PHONETICS OF ENDANGERED LANGUAGES

D. H. Whalen
Speech-Language-Hearing Program
Graduate Center of the City University of New York
365 Fifth Avenue
New York, New York 10016
and
Haskins Laboratories
300 George Street, Suite 900
New Haven, Connecticut 06511
and
Endangered Language Fund

Christian T. DiCanio
Haskins Laboratories
300 George Street, Suite 900
New Haven, Connecticut 06511

Patricia A. Shaw
First Nations Languages Program
University of British Columbia
1866 Main Mall
Vancouver, British Columbia V6T 1Z1, Canada

The world is filled with an astounding array of languages, 6,909, by the count of the *Ethnologue* (Lewis, 2009). Most of these use an acoustic signal as the main element in signal transmission, though vision affects speech even for typically hearing individuals (e.g., McGurk and MacDonald, 1976); sign languages (126 are listed in Lewis, 2009) use the visual channel almost exclusively. The acoustic signal for speech is powered mostly by the larynx and shaped by the vocal tract. Because human populations have essentially the same anatomy, there is a great deal of similarity in the sounds that languages use. However, there is an impressive range of variability as well. The largest survey of sound systems (Maddieson, 1984), for example, lists no sound that occurs in all languages, even though broad patterns are seen. The number of significant sounds, or phonemes, ranges from about 12 (Pirahã, Rotokas) to over a hundred (!Xóõ), and the mechanisms used vary greatly as well.

"Documenting differences among the world's most disparate languages is of central importance to the field of linguistics and to the language community's heritage."

The acoustics of speech have proven to be extraordinarily complex. Early estimates that simple acoustic pattern matching would make automatic speech recognition practical (e.g., Juang and Furui, 2000) proved to be wrong. Current high levels of recognition are founded on acoustic analysis of huge amounts of data combined with statistical inference about common co-occurrences among words and sounds (e.g., Jelinek, 2009). Understanding what it is that listeners are sensitive to in this complex acoustic signal has been fruitfully guided by examining how those sounds are generated in the vocal tract (e.g., Iskarous *et al.*, 2010). For sounds in lesser-studied languages, having articulatory data to help interpret the acoustic signal is even more valuable.

A few of the world's languages have been well-studied, but most of them have yet to be explored in any detail, if at all. Pioneering efforts by Peter Ladefoged and Ian Maddieson to record the sounds of the world's languages resulted in descriptions of many of the less typical sounds used (Ladefoged and Maddieson, 1996) and phonetic sketches of several languages (e.g., Ladefoged *et al.*, 1998; McDonough *et al.*, 1993; Silverman *et al.*, 1995; Taff, *et al.*, 2001). Funded for many years by the National Science Foundation, this work has provided an invaluable basis for further phonetic work, given that it provides an initial understanding of the expected production and acoustic bases of virtually all the sounds that are used in language.

These descriptions are far from complete, however, as can be seen from two trends. First, we are still learning new and important facts about even the best-known languages, including English, as evidenced by the continuing appearance of phonetic studies in the pages of scientific journals. Second, even though the "same" sound may be described in phonetic studies from different languages, we can not assume that it shares the same characteristics across languages. For instance, ejective fricatives occur in several languages (Maddieson *et al.*, 2001), but they vary widely in how they are produced. Such wide variability calls into question the validity of these phonetic categories across languages. There is much left to learn.

Phoneticians have begun stepping up their efforts to study endangered languages while there are still fluent speakers left, especially those who acquired the language as their first. Language documentation efforts have been on the upswing in recent years, but phonetic studies have not been as obviously useful as the collection of texts and the making of dictionaries. As more communities try to revive their languages from documentary sources, it is becoming increasingly clear that phonetic documentation can contribute in valuable ways to describing the pronunciation of the ancestral language. Furthermore, many endangered languages lack writing systems, and good phonetic descriptions can often help guide their development. Contributing to community literacy is a frequent concrete aim. The scientific goals of the academic community often overlap with the revitalization goals of native communities.

Table 1: Tahltan consonants. Highlighted columns participate in the three-way harmony; those left unshaded are transparent to harmony. The orthography used draws on three characters commonly used in Americanist traditions. These differ from the International Phonetic Alphabet as follows: ž for ʒ, š for ʃ, and y for j.

b	d	dl	dθ	dz	dž	g	gʷ	G	
	t	tl	tθ	ts	tš	k	kʷ	q	
	t'	tɬ'	tθ'	ts'	tš'	k'	kʷ'	q'	ʔ
		ɬ	θ	s	š	x	xʷ	X	h
		l	ð	z	ž	ɣ	ɣʷ	ʁ	
m	n				y		w		
	n'								

In this paper, we will discuss two phonetic studies of endangered languages. Both studies use articulatory and acoustic data to examine specific scientific questions. The first study uses ultrasound, while the second uses acoustic data coupled with electroglottography (EGG).

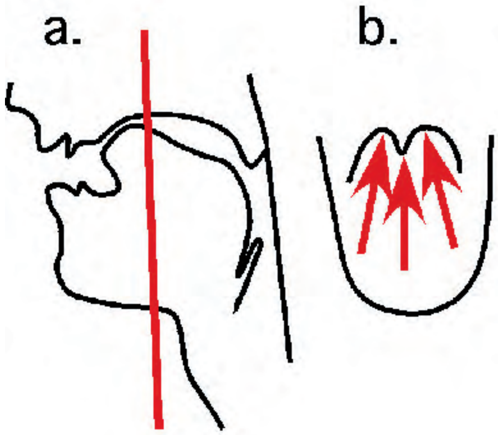
Ultrasound study of tongue shape in Tahltan

Tahltan (ISO 639 code tht) is an Athapaskan language of northern British Columbia, spoken by fewer than 20 elders as their first language. Some younger community members are learning the language as a second language, and the elders are hopeful that the language can be revived. There are three main dialects, and one of them, Telegraph Creek (see Fig. 1), has one of the world's few three-way consonant harmony systems. Harmony is a linguistic process in which sounds in a word have to agree on, or "harmonize with," certain dimensions. Vowel harmony is fairly common across languages, large and small. Some forms are rather limited, such as the umlaut process of German, while systems that affect all vowels can be found in some languages like Turkish. In Turkish, all vowels in a word should have the same frontness (the vowels of "bee" and "bay" are front, those of "boo" and "though" are back), although there are exceptions. Some affixes agree in rounding as well. (The English front vowels are unrounded; the back ones are rounded; Turkish has front rounded and back unrounded vowels.) Less common is consonant harmony, in which one or more features of the non-vowel sounds have to agree. Most such systems have two sounds that have to agree (say an "s" and an "sh"), but a handful have a three-way system of agreement.

All verb stems in Tahltan that contain fricatives (sounds like "s" and "sh") have to come from the same "series"—one that has "th" sounds (like in "thing"), "s" sounds or "sh" sounds (Shaw, 1991). It sounds simple, but Tahltan has 46 consonants, 15 of which participate in the harmony while the others are "transparent" to it in that they neither change nor stop the harmony (Table 1). When vowel harmony is at issue, it is easy to imagine that the consonants are overlaid on top of vowels and that the vowels are really adjacent underneath. With consonants, it seems instead that these segments are acting "at a distance," reaching across vowels that make use of the same articulator—the tongue—that the consonants

Fig. 1. View of the Tuya River crossing heading into Telegraph Creek, British Columbia, Canada.

Fig. 2. Data collection from speaker Margery Inkster with a portable ultrasound machine. The probe is held under the chin, giving a view across the tongue.



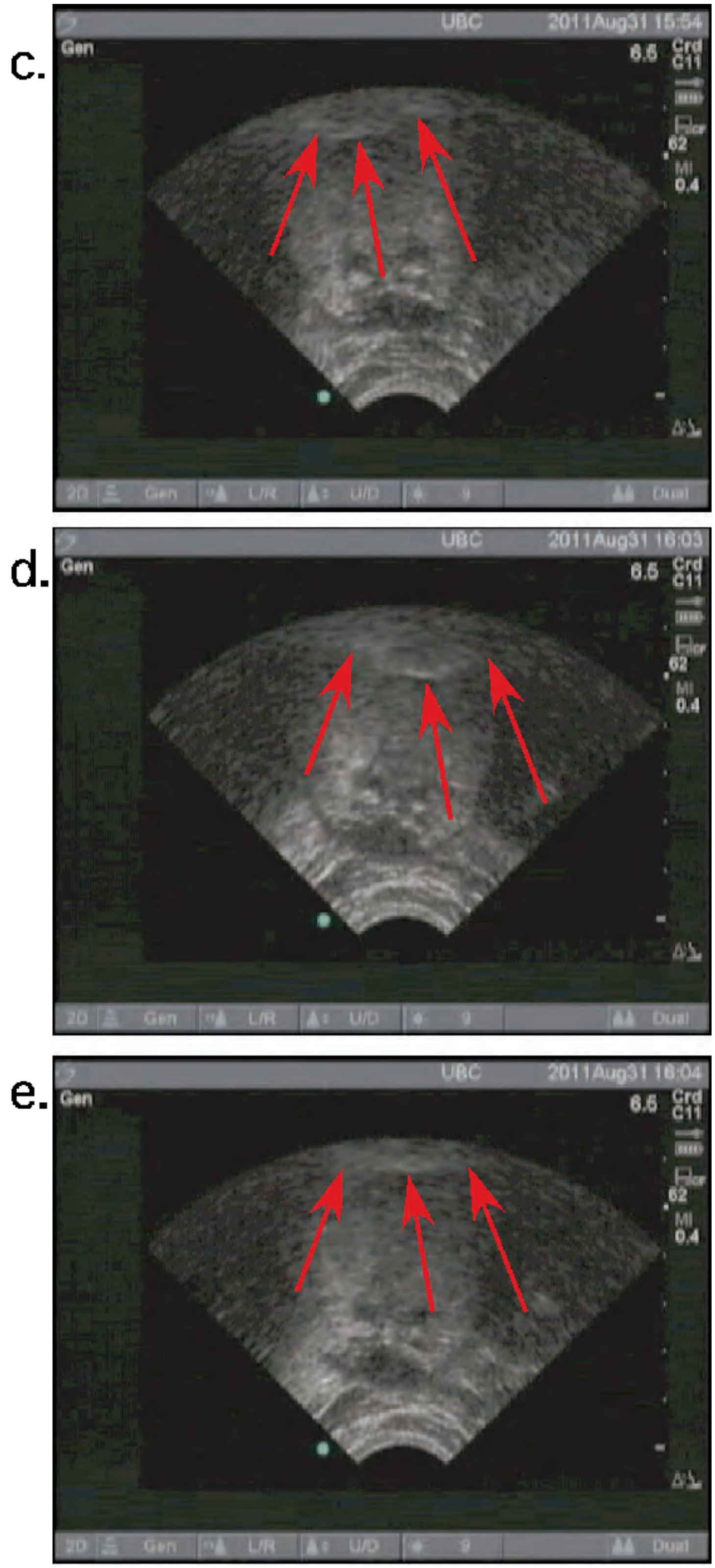


Fig. 3. Ultrasound imaging of the tongue. a: Schematic midsagittal vocal tract for an [š]-like articulation. The red line shows the approximate location of the plane of the ultrasound image. b: Schematic of the tongue shape in the coronal plane shown by the line in (a). The arrows mark the peaks and trough of the tongue groove. c-e: Three ultrasound images of the tongue during c) [θ] d) [s] e) [š]. The speaker's left is on the left and the speaker's right on the right. Selection of the three points was guided by repeated viewing of the video image, where the coherent structure of the tongue is more obvious than it is in any single frame.

depend on. Explaining this long-distance effect has been challenging.

The operation of this harmony can be seen in the following forms, where the underlined letter corresponds to the word segment "-s-", the marker for the first person singular subject:

ɛsk'ɑ:	'I'm gutting fish'
ɛšdžini	'I'm singing'
ɛθdu:θ	'I whipped him'

As can be seen from this last form, this agreement can work across intervening segments (here, /d/ and /u/) that neither promote nor block harmony. It looks like "action at a distance."

What if there is something that makes this process local after all? Gafos (1996) proposed just such a solution. He claimed that the differences in the amount of tongue grooving seen in these fricatives could extend through the intervening vowels and non-participating consonants, thus making the agreement local. Our efforts, supported by the National Science Foundation's Office of Polar Programs, have begun to test this hypothesis using ultrasound images of the tongue.

Ultrasound allows us to look at the tongue during speech (e.g., Stone and Lundberg, 1996) in a minimally invasive way that is appropriate for a broad range of speakers, including the elderly, and in non-laboratory conditions as well (Gick *et al.*, 2005). We generally take a "sagittal" view of the tongue, showing a two-dimensional image from near the tip to the back of the tongue near the uvula. However, we can just as easily take "coronal" sections that go across the tongue. This is ideal for measuring tongue grooves, and it is what we did in our study.

Figure 2 shows one of our speakers, Margery Inkster, with the ultrasound probe under her chin. Figure 3 shows three of the cross-sections of her tongue during the fricatives [θ s ʃ] The groove depth differs among the three, and this same groove persists through the vowel. This is in accord with Gafos's prediction. (See Figs. 2 and 3)

To understand where this pattern might have come from, however, we need to know whether the same kind of persistence appears in languages without the consonant harmony.

Fig. 4. Young Trique women dancing in traditional huipiles (dresses), Oaxaca, Mexico.

For this, we studied a language close at hand, English. Our preliminary results showed that, indeed, English also allows these three different groove depths to persist through the vowel (Whalen *et al*., 2011). Although no language is completely neutral as a comparison to harmony processes, it is useful to compare such patterns to languages for which we have greater phonetic knowledge, like English. So it is plausible that the foundations for a three-way harmony system could be seen in a language that has not yet shown any use of such a process.

We hope to extend this work by measuring images from additional Tahltan speakers and by making a more thorough comparison with English data. Nevertheless, these results are already helping to settle an important issue in linguistic theory: Consonant harmony can be seen as a local process after all.

Electroglottographic study of devoicing in Itunyoso Trique

Itunyoso Trique (ISO 639 trq) is an Oto-Manguean language spoken in Oaxaca, Mexico by approximately 1,400 people (Lewis, 2009). It is one of three Trique dialects, each of which has a unique sound structure and grammar. Like all Oto-Manguean languages, Itunyoso Trique is tonal. This means that the level and direction of pitch in the voice may distinguish the meaning of words. The Itunyoso dialect has nine different tonal melodies that words can carry (DiCanio, 2010). As a comparison, Mandarin Chinese is also tonal, but with only four possible tonal melodies. Most Trique speakers are bilingual, speaking Spanish as a second language. In recent years, children have begun to use Spanish more among their peers (IGABE, 2011). While there are currently many Trique speakers, the language is danger of being replaced entirely by Spanish (See Fig. 4).

The large inventory of tonal melodies is certainly a feature of the Itunyoso Trique language that is seldom found elsewhere, but it is not the only such feature. Itunyoso Trique also has a contrast between long and short consonants found only in the initial position of words. While many languages in the world have consonant length contrasts, such as Japanese *katta* 'bought' and *kata* 'shoulder', this distinction is often restricted to the middle of the word (Muller, 2001). Itunyoso Trique is one of only two languages in the world known to restrict this contrast between long and short consonants to the *beginning* of a word. The other language is Nhaheun (ISO 639 nev), an Austroasiatic language spoken in Laos (Muller, 2001).

One of the intriguing things about this rare contrast in Trique is how voicing functions in the short and long stop consonants. Stops are sounds with a closure in the mouth followed by a sudden release. Many languages distinguish between voiceless stops, like for instance, French "p", "t", and "k," and voiced stops, like French "b", "d", and "g." Voiced stops are produced when the vocal folds are vibrating, voiceless stops when the vocal folds are not vibrating. Yet, in Itunyoso Trique, the long consonants are often preceded by a short puff of air, called *preaspiration*. The short consonants are rarely produced this way, but vary in their production. They may be voiceless, like a "p", or voiced, like a "b." This variability has led researchers to misclassify length contrasts like the one in Itunyoso Trique as strength contrasts, i.e. “fortis” and “lenis” stops.

DiCanio (2012) examined the timing of vocal fold vibration in Trique consonants using electroglottography (EGG) in order to determine exactly what accounted for this variability in voicing in short stops and to investigate if another explanation could account for the pattern. With EGG, sensors are placed on opposite sides of the speaker’s neck, just over the thyroid cartilage (below the Adam’s apple) through which a weak electrical current is passed. When the vocal folds are closed, more of the current can pass from one side to the other. When the vocal folds are open, less current passes through. EGG maxima correspond to the moment of maximum contact between the vocal folds while minima correspond to the moment of minimum contact between the vocal folds (Childers and Krishnamurthy, 1985; Childers and Lee, 1991; Heinrich *et al.*, 2004). The presence of EGG maxima and minima indicates that there is vocal fold vibration. EGG data are typically collected along with acoustic recordings for the identification of acoustic-phonetic boundaries.

The advantage to using an electroglottograph is that the EGG signal is unaffected by acoustic disturbances in field recordings. In rural villages, recording is often done in quiet spaces in private homes. Typically, houses are constructed by community members without the help of an electrician. Thus, in addition to the external noise found in these communities, there are often ground loops due to the use of low current wiring. Ground loops in AC power lines can produce an unwanted signal in acoustic recordings, with a fundamental frequency of 60 Hz and its associated harmonics. To adjust for this effect, a low stop filter can be applied to the recordings; yet, this filter also eliminates low frequency voicing from the signal. Devoicing typically involves low frequency and low amplitude glottal pulses. EGG is ideal for examining devoicing in these less-than-ideal recording conditions because it accurately captures low amplitude voicing.

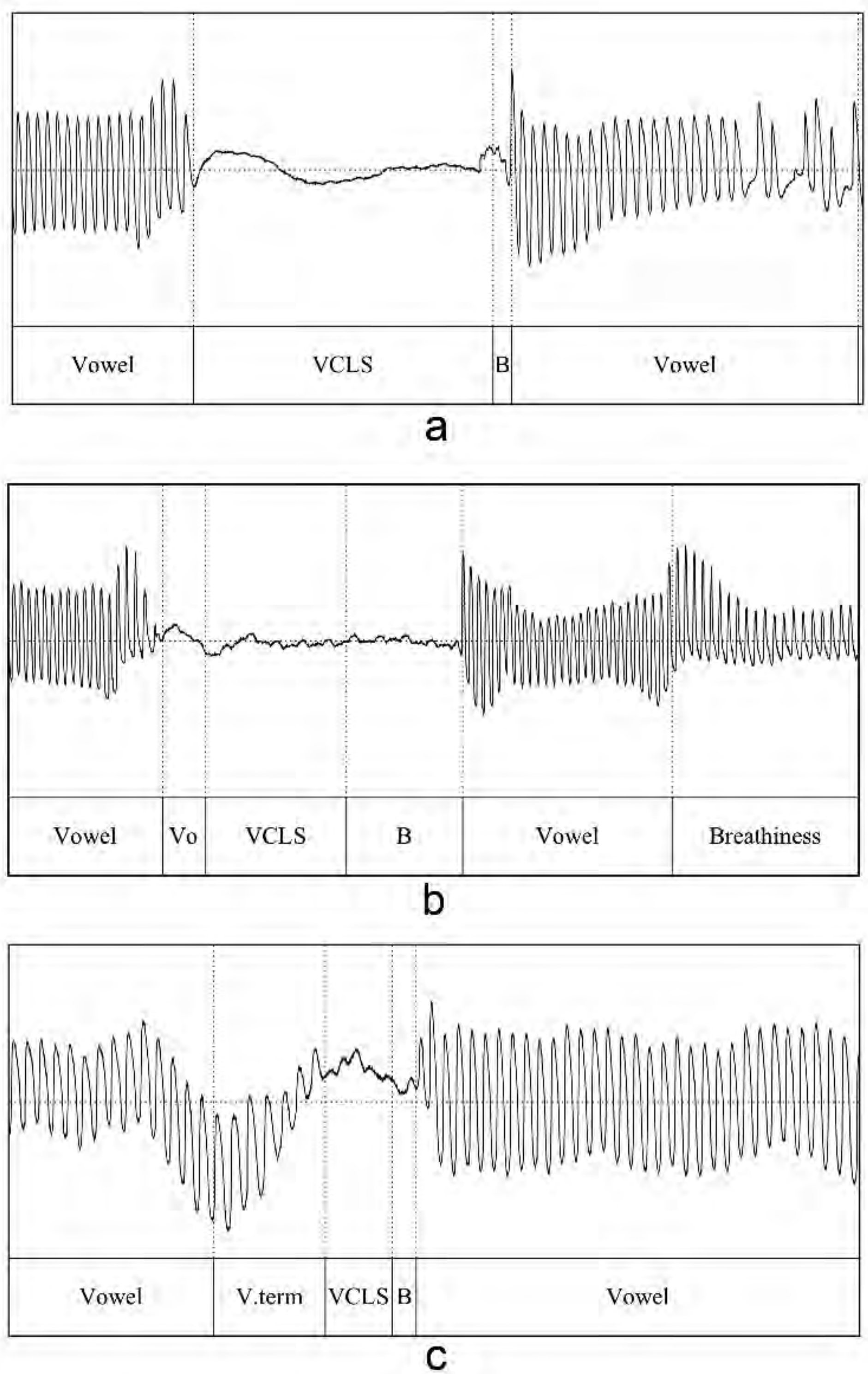


Fig. 5. Oral-glottal alignment configurations from EGG signal. VCLS represents voicelessness during closure. B represents the voiceless burst duration following closure release. Figures from DiCanio (in press). Reprinted with permission.

There are three ways in which vocal fold vibration may be aligned to the closure of a stop in running speech. If the sound preceding the stop is voiced, voicing may cease coincidental with the closure of the following stop. This is simultaneous oral-glottal alignment. Voicing may also cease during the vowel duration prior to stop closure. The time between devoicing and closure is called *voice offset time (Vo)*. This is the gestural configuration used for

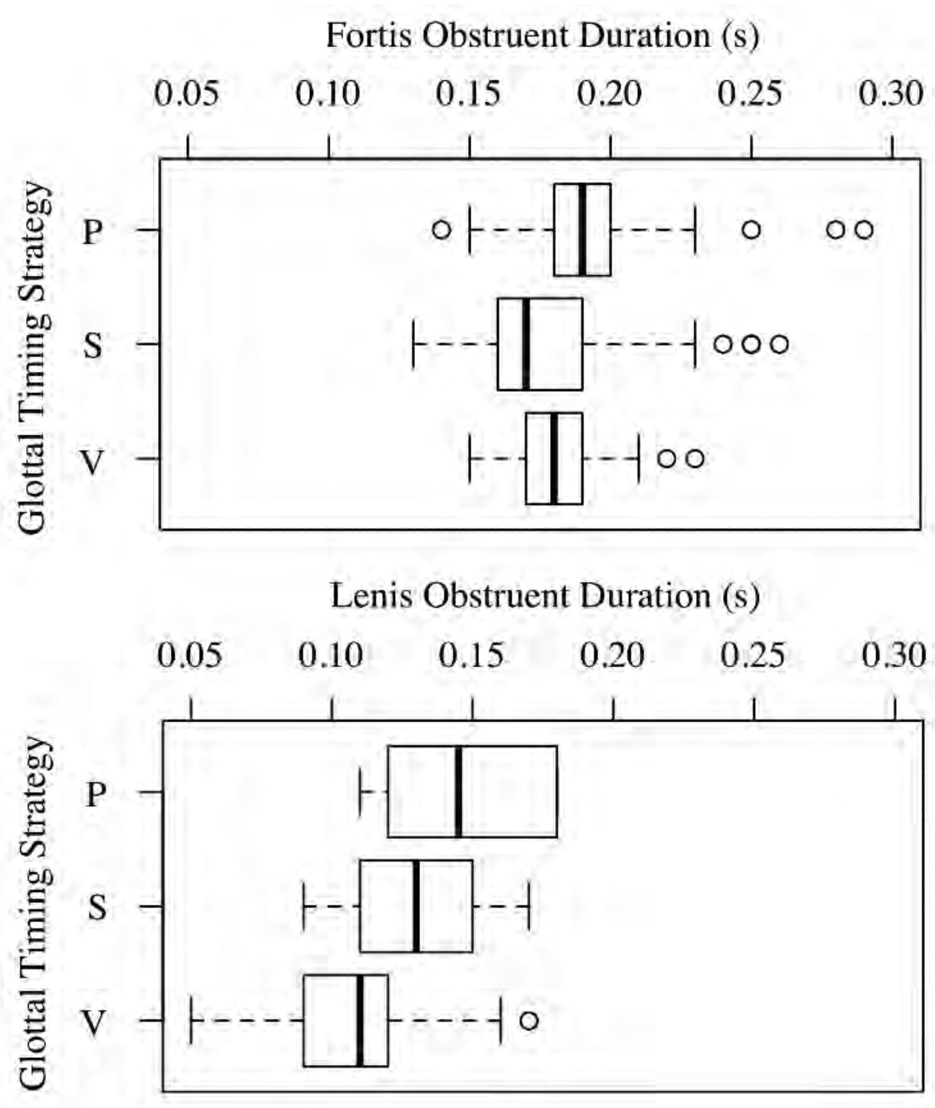


Fig. 6. Effect of total stop duration on glottal timing strategy. Long stop data (left), Short stop data (right). (P = preaspiration, S = simultaneous timing, V = voicing during closure). Figure from DiCanio (in press). Reprinted with permission.

preaspirated stops, as in Icelandic (Helgason, 2002). Finally, voicing may cease during the stop closure duration. This is called *partial voicing* or *voice termination time (V.term)*. Figures 5a-5c show these three oral-glottal alignment configurations, respectively.

In DiCanio (2012), acoustic and EGG recordings were made from four speakers producing short and long stops in Itunyoso Trique. These words were presented in contexts so that the duration of stop closure could be examined from the acoustic signal along with the alignment of devoicing from the EGG signal. The results found significant differences between the long and short stops in relation to the timing of devoicing. Long stops are produced with either simultaneous glottal timing (49.4%), as shown in Fig. 5a, or with devoicing prior to closure (43.8%), as shown in Fig. 5b. Short stops are typically produced with voicing which extends into the stop closure (84%), as shown in Figure 5c, and rarely with simultaneous oral-glottal alignment (14.1%). The amount of voicing during closure for the short (lenis) stops varied in relation to the overall duration of the stop, shown in Fig. 6. No such variation was observed for long (fortis) stops.

Research on related languages like Zapotec is inconclusive as to whether short stops are voiced or voiceless (Avelino, 2001; Nellis and Hollenbach, 1980). These results suggest that such variability is conditioned by within-category changes in consonant duration. The effect of such variability is that, for some short stops with particularly short duration, voicing may extend through the entire stop. This pattern is called passive voicing (Jansen, 2004; Westbury and Keating, 1986). Long stops show a different pattern. They are actively devoiced by an abrupt glottal spreading gesture, the timing of which does not vary with consonant duration.

This work addresses a general descriptive question within Itunyoso Trique: what is the phonetic realization of short and long stops in the language? However, the answer to this question also informs a more general theory of speech production. Voicing is generally considered to be a discrete category within the phonology of a language. Sounds are typically classified as either "voiced," with vocal fold vibration, or "voiceless," without vocal fold vibration. The Trique data suggest that for certain sound types, there is a continuum of voicing that varies due to durational differences. Findings like this illustrate one of the ways in which descriptive phonetic work on endangered languages has a broader impact on the linguistic sciences.

Summary

We have presented just two of the many phonetic investigations currently under way that examine the world's extensive, but shrinking, variety of languages. Documenting differences between the world's most disparate languages is of central importance to the field of linguistics and to the language community's heritage. Such efforts are funded by such governmental agencies as the National Science Foundation and the Administration for Native Americans, and non-profits like the Endangered Language Fund. While the work presented here focuses mainly on the production of consonants, many more aspects of speech acoustics need to be investigated. For instance, prosody (speech timing, intonation) varies substantially across languages; yet, this topic is rarely addressed in studies of endangered languages. As more communities attempt to revive their heritage languages or seek the help of linguists in the development of writing systems, phonetic detail becomes more important. The techniques that are now available for doing phonetic research in field locations are much better than those available even a decade ago, and they continue to improve. The full range of phonetic diversity in human languages should become clearer in the coming years as many more intriguing patterns are discovered.

Acknowledgments

This research was funded by National Science Foundation grants NSF-1128159 and NSF-0966411 to Haskins Laboratories. Reethee Antony, Colleen Leung and Aude Noiray provided crucial support in the collection and analysis of the Tahltan data, as did Bryan Gick and Donald Derrick. Acknowledgments are given to the community of San Martín Itunyoso.**AT**

References

Avelino, H. (**2001**). "Phonetic correlates of fortis-lenis in Yalálag Zapotec, " unpublished Master's thesis, University of California, Los Angeles.

Childers, D. G., and Krishnamurthy, A. K. (**1985**). "A critical review of electroglottography," CRC Critical Reviews in Biomedical Eng. **12**, 131–161.

Childers, D. G., and Lee, C. K. (**1991**). "Vocal quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

DiCanio, C. T. (**2010**). "Itunyoso Trique (Illustrations of the IPA)," J. Intl. Phonetic Assn. **40**, 227–238.

DiCanio, C. T. (**2012**). "The phonetics of fortis and lenis consonants in Itunyoso Trique," Intl. J. Am. Linguistics **78**, 239–272.

Gafos, A. (**1996**). "The articulary basis of locality in phonology," unpublished Ph.D. dissertation, Johns Hopkins University, Baltimore, MD.

Gick, B., Bird, S., and Wilson, I. (**2005**). "Techniques for field application of lingual ultrasound imaging," Clinical Linguistics and Phonetics **19**, 503–514.

Heinrich, N., D'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

Helgason, P. (**2002**). "Preaspiration in the Nordic Languages: Synchronic and diachronic aspects," unpublished Ph.D. thesis, Stockholm University.

IGABE (News Agency) (**2011**, March 9). "Resisten niños triquis a usar lengua materna," Noticias Voz e Imagen.

Iskarous, K., Nam, H., and Whalen, D. H. (**2010**). Perception of articulatory dynamics from acoustic signatures. J. Acoust. Soc. Am., **127**, 3717-3728.

Jansen, W. (**2004**). "Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch," unpublished Ph.D. dissertation, Groningen University.

Jelinek, F. (**2009**). "The dawn of statistical ASR and MT," Computational Linguistics **35**, 483–494.

Juang, B. H., and Furui, S. (**2000**). "Automatic recognition and understanding of spoken language: A first step toward natural human-machine communication," Proc. of the IEEE **88**, 1142-1165.

Ladefoged, P., Ladefoged, J., Turk, A., Hind, K., and Skilton, S. J. (**1998**). "Phonetic Structures of Scottish Gaelic," J. Phonetics **28**, 1–41.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages*. (Blackwell, Oxford, UK; Cambridge, MA) .

Lewis, M. P. (Ed.). (**2009**). *Ethnologue: Languages of the World, Sixteenth edition.* (SIL International, Dallas).

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

Maddieson, I., Smith, C. L., and Bessell, N. (**2001**). "Aspects of the phonetics of Tlingit," Anthropological Linguistics **43**, 135–176.

McDonough, J., Ladefoged, P., and George, H. (**1993**). "Navajo vowels and phonetic universal tendencies," UCLA Working Papers in Phonetics **84**, 143–164.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature **264**, 746–748.

Muller, J. S. (**2001**). "The phonology and phonetics of word-initial geminates," unpublished Ph.D. thesis, The Ohio State University.

Nellis, D. G., and Hollenbach, B. E. (**1980**). "Fortis versus Lenis in Cajonos Zapotec Phonology," Intl. J. Am. Linguistics **46**, 92–105.

Shaw, P. A. (**1991**). "Consonant harmony systems: The special status of coronal harmony," in C. Paradis and J.-F. Prunet (Eds.), Phonetics and Phonology 2: The Special Status of Coronals (pp. 125–157) (Academic Press, San Diego).

Silverman, D., Blankenship, B., Kirk, P., and Ladefoged, P. (**1995**). "Phonetic structures in Jalapa Mazatec," Anthropological Linguistics **37**, 70–88.

Stone, M. L., and Lundberg, A. (**1996**). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. **99**, 3728–3737.

Taff, A., Rozelle, L., Cho, T., Ladefoged, P., Dirks, M., and Wegelin, J. (**2001**). "Phonetic structures of Aleut," J. Phonetics **29**, 231–271.

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

Whalen, D. H., Shaw, P. A., Noiray, A., and Antony, R. (**2011**). "Analogs of Tahltan consonant harmony in English CVC syllables," in W.-S. Lee and E. Zee (Eds.), *Proc. of the 17th International Congress of Phonetic Sciences* (pp. 2129–2132). (City University of Hong Kong, Hong Kong.)

Christian T. DiCanio is a research associate at Haskins Laboratories. He received his Ph.D. in Linguistics from the University of California, Berkeley, in 2008. Prior to his employment at Haskins, he was a Fyssen Foundation postdoctoral fellow at Université Lumière in Lyon, France. His research encompasses both descriptive phonetic work on endangered languages and experimental work in speech production and perception. He has done field work on a variety of endangered languages in Mexico, including Trique, Ixcatec, and Mixtec. Apart from his descriptive linguistic work, his areas of research include tone perception, voice quality perception, voice quality production, and the coarticulatory dynamics of tone. Together with colleagues at Haskins Labs, he is investigating how automatic speech recognition software can be used to align text to speech in corpus data from different endangered languages.

Patricia A. Shaw received her Ph.D. in Linguistics from the University of Toronto, and is the Founder and Chair of the First Nations Languages Program at the University of British Columbia in Vancouver, BC, where she provides training for language documentation and revitalization and engages in research on and teaching of Kwak'wala, Musqueam and Cree. She has worked in close collaboration with several critically endangered language communities (Salish, Wakashan, Siouan, Athapaskan, Tsimshian) to record and analyze extant grammatical knowledge, to teach research skills and archiving methodologies, and to develop pedagogical materials for language revitalization. She is currently President of the Society for the Study of the Indigenous Languages of the Americas (SSILA), an organization that brings linguists and community members together to work on languages of North, Central, and South America.

D. H. Whalen received his B.A. from Rice University and his Ph.D. from Yale University. He is currently Distinguished Professor of Speech-Language-Hearing Sciences at the Graduate Center of the City University of New York. He also serves as Vice President of Research at Haskins Laboratories, a private non-profit research institute in New Haven, CT. He is Founder and President of the Endangered Language Fund, a nonprofit organization dedicated to the description and revitalization of languages at risk of falling silent. You can learn more at www.endangeredlanguagefund.org. He is a Fellow of the Acoustical Society of America.