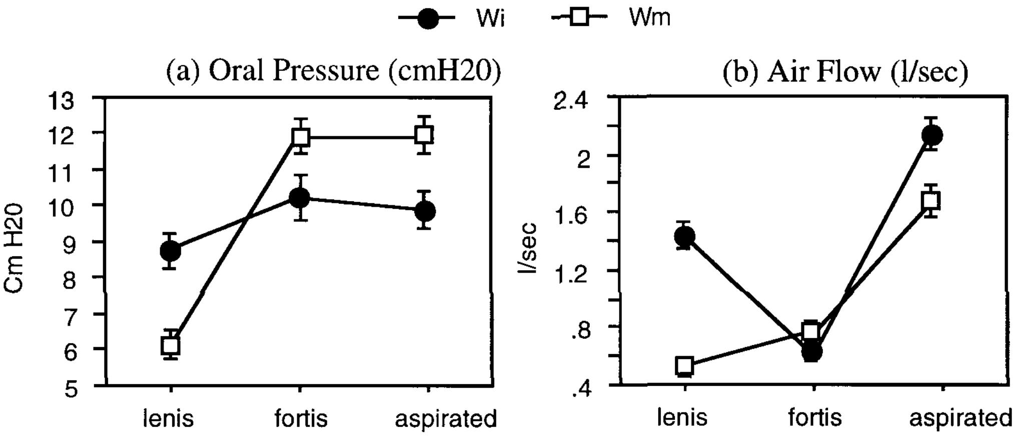


Figure 12. Difference between word-initial (Wi) and word-medial (Wm) positions for oral pressure (a) and flow (b) for three stop types, pooled across Position-in-Sentences and speakers.

The word-initial lenis stop has a greater oral pressure and flow than the word-medial one, regardless of where it occurs in isolation or sentence-medially. The lower oral pressure and flow for the word medial lenis stop can be accounted for by the intervocalic voicing for the lenis stop. This clearly suggests that there is an effect of weakening in medial position. Seen from a different angle, however, this effect can be understood as strengthening in initial position. The glottal opening gesture is larger in word-initial position, which explains the greater oral pressure and flow, compared to word-medial position, where there is voicing. This agrees with Jun et al.'s (1998) fiberscopic study of glottal configurations of Korean obstruents, where they found that Accentual Phrase initial lenis and aspirated stops are produced with a larger glottal aperture than Accentual Phrase medial ones. A similar case was found for English by Cooper (1991).

For both fortis and aspirated stops the oral pressure is smaller for word-initial stops than for word-medial ones. This appears to be due to the geminate characteristics of the fortis and aspirated stops word-medially. In Korean, it has been reported that word-medial tense and aspirated stops in general lengthen (e.g., Silva, 1992; Han, 1996; Cho, 1998; Cho & Keating, 1999). Cho & Keating (1999) also found that linguopalatal contact for the fortis and aspirated stops tends to be greater word-medially than word-initially (but Accentual Phrase medially), which could perhaps be correlated with more forceful articulation, compared to word-initial ones.

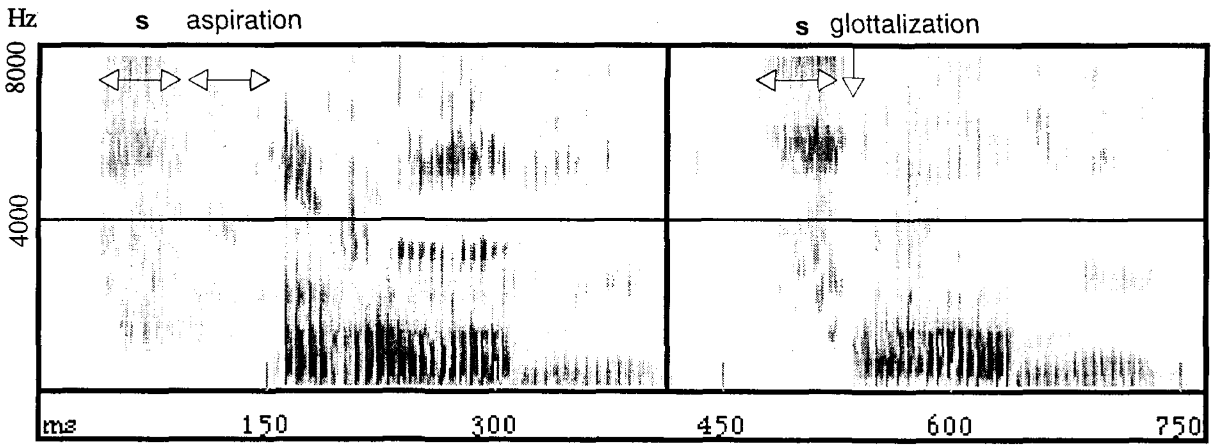
For the air flow, there is no significant difference between initial and medial fortis stops, as shown in Figure 12b. This is presumably because the glottal aperture is too small to be effectively different between the word-initial and word-medial positions. This is also supported by the fact that there is no significant difference in VOT between word-initial and word-medial fortis stops (Cho & Keating, 1999). For the aspirated stops, however, the word initial stops have a greater flow than word medial ones. This is presumably because the glottal aperture is larger at word initial position than medial position (Kagaya 1974, Jun et al. 1998).

6. Results and discussion for fricatives, s and s*

Let us first examine the acoustic characteristics for s and s* that can be qualitatively observed in the spectrograms. Spectrograms of word-initial s and s* are given in Figure 13. The plain s consists of both frication and aspiration, making it a complex, aspirated segment. This complex nature of the plain s was observed in most of the tokens examined. This suggests that the plain s is phonetically aspirated, agreeing with the previous studies (e.g., Kagaya, 1974; Jun et al, 1998; H. Park, 1999). However, it does not behave like the aspirated stops in that it does not retain a period of aspiration word-medially. The plain s is unaspirated when it occurs word-medially, as shown in Figure 14a. In these circumstances it has intervocalic weakening, like the lenis stops. Furthermore, although it is commonly supposed that s in general does not become voiced intervocalically, we observed about 38% of tokens for s being voiced in this position—further evidence that s behaves as a lenis segment. Two examples of the fully voiced s between vowels are given in Figure 15.

The fortis s* also has two components. As shown in Figures 13b and 14b, it has a frication portion similar to that of the plain s, followed by a vertical gap immediately before the vowel onset. The gap appears to be an indication of the glottalization for s*. Figure 16 shows examples of glottalization spanning over the first several cycles of the following vowel. There were a total of 9 out of 16 tokens showing a clear gap at word-initial position. A similar gap was found for word-medial tokens in 8 out of 16 tokens.

Such a glottalization gap was not generally observed when we examined fortis stops. In fortis stops the vocal folds close just before the oral release and start vibrating immediately for the following vowel, as implied by a short VOT. However, as noted by Jun et al. (1998), for the Korean fortis fricative (as well as the plain one) the glottis must remain open to a certain degree in order to permit sufficient air flow for the frication noise. Consequently it seems that it is only in the last part of s* that the glottal characteristics of a fortis consonant are present.



(a) plain S

(b) fortis S*

Figure 13. Spectrograms of s and s* in word-initial positions (speaker S6).

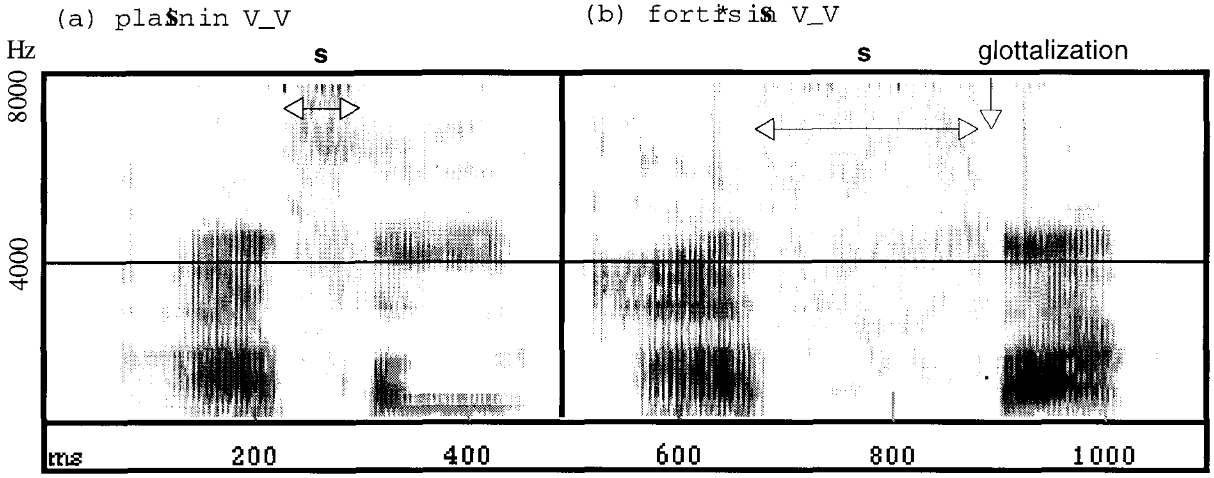


Figure 14. Spectrograms for s and s* word-medially between vowels.

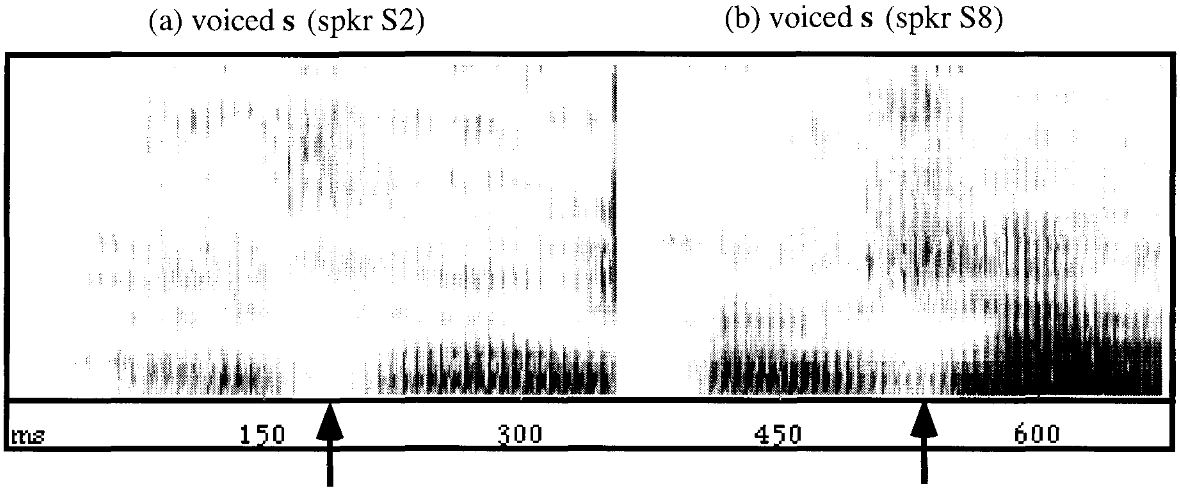


Figure 15. Spectrograms for fully voiced intervocalic s.

6.1 *Fricative duration*

For the word-initial position, the duration of the plain s (frication plus aspiration) is significantly longer than that of the fortis s* (F(1,30) = 11.097, p = .0023), as shown in Figure 16. This is the opposite of the difference in closure duration between the lenis and the fortis stops; that is, the closure of the fortis stop is longer than that of the lenis stop at word initial position (Cho & Keating, 1999). However, if we exclude the aspiration period, the difference becomes insignificant as shown in the figure. When the data were separated by the speaker, only two of eight speakers (S4, S6) show a significant duration difference between the plain s (excluding aspiration) and the fortis s*. This is contradictory to Yoon (1998) who found that the fortis s* is significantly longer than the plain s in non-high vowel context.

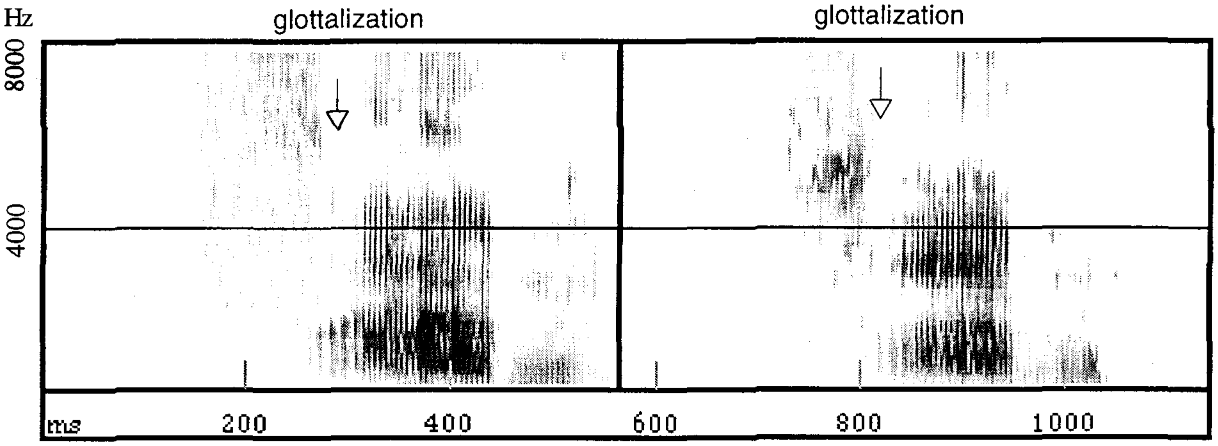


Figure 16. Spectrographic evidence of glottalization for s* in word-initial positions in **s*amtʃə**, produced twice by speaker S4.

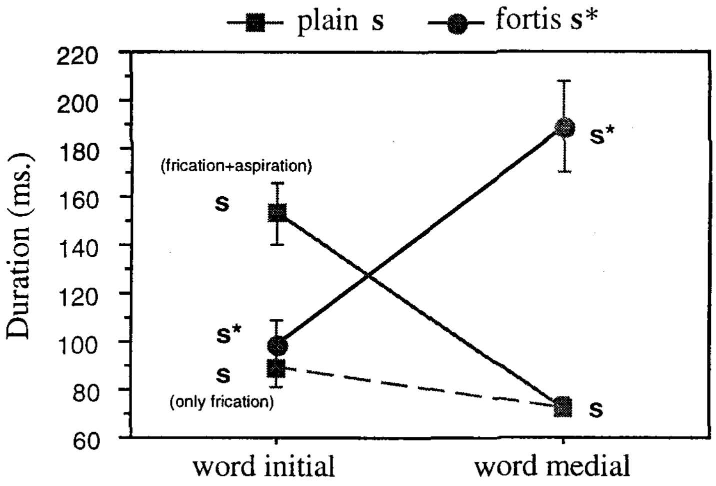


Figure 17. Duration of the plain s and the fortis s* in word-initial and word-medial positions. For the word-initial plain s, the duration with and without aspiration is shown together, with the latter marked by a dashed line. Error bars indicates standard errors.

The duration difference between s and s* is even larger in word-medial position (F(1, 30) = 36.415, p < .0001), but in the opposite direction, as also shown in Figure 17. The word-medial, intervocalic s* is about twice as long as the word-initial one, showing the geminate nature of the fortis segment, which is phonetically realized with a longer duration intervocalically (see Cho & Keating, 1999, for the seal closure duration difference between word-initial and word-medial fortis stops in Korean). In contrast to the increase in duration for the word-medial fortis s*, there is a decrease in duration for the plain s. The decrease is primarily due to the absence of the aspiration intervocalically. The duration of the word-medial s is not significantly longer than that of the word-initial s excluding the aspiration period.

6.2 *Centroid Frequency*

There is a trend toward a higher centroid frequency for **s*** than for s, as can be seen in Figure 18. However, there is also a substantial speaker variation--only four out of eight speakers produce s* with a higher centroid frequency as compared to **s**. We infer that, at least for those who show this difference, **s*** is produced with a relatively smaller front cavity, which presumably results in a higher centroid frequency. Another interesting fact lies in the similarity between **h** and the aspiration portion for **s**, (i.e. **h** in **s**) as can be seen in the figure. This indicates that the plain **s** in Cheju (as well as most of dialects in Korean) has two components, frication and aspiration, and has a similar acoustic pattern as other aspirated obstruents.

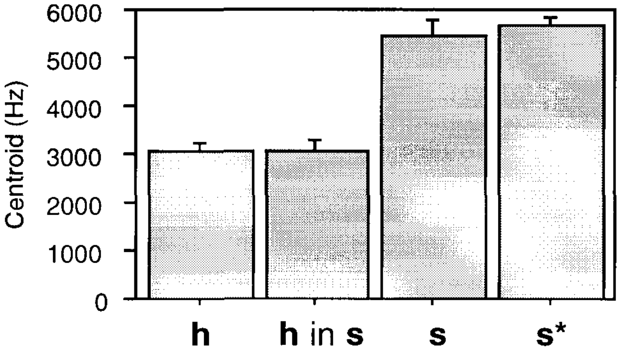


Figure 18. Centroid frequencies for **h**, **h** in **s**, **s**, and **s*** in word-initial positions.

6.3 *RMS Energy of frication*

There is no RMS energy difference between the plain **s** and the fortis **s***.

6.4 *Fundamental Frequency (f0)*

f0 in the onset of the vowel is significantly higher after the fortis **s*** than after the plain **s** (t=-2.242, p <.05), as shown in Figure 19. However, there is no significant difference at the midpoint of the vowel. As we have seen, the plain **s** is often aspirated, but it does not have a high f0 as other aspirated obstruents (see Figure 9). It also has a lower f0 than fortis **s***, unlike the higher f0 for aspirated stops than fortis stops. Although the effect of the aspiration of plain **s** on f0 is not as great as that of the aspiration in aspirated stops, there still remains some effect of the aspiration in plain s. In a separate t-test comparing the fricative and stops, f0 after the plain **s** is significantly higher than that after the lenis stop (t=2.411, p < .05). f0 values after the plain **s** are closer to the fortis **s*** in Jun (1999) where Seoul Korean female speakers, in their 20s and 30s, produced fricatives and stops in an Accentual Phrase initial position. Jun also shows that the average f0 of both fricatives is higher than that of fortis stops. A low f0 value after the plain **s** in Cheju may be due to a dialect and/or an age difference.

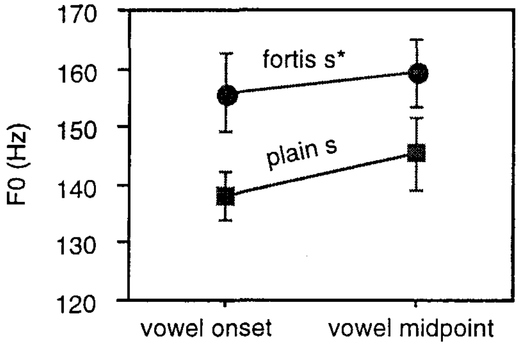


Figure 19. f0 difference between s and s* in the onset and the midpoint of the following vowel. Error bars indicate standard errors.

6.5 *H1-H2 and H1-F2*

As shown in Figure 20, there is a significant H1-H2 difference between s and s* in the onset of the vowel (t=4.04, p =.0003) — H1-H2 is positive after s and negative after s*. This suggests that the vowel onset has breathy voicing after s but 'pressed' voicing after s*. The average H1-H2 value after the plain s (2.5 dB) is closer to that after the lenis stop (3.5 dB) than that after the aspirated stop (-0.3), showing that the plain s has breathy voicing in the onset of the vowel in common with the lenis stop, not with the aspirated stop. There is no difference in H1-H2 between s and s* in the midpoint of the following vowel.

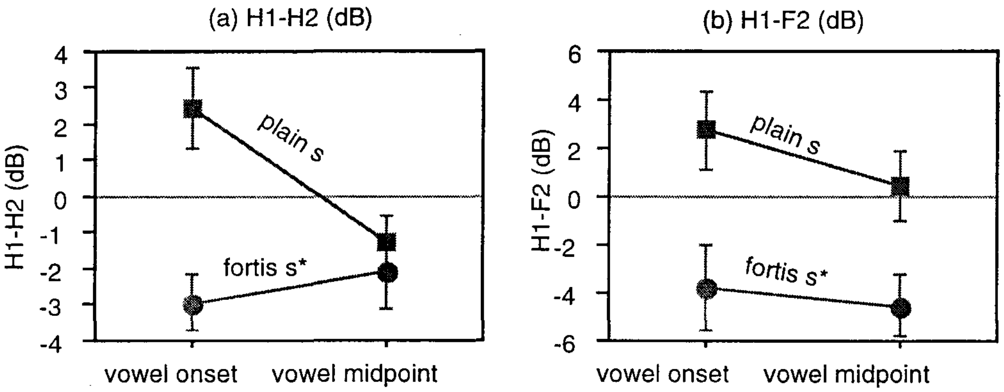


Figure 20. (a) difference in dB between the first and the second harmonic (H1-H2) and (b) difference in dB between the first harmonic and the second formant (H1-F2). Error bars indicate standard errors.

The plain s and the fortis s* differ also in H1-F2, an indicator of the abruptness of the vocal fold closure — H1-F2 is positive after s and negative after s*. The difference is maintained significantly in both the onset and the midpoint of the vowel (t= 2.745, p =.0101 for the onset; t=2.499, p = .0194 for the midpoint). Put differently, vowels after the fortis s* are produced with abrupt vocal fold closure as are vowels after fortis stops, and the abruptness after s* continues into the middle of the vowel. That is, the pressed nature of the vowel seems to be more evident after s* than after fortis stops.

6.6 *Summary on phonetic natures of s and s**

We have examined some phonetic properties that may contribute to distinguishing the contrasting segments **s** and **s*** in Cheju. The phonetic nature of the fortis **s*** is rather straightforward: it is produced with (1) frication and glottalization immediately before the vowel, (2) higher f0 in the vowel onset, (3) 'pressed' voicing (negative H1-H2) at the vowel onset, and (4) abrupt vocal fold closure during the following vowel (negative H1-F2).

The phonetic nature of the plain **s** is quite complex and ambiguous. It is produced with frication, aspiration, and f0 at the vowel onset relatively high as compared to a lenis stop. This might make one believe that it is an aspirated **s** appropriately transcribed as [s^h]. However, we claim that the plain **s** may well be categorized as a lenis segment for several reasons. First, its counterpart lenis stop is also produced with a fair amount of aspiration. Second, the onset of the vowel after **s** has a similar breathy voice quality to vowels after the lenis stop. Third, there is a relatively low f0 in the vowel onset after **s**, compared to f0 in the vowel onset after an aspirated stop, though higher compared to f0 after a lenis stop. Fourth, interestingly enough, the plain **s** loses its aspiration word-medially as does the lenis stop. Finally, though in general it does not become voiced, we have observed about 38% of tokens fully voiced word-medially. Considering all these observations, it seems that the plain **s** can be better labeled as a lenis fricative, rather than an aspirated fricative.

The phonetic nature of the frication is much the same for both plain and fortis fricatives in word-initial positions. There is no significant difference in centroid frequency, RMS energy, or the duration of frication. Yoon (1998) reports inconsistency in parameters such as the acoustic energy of the frication and the frequency distribution of the spectral peaks. It appears that the substantial difference then lies only after the frication period and in the onset of the following vowel, as just discussed above.

7. Other consonants

The Cheju segments that we have not yet discussed are the nasals and the lateral exemplified in Table 6, and the glides exemplified in Table 7. The velar nasal **ŋ** occurs only in coda position, and the same is true for the lateral **l**, except for its occasional occurrence initially in loan words. The lateral is usually realized as a flap word-initially and intervocalically. We did not find any unusual phonetic features specific to Cheju nasals and lateral.

Table 6. Word list for nasals and lateral

	Wd initial		Wd medial	
m	**mataŋ**	ground	**samɨla**	to weave straw
n	**naamtʃ*ə**	to give a birth	**anɨla**	to hug
ŋ			**paŋe**	grinding tool
l	**latio**	radio	**ɑlamtʃ*ə**	to know
			tallɨtʃu	different

The glides are similar to those in Korean. We recorded each of the glides **w, j, ɰ** before each of the vowels, as permitted by the phonological constraints. (In Korean, glides are sometimes analyzed as a consonant and sometimes as part of the vowels. In this paper we take no stand on either of these positions.)

Table 7. Word list for glides

wi	**wi**	stomach
we	**we**	cucumber
wæ	**wæullɨmsə(ke)** **wæullɨmtʃə**	to shout
wə	**wəllamɐl**	horse (colorful)
wa	**wa**	to stop (to horse)
je	**jesun**	sixty
jæ	**jæja**	Child! (vocative)
jə	**jə**	stone
ja	**jakatʃi** / jakeki	neck
jɐ	**jɐtap**	eight
jo	**jo**	blanket
ju	**juktʃi**	land
ɰi	**ɰisa / ɰitʃa** **ɰikwili**	doctor / chair name of a village

As in nasals and lateral, there is nothing particularly noteworthy about these glides except for the back unrounded glide, **ɰ**, a rare glide in the languages of the world. The back unrounded glide occurs only with **i**, as shown in Table 7. In casual speech, however, **ɰi** is usually monophthongized as either [i] or [ɨ] by most speakers. A few rural speakers in the experiment produced this glide and the labio-velar glide, **w**, distinctively as shown in Figure 20. In Figure 20a, for **wi** 'stomach,' formants (F2, F3, F4) are lowered at the beginning, an indication of the rounded glide **w**, followed by the steady state of **i**. In Figure 20b, for **ɰi** in **ɰikwili** 'name of a village', there is a much lowered F2 at the beginning but no observable lowering of the higher formants. A clear transition in F2 from **ɰ** to ɪ **is shown in the figure.**

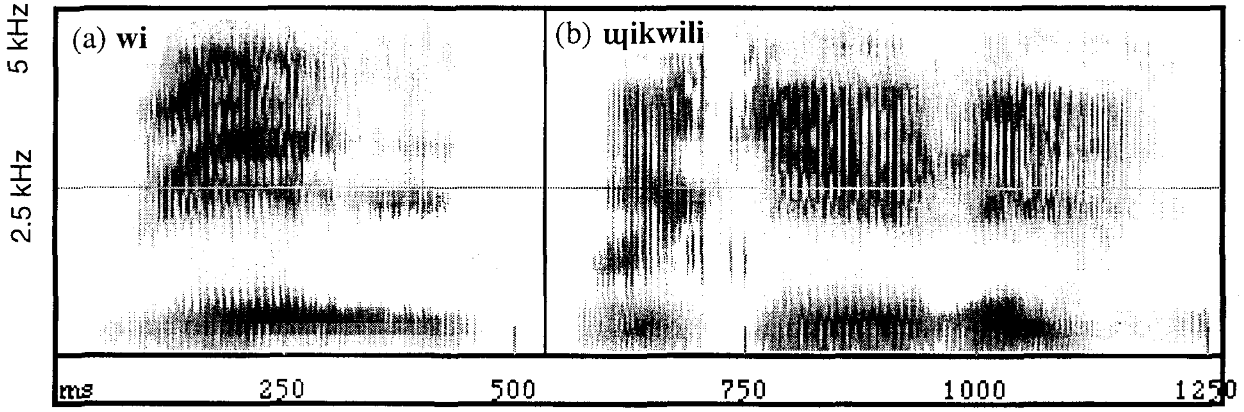


Figure 20. Spectrograms of (a) **wi** 'stomach' and (b) **ɰikwili** 'name of a village.'

8. Closing Remarks

Cheju and Korean are not mutually intelligible, but the consonants of the two languages seem to be the same in every respect. The opposition between the fortis, lenis and aspirated obstruents is realized in the same way, and the phonetic realization of all consonants are the same as those found in Korean.

Among findings reported in this paper, three points are particularly worth recapitulating. First, the voice quality of the vowel varies systematically depending on the preceding consonant type. This suggests that the voice quality may be one of the strong perceptual cues for listeners to use in differentiating the three way contrast in Korean stops. This finding is in agreement with Cho (1995, 1996) who reported that Korean listeners were able to identify the stop category at a far better than chance level by only listening to vowel stimuli following the stop. Second, the Cheju stops can be differentiated by aerodynamic characteristics. While two 'strong' strops (i.e., aspirated and fortis) are equally strongly articulated on a supralaryngeal level (cf. Cho & Keating, 1999), aerodynamics suggests that different mechanisms are employed in producing these stops. Finally, our study suggests that the fricative **s** is closer to a lenis category than an aspirated category in terms of its phonetic realization. The breathiness of the vowel onset after **s** and the property of intervocalic lenition are most appealing evidence in favor of the lenis category.

Acknowledgements

We appreciate Drs. Chang-Myong Oh and Seung-Chul Jung helping us record in Cheju city, and Dr. Won-Bo Kim who not only helped us record in the rural areas, but also gave us great assistance in traveling around the island. We are also grateful to all the Cheju speakers who recorded data for us. May they and their language continue to flourish.

References

Abberton, E. 1972. "Some laryngographic data for Korean stops." *Journal of International Phonetic Association*, 2, 67-78.

Bickely, C. 1982. "Acoustic analysis and perception of breathy vowels." *Speech Communication Group Working Papers I.* Research Laboratory of Electronics, MIT, 73-83

Blankenship, B. 1997. *The time course of brethiness and laryngealization in vowels.* PhD dissertation. UCLA.

Childers, D.G. & C.K. Lee. 1991. "Vowel quality factors: analysis, syntheses, and perception." *Journal of the Acoustical Society of America*, 90(5), 2394-2410.

Cho, T. 1995. "Korean stops and affricates: Acoustic and perceptual characteristics of the following vowel." Paper presented at 130th Meeting of the Acoustical Society of America. St. Louis, MO. 1995.

________. 1996. *Vowel Correlates to Consonant Phonation: An Acoustic-Perceptual Study of Korean Obstruents.* M.A. Thesis, University of Texas at Arlington

________. 1998. "Domain-initial strengthening in the prosodic hierarchy: an EPG study." *Proceedings of 11th International Conference on Korean Linguistics (ICKL 11),* July 6-9, 1998. University of Hawaii at Manoa.

Cho, T. & P. Keating. 1999. "Articulatory and acoustic studies of domain-initial strengthening in Korean." *UCLA Working Papers in Phonetics, 97,* 100-138.

Cho, T. & P. Ladefoged. 1999. "Variations and universals in VOT: evidence from 18 languages." *Journal of Phonetics*, 27(2), 207-229.

Jung, S.-C. 1995 *Cheju-do Pang-eon-euy Thong-shi-eum-un-lon (A Diachronic Approach to the Phonological Processes of Cheju Dialect)* Kuk-Eo-Hak-Hoi, Seoul, Korea.

Cooper, A. 1991. *Glottal gestures and aspiration in English.* PhD dissertation. Yale University.

Dart , S. 1987. "An aerodynamic study of Korean stop consonants: Measurements and modeling." *Journal of the Acoustical Society of America*, 81(1), 138-147

Fischer-Jørgensen, E. 1954. "Acoustic analysis of stop consonants." *Miscellanea Phonetica*, 2, 42-59.

Han, J.-I. 1996. *The Phonetics and Phonology of "Tense" and "Plain" Consonants in Korean.* Ph.D. Dissertation. Cornell University.

Han, N. 1998. *A comparative Acoustic study of Korean by native Korean children and Korean-American children.* MA thesis. UCLA.

Han, M.S. & R.S. Weitzman. 1970. "Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/." *Phonetica*, 22, 112-28.

Hardcastle, W.J. 1973. "Some observations on the Tense-Lax Distinction in Initial Stops in Korean." *Journal of Phonetics*, 1, 263-271

Hirose, H., C.Y. Lee & T. Ushijima. 1974. "Laryngeal control in Korean stop production." *Journal of Phonetics*, 2, 145-152.

Hombert, J.-M. 1978. "Consonant types, vowel quality, and tone." In V. Fromkin (ed.), *Tone: A linguistic survey*, pp. 77-111. NY: Academic Press.

Hombert, J.-M., J. Ohala and E. William. 1979. "Phonetic explanations for the development of tones." *Language*, 55, 37-58

Hong, Y. 1991. *A sociolinguistic study of Seoul Korean.* PhD dissertation, University of Pennsylvania. (Published by Research Center for Peace and Unification of Korean.)

Huffman, M. 1987. "Measures of phonation type in Hmong." *Journal of the Acoustical Society of America*, 81,495-504.

Huh, W. 1980. *Kuk-Eo-Um-Un-Hak (Korean Phonology).* (In Korean) Sam-Mun-Hwa-Sa, Seoul, Korea.

Jun, S.-A. 1993. *The phonetics and phonology of Korean prosody.* Ph.D. dissertation, Ohio State University.

________. 1995. "Asymmetrical prosodic effects on the laryngeal gesture in Korean." In B. Connell & A. Arvaniti (eds.), *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*, pp. 235-253. Cambridge, UK: Cambridge University Press.

________. 1996a. *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure.* Garland Press. New York.

________. 1996b. "Influence of microprosody on macroprosody: a case of phrase initial strengthening." *UCLA Working Papers in Phonetics*, 92, 97-116

________. 1998. "The Accentual Phrase in the Korean prosodic hierarchy." *Phonology*, 15(2), 189-226.

________. 1999. "Microprosody of English, French, and Korean: phonologization of microprosody in Korean." Unpublished manuscript. UCLA.

Jun, S.-A., M. Beckman & H. Lee. 1998. "Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents." *UCLA Working Papers in Phonetics,* 96, 43-68

Kagaya, R. 1974. "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives." *Journal of Phonetics*, 2, 161-80.

Keating, P., T. Cho, C. Fougeron & C. Hsu (1999, to appear) "Domain-initial strengthening in four languages." *UCLA Working Papers in Phonetics*, 98, 139-151. To appear in *Laboratory Phonology 6.* Cambridge University Press.

Kim, W.-J. 1963. *Kuk-eo Mo-eum-che-kye-eui Shin-ko-chal (Re-analysis of Korean Vowel System).* Chin-tan-hak-po 24. Il-cho-kak, Seoul Korea.

Kim, H.-K. 1980. "An acoustic study of the Cheju dialect in Korean." *Eon-Eo (Language)*, 5(1), 25-61. (The Journal of Linguistics Society of Korea, Seoul Korea.)

Kim, C.-W. 1965. "On the autonomy of the tensity feature in stop classification." *Word*, 21, 339-359.

Kim, C.-W. 1970. "A theory of aspiration." *Phonetica*, 21, 107-116.

Kim, M.-R. Cho. 1994. *Acoustic Characteristics of Korean Stops and Perception of English Stop Consonants.* Ph.D. dissertation. University of Wisconsin-Madison.

Kim, Y. 1995. *Acoustic Characteristics of Korean Coronal.* M.A. thesis. University of Texas at Arlington

Klatt, D. 1975. "Voice onset time, frication, and aspiration in word-initial consonant clusters." *Journal of Speech and Hearing Research*, 18, 686-706

Klatt, D. & L. Klatt. 1990. "Analysis, synthesis, and perception of voice quality variations among female and male talkers." *Journal of the Acoustical Society of America* , 87, 820-857

Ladefoged, P. 1964. *A phonetic study of West African languages.* Cambridge: Cambridge University Press.

___________. 1983. "The linguistic use of different phonation types." In D. Bless & J. Abbs (eds.), *Vocal fold physiology: Contemporary research and clinical issues*, pp. 351-360. San Diego: College Hill Press.

Ladefoged, P., I. Maddieson & M. Jackson. 1987. "Investigating phonation types in different languages." In O. Fujimura (ed.), *Vocal Physiology: Voice Production, Mechanisms and Functions,* pp. 297-317. New York: Raven.

Lehiste, I. & G.E. Peterson. 1961. "Some basic considerations in the analysis of intonation." *Journal of the Acoustical Society of America* , 33, 419-425.

Lieberman, P. 1963. "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges." *Journal of the Acoustical Society of America*, 35, 344-353.

Lisker, L. & A. S. Abramson. 1964. "Cross-language study of voicing in initial stops: Acoustical measurements." *Word*, 20, 384-422.

Park, H.S. 1999. "The phonetic nature of the phonological contrast between the lenis and fortis fricatives in Korean." In *Proceedings of the 14th International Congress of Phonetic Science (ICPhS 99)*, vol. 1, 424-427.

Pierrehumbert, J. & D. Talkin. 1992. "Lenition of /h/ and glottal stop." In G. Docherty & D. R. Ladd (eds.) *Papers in Laboratory Phonology II: gesture, segment, prosody,* pp. 90-117. Cambridge, UK: Cambridge University Press.

Silva, D. J. 1992. *The Phonetics and Phonology of Stop Lenition in Korean.* Ph.D. dissertation. Cornell University

Stevens, K. N. (1999) *Acoustic Phonetics*. Cambridge: MIT Press

Stevens, K. N., S. J. Keyser & H. Kawasaki. 1986. "Toward a phonetic and phonological theory of redundant features." In J.S. Perkell & D.H. Klatt (eds.), *Invariance and Variability in Speech Processes*, pp. 426-463. New Jersey: Lawrence Erlbaum Associates.

Yang, B. 1996. "A comparative study of American English and Korean vowels produced by male and female speakers." *Journal of Phonetics*, 24, 245-261.

Yoon, K. 1998. *An acoustic analysis of Korean aspirated and tense alveolar fricatives using Multi-Speech.* MA thesis. University of Kansas.

The vowels of Cheju

Taehong Cho[1], Sun-Ah Jun[1], Seung-Chul Jung[2] and Peter Ladefoged[1]

This paper investigates acoustic qualities of monophthong vowels of Cheju spoken by rural and urban speakers mostly in their 60 - 70s. We examined the first and second formants (F1 and F2) of vowels with reference to how vowels are spaced in the bark scaled vowel space. One of the obvious findings is that quite a few Cheju speakers do make clear distinctions between **e** and **æ**, and **o** and **ʌ**. When the urban and rural data were considered separately, and were compared with those about 50 years ago, a few generalizations appear to fall out: (1) the mergers of **e/æ** and **o/ʌ** are in progress in Cheju; (2) the mergers spread from the urban to the rural area; (3) the **e/æ** merger started earlier than the **o/ʌ** merger; and (4) at least for the urban Cheju speakers, the merger is best accounted for by the *merger-by-transfer* model, a unidirectional change in which one phonemic category becomes another (cf. Labov, 1994). Further, when our data are compared with other acoustic data available (including studies of the standard Korean in the 1960s and 1990s), it suggests that the directionality of the diachronic sound change is guided by both auditorily and articulatorily based principles such as contrast maximization and effort minimization principles.

1. Introduction

Cheju is the form of Korean spoken on Cheju Island, one of the eight provinces of the country. Although Cheju is not mutually intelligible with Standard Korean, it has been considered to be one of the six dialects in South Korea. In this paper we will take no stand on whether Cheju is a dialect of Korean or a separate language. Our purpose here is simply to compare the vowels of Cheju with those of what we will refer to as Korean, meaning Standard Korean as spoken in Seoul. It is generally agreed that Cheju is currently in the process of changing from a 9 vowel system to a 7 vowel system, as shown in Figures 1a and b. The 9 vowel system is found in the dialect of the older generation of Cheju speakers, and the 7 vowel system is found in the dialect of the younger generation (Kim 1963, Hyeon 1964a, 1985, Jung 1995).

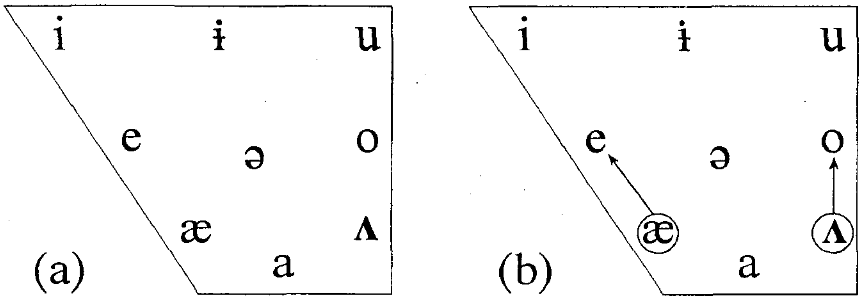


Figure 1. Two vowel systems in Cheju. The 9 vowel system in (a) is found in the older generation and the 7 vowel system in (b) is found in the younger generation (Jung 1995)

---

[1] Department of Linguistics, UCLA, [2] Department of Korean, Inha University, Korea.

Historically, in Middle Korean, **ʌ** (known as *alae a* 'below a', named after its location in the old hangul paradigm) contrasted with **o** in rounding and in vowel height (Jung 1995), but it disappeared in Standard Korean in the late 1800s (Lee 1972, Huh 1980). (Note that the symbol **ʌ** is used to refer to the unrounded counterpart of **o** simply for the sake of simplicity, but its exact phonetic values in the vowel space may not correspond to those of **ʌ** in the IPA vowel chart.)

In Cheju, however, it has been claimed that **ʌ** has not been completely lost, and still contrasts with **o** in the speech of the older generation of Cheju speakers, as exemplified by the minimal pair ***tol*** 'rock' and ***tʌl*** 'moon' (Jung 1995). The presence of **ʌ** in the speech of these speakers would allow us to have some insight into the nature of this sound in middle Korean.

The other vowel that exists in the 9 vowel system but not in the 7 vowel system in Cheju is the mid-low or low front vowel **æ**, which is sometimes transcribed as **ɛ**. In this paper, we will use **æ** for this mid-low front vowel, noting, however, that we are not suggesting that this vowel is the equivalent of the English vowel 'a' as in 'apple' which is often transcribed with the symbol **æ**. The merger of **e** and **æ** is in progress in Cheju Korean (Jung 1995), and this is believed to be also true in Seoul Korean (Hong 1991, Kang 1996).

The phonetic structures of Cheju vowels have not been intensively studied. The only acoustic study that we have found in the literature is by Kim (1980). Kim reported F1 and F2 values of vowels in various phonetic contexts from one male speaker, 80 years of age at the time of recording in 1969. The speaker had not been exposed very much to (Seoul) Korean. The average values of F1 and F2 for monophthong vowels as reported by Kim are plotted in Figure 2, where the vertical axis shows F1 values and the horizontal axis shows F2 values in Hz. As can be seen in the figure, these vowels are quite symmetrically distributed in the F1-F2 plane, and there are clear differences between **e** and **æ** and between **o** and **ʌ**.

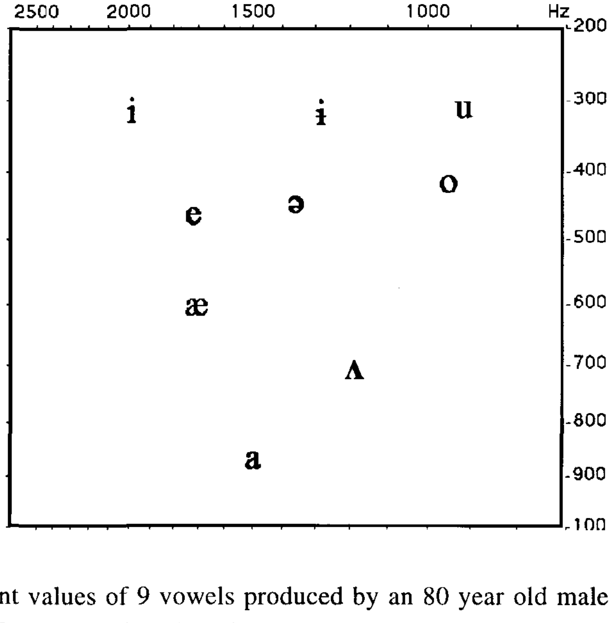


Figure 2. The formant values of 9 vowels produced by an 80 year old male speaker as of 1969. The formant values of each vowel are based on the mean values in H.K. Kim (1980).

2. Procedure

This paper is based on recordings of eight male speakers of Cheju. Three speakers (S1-S3), aged 55, 61 and 68, were recorded in Cheju city, and the other five speakers (S4-S8) were recorded in the more mountainous area of southern Cheju (Namcheju-kun), in two villages: Shinrye-ri and Uikwi-ri. The five men in this region were between 66 and 74 years old. All the speakers understood Korean, were literate, and above average in their socioeconomic status. Each of the eight speakers was recorded using a close-talking, noise-canceling Shure microphone and a Sony DAT recorder. Further details of the recording procedures are given in Cho, Jun and Ladefoged (to appear, this volume), a paper discussing the consonants of Cheju based on data recorded at the same time.

The three sets of words used for the analysis of vowel quality in this study are shown in Table 1a-c, in a broad phonetic transcription, using the diacritic '*' to mark fortis obstruents. The first set has vowels after a velar stop, the second after a denti-alveolar stop. The third set contains additional words suitable for the **e/æ** comparison. The third column in Table 1a and 1b indicates which speakers produced which words. Two different words were used for the vowels **i, e, æ, o, u** because the speakers from the rural area were not comfortable pronouncing the words that the speakers in the town had found satisfactory.

Recordings were digitized into a computer at a sampling rate of 10 kHz. The first two formant frequencies were measured using the Kay Elemetrics MultiSpeech system. For each token, a steady state portion was found near the mid point of the vowel, and superimposed LPC and FFT spectra were calculated with 30 ms and 25.6 ms frames, respectively. The formant values were usually determined from the LPC spectra (with 12 or 14 coefficients), using the FFT spectra (and sometimes formant history tracking) as supplementary checks.

Table 1. Words recorded for the analysis of vowels in a word initial syllable.

(a) After velar stop

Vowel	Words in isolation	Speakers	Gloss
i	**k*iəmtʃə** **k*itʃimalla**	S1-S3 S4-S8	'to fit' 'don't fit it'
e	**k*eəmtʃə** **k*etʃimalla**	S1-S3 S4-S8	'to break (glass)' 'don't break it'
æ	**kæəmsə** **kætʃimalla**	S1-S3 S4-S8	'to clear (sky)' 'don't fold (a blanket)'
ɨ	**k*ɨla**	all	to turn off (light)
ə	**kəlla(ke)**	all	to hang
a	**kalk*əjə**	all	to plow
ʌ	**kʌlla**	all	to grind (knife)
o	**k*oamtʃə** **k*otʃimala**	S1-S3 S4-S8	to twist (knot) not to twist (knot)
u	**k*uəmtʃə** **k*utʃi anjəttʃə**	S1-S3 S4-S8	to dream (I) did not dream

Table 1. Cont'd

(b) After denti-alveolar stop

Vowel	Words in isolation	Speakers	Gloss
i	**t*i**	all	band (waist)
e	**teəttʃə** **t*etʃimalla**	S1-S3 S4-S8	to get burned not to take off
æ	**tæəttʃə** **tætʃimalla**	S1-S3 S4-S8	to touch do not touch
ɨ	**tɨləttʃə**	all	to encumber money
ə	**tʰələttʃə** **tʰələpulla**	S1-S2 S3-S8	to dust
a	**tʰattʃə**	all	to burn
ʌ	**tʌlattʃə**	all	to sew on a button
o	**tolattʃə**	all	to go around
u	**tuəstʃə** **tutʃimalla**	S1-S3 S4-S8	to place do not place

(c) Front mid/low vowel pairs in monosyllabic words

Vowel	Words in isolation	Gloss
e	**pe**	hemp cloth
æ	**pæ**	thick rope
e	**se**	tongue
æ	**sæ**	material (roof)
e	**kʰe(ta)**	to chap
æ	**kʰæ(ta)**	It is kʰæ (game)!

3. Results

Figure 3 shows the vowel distribution across speakers and consonants (**t** and **k**) in the plane of the first two formants of the vowels, plotted on a Bark scale, using the UCLA Plot Formants program. This program was also used to draw ellipses with a radius of two standard deviations along the axes of the first two principal components of the distribution.

This plot appears to show that there is considerable confusion among these vowels. The vowel space between **e** and **æ** and between **o** and **ʌ** is substantially overlapped and there is a large variation for **u** in the F2 dimension. In what follows, we will examine the distribution of vowel space and the variability of front/back mid vowels and **u**, considering factors such as speaker, region, and segmental context. When we do so, it will be apparent that much of the confusion can be resolved.

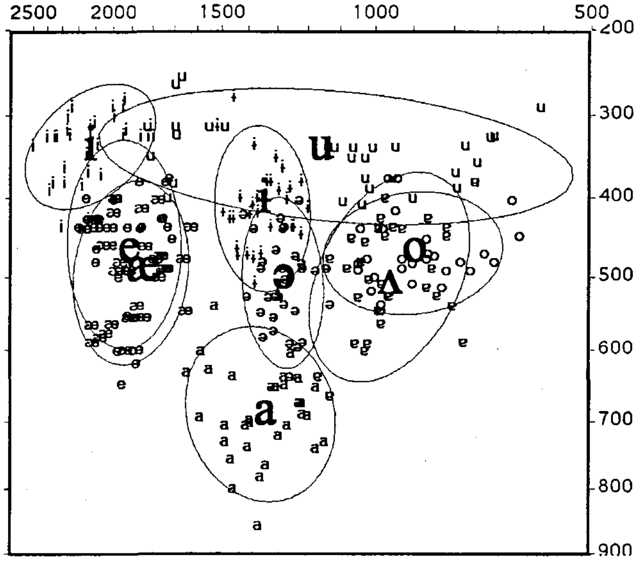


Figure 3. Formant plot of Cheju vowels across speakers and consonantal contexts.

3.1 *Front mid vowels: e and æ distinction*

Part of the spread in the data for the **e** and **æ** vowels can be ascribed to a problem with the word list. Owing to confusions in the mapping between Cheju sound and Korean orthography, the words **k*eəmtʃə** (recorded by the urban speakers) and **k*etʃimalla** (recorded by the rural speakers) were written in a way that indicated a pronunciation that might not correspond to their pronunciation when not reading. . As we do not know how our speakers were affected by the orthography, we leave these words out of our detailed analyses, focusing on the **e** and **æ** in /t_/ context.

In order to see whether there were differences between the urban and rural speakers, the data were submitted to a two way ANOVA with factors Vowel, and Region. With the first formant (F1) as the dependent variable, there were significant effects for both factors (F(1, 28) = 12,085, p < .001 for the **e/æ** difference; F(1, 28) = 4.253, p < .05 for the Urban/Rural difference. For the second formant (F2), none of the factors had a significant effect. Since the Region significantly influenced the vowel production, the vowels are plotted separated by Region in Figures 4a and 4b, for the urban speakers and the rural speakers, respectively.

Pairwise *posthoc* comparisons show that urban speakers did not distinguish between **e** and **æ** in both F1 and F2, **as seen in** Figure 4a, **while rural speakers did** make a significant distinction between these two vowels in F1 (p <.0001) as seen in Figure 4b— **æ** has a higher average F1 value than **e**, i.e. **æ** is a lower vowel than **e**.

When we examine the formant values of **e** and **æ** in the monosyllabic words (Table 1c, Figure 6), it is clear that the rural speakers fully distinguish these two mid front vowels, but the urban speakers, even though they were in their 60's, do not. For the rural speakers, F1 is significantly greater for **æ** than for **e**, and has a similar value to that in the **æ** produced by the 80 year old speaker in 1969, as shown in Figure 2. However, our data suggest that even the rural speakers of Cheju do not make a distinction between e and **æ** as clearly as the distinction made 50 years ago.

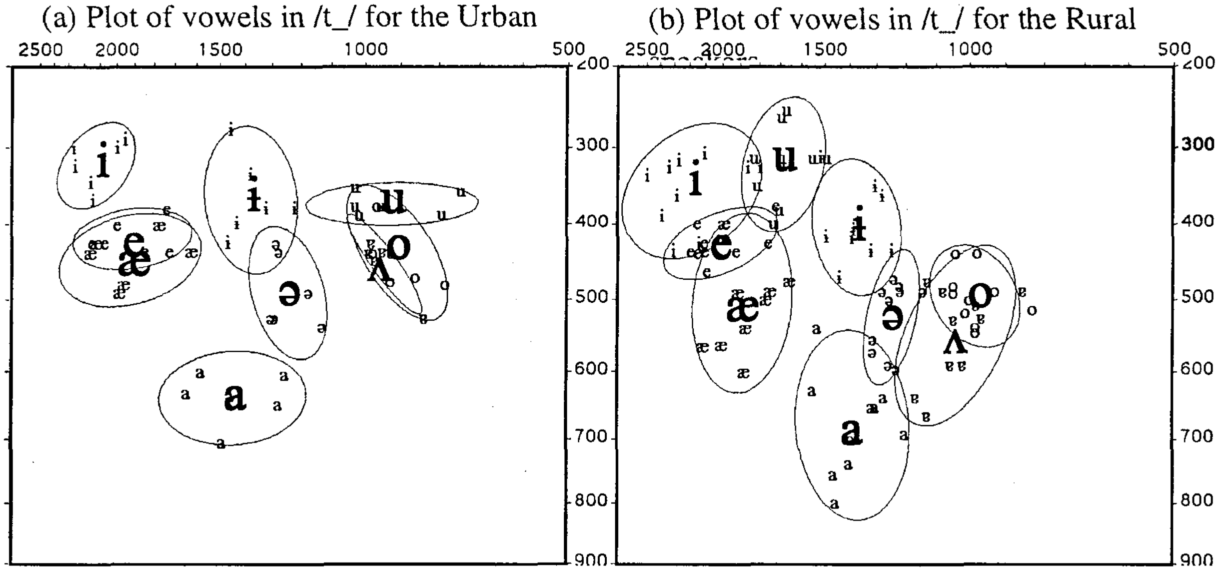
3.2 *Back mid vowels: o and ʌ distinction*

Figures 4 and 5 show that there is a difference between **o** and **ʌ** for both urban and rural speakers. The urban speakers make the **o/ʌ** distinction by employing an F2 (backness) difference. As can be seen in Figures 4a and 5a, they tend to make **o** less front than **ʌ**. This difference was significant only in the /k_/ context (t=3.107, p = 0.01), but not in the /t_/ context. On the other hand, the rural speakers make the **o/ʌ** distinction by employing an F1 (height) difference. As shown in Figures 4b and 5b, **ʌ** is generally lower than **o** and the differences are significant in the /t_/ context for all six speakers. In the /k_/ context, all but one speaker (S6) showed a significant difference in F1 (t=2.522, p = .0244). It seems that this distinction is in the process of being lost among urban speakers, even those in their 60s, but is better preserved among rural speakers of a similar age, although one of the rural speakers did not show any difference at all—again indicating a move towards a merger, even by the rural speakers. However, the distinction is not as clear as it was a half a century ago (see Figure 2 for the result reported in Kim, 1980). The rural speakers seem to lower the rounded mid back vowel, **o**, and raise and retract the unrounded vowel, **ʌ**, compared to the earlier data, thus weakening the contrast between these two vowels in both the height and the front/back dimension.

3.3 *The variability of **u***

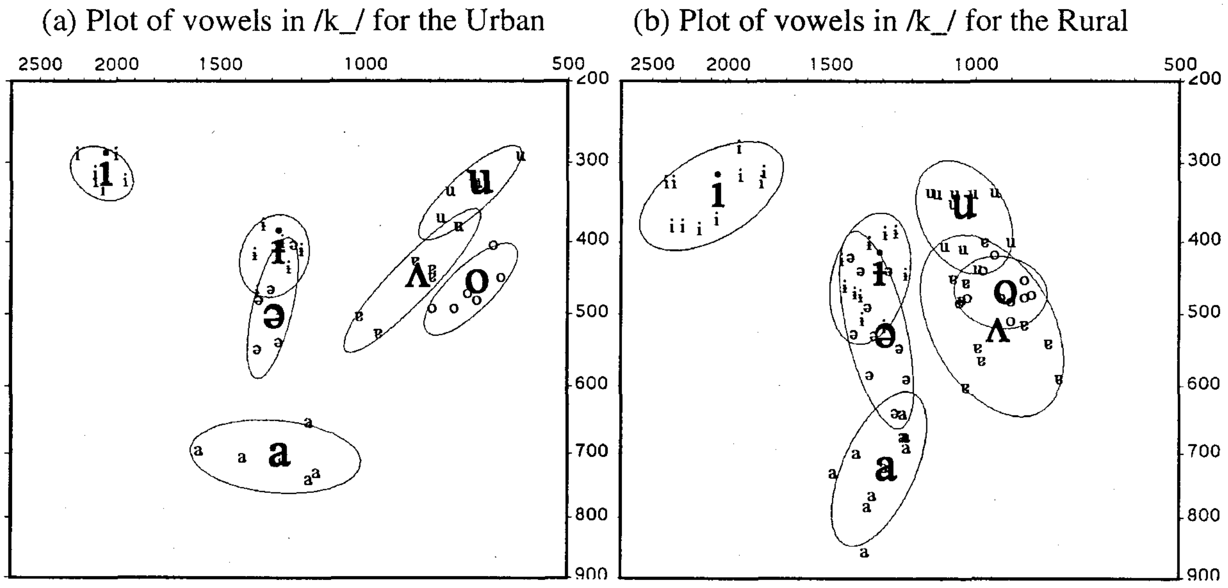
As noticed earlier in Figure 3, there is great variation in the distribution of tokens for **u**. A comparison of the four graphs in Figures 4 and 5 shows that rural speakers have more variation than urban speakers. In both speaker groups, **u** is fronter (i.e., higher F2 values) in t_ context than k_ context (F (1, 28)=17.532, p < 0.001), but the degree is much stronger in the rural speakers' data. In addition, the frontness of **u** in each context is stronger in rural speakers' data than in urban speakers' data. Much of this variation is due to the coarticulatory effect of the adjacent segment. Figure 7(a) shows the effect of the preceding consonant on the formant value of **u** in two different following contexts (before **ə** in the left panel, and before **tʃi** in the right panel). The arrows indicate fronting of **u** when it is in t_ context compared to when it is in k_ context. In other words, speakers make **u** more fronted after the front consonant **t**, and such fronting is not made when the preceding consonant is **k**.

In addition to the fronting due to the preceding context, the data also show that F2 values of **u** are influenced by the following context. The urban speakers produced **u** in words where the following context is **ə** as in **tuəttʃə** ('to place') and **k*uəmtʃə** ('to dream') while the rural speakers produced **u** followed by **tʃi** as in **tutʃimalla** ('do not place') and **k*utʃi anjəttʃə** ('(I) did not dream'). This different context significantly influenced the production of **u.** Figure 7b illustrates that **u** is more fronted (with a greater F2) when followed by **tʃi** than when followed by **ə,** and this is statistically significant (F (1, 28) = 31.124, p < .0001 for F2). This suggests a significant effect of anticipatory coarticulation between **u** and the following segment, and explains why rural speakers show more context effect in the F2 dimension than urban speakers. In sum, the variation of **u** is due to the coarticulatory effect of adjacent segments, and the difference between urban and rural speakers is not related to a regional feature, but is caused by different word sets. Our data show that in general other back vowels are also fronted by the following **tʃi** context, but the effect is especially large in **u**.


(a) Plot of vowels in /t_/ for the Urban

(b) Plot of vowels in /t_/ for the Rural

Figures 4. Formant plots of vowels in /t_/ context for (a) the urban speakers and (b) the rural speakers.

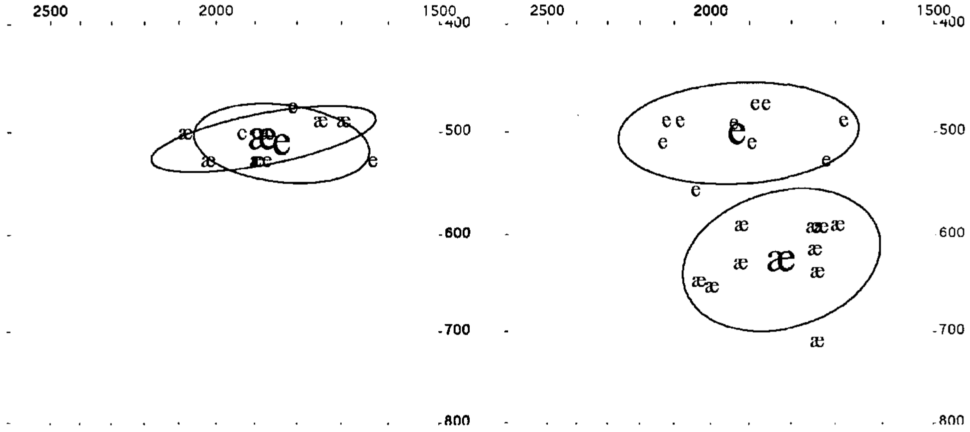

(a) Plot of vowels in /k_/ for the Urban

(b) Plot of vowels in /k_/ for the Rural

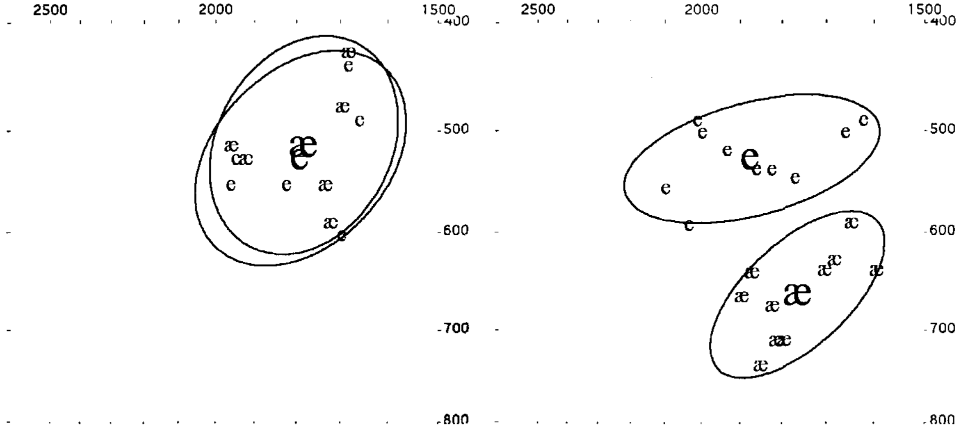
Figures 5. Formant plots of vowels in /k_/ context for (a) the urban speakers and (b) the rural speakers.

Urban speakers (S1-S3) Rural speakers (S4-S8)

(a) in /p_/ context



(b) in /s_/ context



(c) in /k_/ context

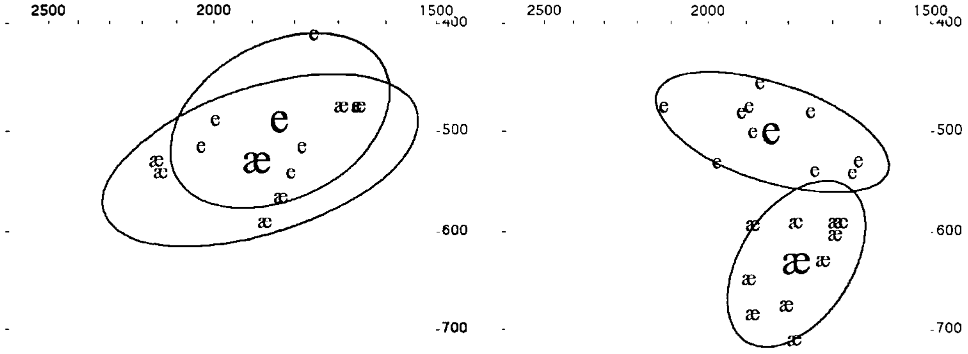
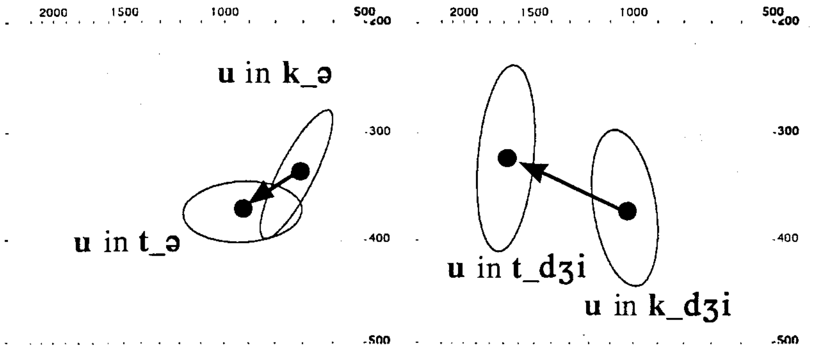


Figure 6. e/æ difference in monosyllabic words.

(a) Effect of the preceding C on **u**



(b) Effect of the following V on **u**

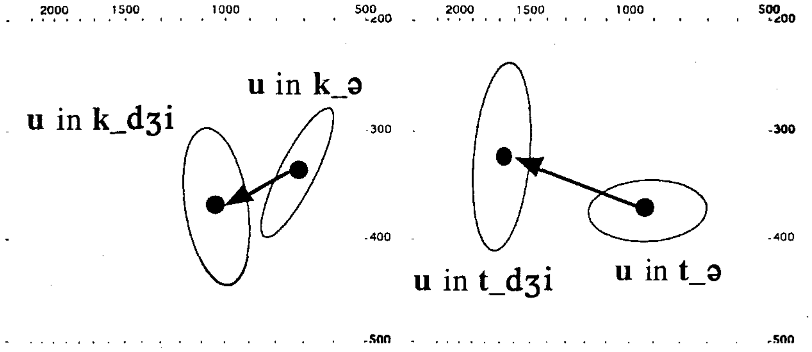


Figure 7. Variation of **u** due to the adjacent context, showing coarticulatory effect.

4. Summary and discussion

In this paper, we examined the vowel space of Cheju with a primary focus on **e/æ** and **o/ʌ** differences and the variation of **u**. There is strong evidence that the rural speakers have an **e/æ** distinction in their clear speech, whereas the urban speakers, even in their 60s, seem to have lost it entirely. Second, both groups make the **o/ʌ** distinction to some extent. The rural speakers made a consistent distinction in F1, and the urban speakers made a distinction in F2 for half of the data recorded. Third, we observed a great deal of variability for **u**, which is primarily due to coarticulatory effects.

Our data suggest that the mergers of **e/æ** and **o/ʌ** are in progress in Cheju. Such transitory status can be seen in that **e/æ** or **o/ʌ** distinctions were made by the rural speakers most of the time, but rarely by the urban speakers of a similar age. This is presumably because the rural speakers are more conservative, and, as a result, tend to preserve the contrasts. This is reminiscent of a case reported by Trudgill (1974) where a pattern of merger is detected around an urban center and spreads to the rural area. Labov (1994, p. 313) explains that "dialect geography is one of the most powerful tools available for interpreting the past by means of data drawn from the present." In other words, by the investigation of the forms spoken in an isolated rural area, we could possibly infer what the old forms would be like. In fact, some of the **e/æ** distinctions made by our rural speakers are very similar to those made by the old speaker in Kim's 1969 study. In addition, the direct comparison between the present day urban speakers's data and the data in Kim's 1969 study shows that the merger is in progress.

Our data also suggest that the **e/æ** merger started earlier than the **o/ʌ** merger in Cheju. Urban speakers in our study completely lost the **e/æ** distinction but still partially preserve the **o/ʌ** distinction. A supporting example is that there is a complete overlapping for the **e/æ** pair while a clear **o/ʌ** distinction is made by the same urban speakers, as seen in Figure 5a. This supports the informal observation shared by the researchers of Cheju; it has been believed that in general the loss of **o/ʌ** distinction is found among speakers in their mid 50s and younger while the loss of **e/æ** distinction is found among speakers in their 60s and younger.

It is interesting to note that the order of merger in Cheju between these two pairs is the opposite of that in Standard Korean. In mainland Korean, the **o/ʌ** merger occurred in late 1700s (e.g., Lee 1972; Huh 1980), and the **e/æ** merger has been in progress in the past several decades (cf. Hong 1991). Jung (1995) posits that the loss of contrast in non-high front vowels (the **e/æ** merger) that is in progress in Standard Korean is driven by pressure from the system. With only two back non-low vowels (**u** and **o**, after the loss of **ʌ**) and three front non-low vowels (i, e, and æ), the system is loosing the **e/æ** distinction to make the vowel system more symmetric. Along the same line, we could posit that in Cheju, the **e/æ** merger started first and the **o/ʌ** merger follows to make the vowel system symmetric. It is believed that Cheju lost the **e/æ** distinction earlier than mainland Korean due to the confusion in the mapping between sound and the standard orthography. That is, as mentioned earlier, the vowel **æ** symbol in Korean is sometimes mapped to Cheju **æ** and sometimes to Cheju **e**.

4.1. *Mechanism of mergers*

Now a question arises as to by what route two phonemes become one. Labov (1994, pp. 321-323) suggested three models to account for different kinds of mergers that have been observed in sound changes. The first type of merger processes is *merger by approximation*—i.e., phonetic targets of two distinctive sounds are gradually approximated until they become non-distinctive. In this type we expect the two sounds are merged into an intermediate phonetic form. The second type of merger occurs by means of *merger by transfer*. This process is a unidirectional change in which one phonemic category becomes another. In general, merger by transfer takes place when one form becomes a social prestige, being dominantly used in an established standard language. The last process is *merger by expansion* —i.e., the new phonetic space is expanded to range over phonetic spaces of two phonemes that are merged. In other words, the new phonetic space falls neither on an intermediate range (as in merger by approximation) nor on the phonetic space that belongs to either one of the phonemes that are merged (as in merger by transfer).

It is not entirely clear from the data available to us what kind of merger process is taking place in Cheju. On the one hand, the new phonetic space for the **e/æ** for urban speakers as seen in Figure 4a appears to be best accounted for by the merger-by-transfer model which predicts a unidirectional change. In other words, the phonetic space of **æ** seems to re-map to that of **e**. This can be more clearly seen in monosyllabic words in Figure 6. On the other hand, the **o/ʌ** merger show inconsistent patterns between urban and rural speakers. The merger in Figure 4a (as produced by urban speakers) fits into the merger-by-transfer model—i.e., the tokens are clustered around the phonetic space for **o**. By contrast, the merger shown in Figure 4b (rural speakers) indicate that, although there is no complete merger, merger-by-approximation appears to be in progress,—i.e., the phonetic space for **o** is relatively lowered. Though more systematic data are required to examine the exact mechanism of the merger that is in progress in Cheju, what emerges from the data available is that at least the merger for the urban speakers is best accounted for by the merger-by-transfer model.

In the next section we will examine vowel spaces in Seoul Korean and Cheju synchronically and diachronically, and discuss the sound change related to the principles of the maximal perceptual contrast and minimal articulation.

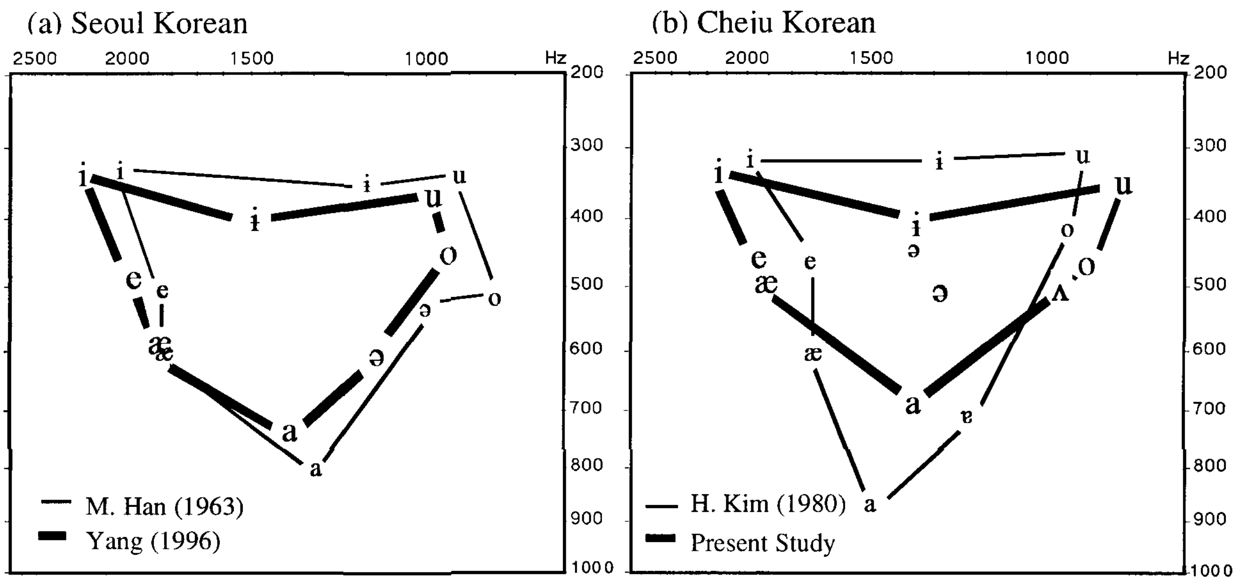
4.2 *Dispersion theory and sound change*

So far, we have compared the two pairs of mid vowels in Cheju at two time frames separated by about 50 years. In this section, we will examine the vowel system of Seoul Korean roughly at the same time frame as Cheju data, and compare Seoul Korean data with Cheju data. As the vowel system of Seoul Korean at the two time frames, we will use data from Han (1963) and Yang (1996).

Middle Korean had **ʌ**, as in present day Cheju. In late 1700s Seoul Korean lost **ʌ** (Huh 1980) and now has an 8 vowel system, if they have **e/æ** distinction, or a 7 vowel system, if not (Kang 1997). Figure 8 illustrates vowel systems of (a) Seoul Korean and (b) Cheju. Figure 8a compares the present day vowel system of Seoul Korean (Yang 1996) with that in 1960s (Han 1963), and Figure 8b compares the present Cheju data with Kim's (1980) Cheju data. Note that the Cheju data from the present study are pooled from both urban and rural speakers except for [u] which is based on urban speakers' data only, due to the extreme fronting of rural speakers' data influenced by the following segments (see section 3.3 and Figure 4b). As a result, both **e/æ** and **o/ʌ** distinctions are merged in the Present Study plot in Figure 8b.

In Figure 8a, there seems to be a clear distinction between **e/æ** in Yang's (1996) data (the mean difference is 99 Hz in F2). However, the large standard deviations (75 Hz and 105 Hz for **e** and **æ**, respectively) suggests that the difference is not statistically significant. In fact, Lee, Kim, and Hong (1997) showed in their study of the vowel qualities of media language that there is no substantial difference between **e** and **æ** in the present day Seoul Korean. They also compared their own media data with Han's and Yang's data and showed that the vowel space is smaller for the normally spoken media data, due to the reduction and the contextual influence. In sum, Seoul Korean had 8 vowels in 1960s (i.e., Han's data), and 7 vowels in 1990s (i.e., Yang's data), while Cheju had 9 vowels in Kim's study and 7 vowels (or 8 for rural speakers) in the present study.

Lindblom's revised dispersion theory (Lindlbom 1986; Lindblom and Maddieson, 1988) predicts that, in a given system, contrastive vowels are spaced with a sufficient contrast. This principle of sufficient contrast can be observed in the location of the mid vowel **ə** and the low vowel **a** in Seoul Korean and Cheju. First, the vowel **ə** in Cheju is mid-central, whereas it is further back and lowered in Seoul Korean, taking the space of open-mid back vowels in the IPA chart (1999). This mid-centralness of the vowel **ə** is preserved in both the earlier and present day vowel systems of Cheju. We believe that this is due to the presence of contrast between **ə** against **o** or **ʌ** in Cheju. In Seoul Korean, on the other hand, it would not be necessary to keep **ə** in the central position. It seems that **ə** is instead moved to the open-mid back position, so that it could enhance the front/back contrast among vowels and make the system more symmetrical. That is, the comparison between Seoul Korean and Cheju allows us to infer that **ə** has been shifted from the central vowel space to the mid-back area in Seoul Korean, in order to achieve a sufficient contrast among vowels, a phenomenon in which the sound change is driven by the principle of contrast maximization.


(a) Seoul Korean

(b) Cheju Korean

Figure 8. Comparisons between (a) Seoul Korean and (b) Cheju vowel systems at two time frames, current vs. 1960s. The vowel chart for Han (1963) was drawn based on her Informant 1's average values and Yang's (1996) data were based on 30 male Seoul speakers. The Kim (1980) data were collected in 1969.

Second, as shown in Figure 8b, within Cheju, **a** in the present study is much higher than that in Kim's study (by 139 Hz in F1). As discussed earlier, Cheju in 1960s had a 9 vowel system and had four vowels contrastive in vowel height, or F1 (e.g., **i-e-æ-a** or **u-o-ʌ-a**). This 9 vowel system seems to use a larger vowel space especially in the vowel height dimension, compared to the present Cheju vowel system in which there is a three way contrast in vowel height (i.e., **i-e/æ-a** or **u-o/ʌ-a**) due to the merger of **e/æ** and **o/ʌ**. Reducing the vowel space in the height dimension in the current Cheju system supports the sufficient contrast principle instead of the maximal contrast principle as originally proposed by Liljiencrants and Lindblom (1972). The space reduction in the current Cheju system also supports the principle of articulatory effort maximization (e.g., Lindblom 1986, 1990; Flemming, 1995). These two principles together predict that speakers minimize the articulatory effort to an extent that allows a sufficient contrast. Our data follow these principles. In the 9 vowel system, speakers make more effort to make an extreme **a**, or to open their mouth wider, in order to achieve a sufficient contrast between four vowels that are distinctive in vowel height, whereas in the 7 vowel system, speakers do not need to make such articulatory effort since less vowel space in height is required for sufficient contrast between three vowels. These auditorily and perceptually based principles together with the distribution and the inventories of the vowel system seem to guide the directionality of the diachronic sound change.

Acknowledgements

We thank Dr. Doheung Ko for providing us with the paper by Kim (1980), and Dr. Kee-Ho Kim for the paper by Lee, Kim, and Hong (1997). We also thank Dr. Chang-Myong Oh who helped us record in Cheju city, and Dr. Won-Bo Kim who helped us record in the rural areas and provided information of Cheju morphology. We are also grateful to all the Cheju speakers who recorded data for us.

References

Cho, T., S. Jun, & P. Ladefoged (this volume, to appear) "An acoustic and aerodynamic study of Cheju Consonants." UCLA Working Papers in Phonetics, this volume. (To appear in *The Journal of the Korean Association of Speech Sciences.*)

Jung, S.-C. (1995) *Cheju-do Pang-eon-euy Thong-shi-eum-un-lon (A Diachronic Approach to the Phonological Processes of Cheju Dialect)* Kuk-Eo-Hak-Hoi, Seoul, Korea.

Flemming, E. (1995) *Auditory Representations in Phonology.* PhD. dissertation, UCLA.

Han, M. (1963) *Acoustic Phonetics of Korean.* UCLA.

Hong, Y. (1991) *A sociolinguistic study of Seoul Korean.* PhD dissertation, University of Pennsylvania. (Published by Research Center for Peace and Unification of Korean.)

Huh, W. (1980) *Kuk-Eo-Um-Un-Hak (Korean Phonology).* (In Korean) Sam-Mun-Hwa-Sa, Seoul, Korea.

Hyeon, Pyeong-hyo (1964a) *Chejudo Pang-eon-ui Tan-mo-um Seol-jung (Establishing Monophthongs of Cheju dialect)* (In Korean). Hankuk Eoneo Munhak (Korea Language and Literature) 2, Korea.

Hyeon, Pyeong-hyo (1964b) *Cheju Pang-eon ay-um-e-tae-ha-yo* (On Cheju dialect /æ/). Kukmun Hakpo (Research on Korean Literature) 2, Cheju University.

Hyeon, Pyeong-hyo (1985) *Cheju-do Pang-eon Yeon-ku Non-ko-pyun* (A Study of Cheju Dialect), (In Korean), Eewu Publishing Co., Korea.

IPA chart (1999) *Handbook of the International Phonetic Association.* Cambridge University Press, Cambridge, UK.

Kang, Soon-Kyung (1996) "Acoustic analysis of South and North Korean vowels." *Language Research,* 32 (1), 1-18.

Kim, Wan-Jin (1963) *Kuk-eo Mo-eum-che-kye-eui Shin-ko-chal (Re-analysis of Korean Vowel System).* Chin-tan-hak-po 24. Il-cho-kak, Seoul Korea.

Kim, Han-Kon. (1980) "An acoustic study of the Cheju dialect in Korean." *Eon-Eo (Language) 5(1):* 25-61. (The Journal of Linguistics Society of Korea, Seoul Korea.)

Labov, W. (1994) *Principles of Linguistic Change.* Blackwell Publishers, Cambridge, MA.

Lee, K.-M. (1972) *Kuk-eo-sa-ke-seol (Introduction to Korean History)* (In Korean), Min-Jung Seo-Kwan Publisher, Korea.

Lee, Y., Kim, K., and Hong, J. (1997) "A compariative stuydy on vowel qualities of North/South Korean media language." (In Korean.) Seong-Kok-Hak-Sul-Mun-Hwa-Chae-Tan.

Liljencrants, J. and B. Lindblom. (1972) "Numerical simulation of vowel quality systems: The role of perceptual contrast." *Language, 48,* 839-862.

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

Lindblom, B. (1990) "Phonetic content in phonology." *PERILUS11.*

Lindblom, B. and I. Maddieson. (1988) "Phonetic universals in consonant systems." In L. Hyman and C. Li (eds.), *Language, Speech and Mind,* Routledge, pp. 62-78.

Trudgill, P. (1974) "Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography." *Language in Society, 3,* 215-46.

Yang, B. (1996) "A comparative study of American English and Korean vowels produced by male and female speakers." *Journal of Phonetics, 24*, 245-261.