Click Cavity Formation and Dissolution in IsiXhosa: Viewing Clicks with High-Speed Ultrasound

Amanda L. Miller

University of British Columbia and Cornell University
E-mail: amandami@interchange.ubc.ca

Abstract

The Corrected High-Speed Ultrasound with Software Alignment (CHAUSA) architecture and method for collection of portable high-speed ultrasound data is presented. Simultaneous head to probe anchoring using Articulate Instruments' Ultrasound Stabilization Headset and Head-correction using the Palatoglossatron method are used. Results from a preliminary experiment on the Bantu language IsiXhosa post-alveolar click show complete click cavity formation and dissolution in this click. Results are compared with X-ray data from the Khoesan language !Xóõ at 30 fps, and 30 fps ultrasound studies on the Khoesan languages Khoekhoe and N|uu. The posterior constriction in the Xhosa post-alveolar click starts out as velar, similar to the pulmonic dorsal stop [g]. Tongue dorsum retraction is shown to be present during click cavity dissolution, with the posterior constriction being uvular at the time of the posterior release. X-ray data on !Xoo [!] click showed it to have a velar posterior constriction, while the slower US data on Khoekhoe and N|uu showed these clicks to have a uvular posterior constriction, involving tongue root retraction. Differences may be due to differences in frame rate used in the earlier studies. Tongue tip recoil is shown to be present following the fast anterior apical release of the post-alveolar click in IsiXhosa.

1 Introduction

A Corrected High-speed Anchored Ultrasound with Software Alignment (CHAUSA) architecture and method is presented. This method has been developed to be portable for use in linguistic fieldwork. An experiment undertaken on the production of the IsiXhosa post-alveolar click shows that the method is capable of viewing tongue dorsum retraction and tongue tip recoil.

2 Background

2.1 Speed

While ultrasound machines typically are able to collect data at rates far higher than 30 fps; in practice, speech production data has been limited to the standard 30 fps video rate as it exits the machine. This is because most researchers have been using external monitor ports or s-video ports to get the data out of the machine.

2.2 Alignment

Ultrasound machines have not been outfitted to collect simultaneous tongue video and audio input important for the analysis of speech. Most researchers have achieved articulatory – acoustic alignment by exporting the video from the machine, and mixing it with a separate audio signal using a video mixer or a VCR. However, this requires digital to analog conversion in getting the US signal out of the machine. Thus, results have been tied to the standard 30 fps analog video signal. Further, some machines, such as the GE Logiqbook result in a 3-4 frame delay at 30 fps when mixed in a video mixer.

2.3 Head and Probe Stabilization and Correction

Ultrasound differs from X-ray and MRI data in that the US signal does not image hard bony structures of the vocal tract. Thus, movement of the tongue seen in US images can also be due to movement of the head. There have been two classes of approaches to solve this problem: head stabilization and head movement correction. Head stabilization has been achieved for a laboratory setting with the HATS system by Stone et al. [1]. Gick et al. [2] and Davidson [3] have developed simpler approaches. The system in [2] is portable.

Head-movement correction is developed by Whalen et al. [4] in the HOCUS system; wich uses optical tracking to measure head position. Mielke et al. [5] have developed a simpler, cheaper and portable system for use in linguistic fieldwork. The system uses two sticks: one attached to the ultrasound probe, and one attached to a pair of glasses on the head. The sticks are fitted with pink dots, and the dots are tracked with a video camera as the person speaks to measure head and probe movement. Software has been developed to correct the traced tongue images for head and probe movement.

3 CHAUSA Architecture and Method

The CHAUSA architecture and method was developed to attain high-speed data, while controlling for all of these problems in a way that would be portable, and could be used for linguistic fieldwork. The new architecture is provided in Figure 1. High-speed ultrasound data is collected in the portable GE LogiqE ultrasound machine, and is transferred to the laptop using DICOM (Digital Communications in Medicine) file transfer protocol via the Ethernet port in a post data collection step. During US data collection, low speed US data is simultaneously mixed with audio in the Canopus Twin 100 audio-video mixer. This path, called the CANOPUS path, has low-speed US data, as the signal is converted to the standard Analog video rate of 30 fps when it comes out of the external monitor port of the LogiqE. The high-speed US video collected via the DICOM path, and the low-speed US video and audio collected via the CANOPUS path are aligned post-hoc using Adobe Premiere Pro. Alignment is undertaken using multiple stop bursts. Acoustic stop bursts are aligned with the downward motion of the tongue in the release of the stop. Multiple bursts improve the accuracy of the alignment. We therefore collect 3 repetitions of the frame sentence including the target word in each repetition.

Figure 1: *CHAUSA Architecture*

Articulate Instrument's Ultrasound Stabilization Headset [6] assures that the ultrasound transducer is kept in the same position throughout the recording session. A pilot study showed that the head still moved approximately 4-5 mm during data collection, and thus head correction must also be undertaken. The Palatoglossatron head correction method [5] is usable in a fieldwork system. A Figure 2 provides a picture of a Xhosa subject wearing the ultrasound stabilization headset, combined with the Palatoglossatron hardware. An image from this viewpoint is recorded with a video camera in order to record the movement of the head and probe.

4 Xhosa Click Cavity Formation and Dissolution

4.1 Introduction

Previous studies on the production of clicks have all been undertaken with a sampling rate of 30 fps. Traill [7] used X-ray cineradiography for all five !Xóõ clicks. Thomas-Vilakati [8] used electropalatography for all 3 IsiZulu clicks, and Miller, Namaseb and Iskarous [9] and Miller et al. [10] used 30 fps ultrasound captured with just the CANOPUS path t to image Khoekhoe and N|uu palatal and alveolar clicks.

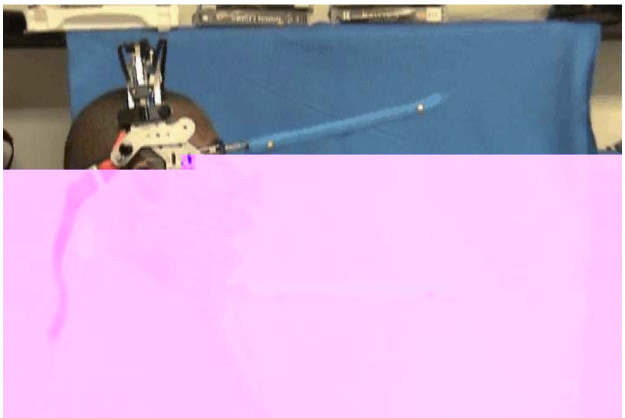

Figure 2: *Xhosa subject wearing the ultrasound stabilization headset and Palatoglossatron hardware*

While [7] called the posterior constriction of the post-alveolar click velar, images show the constriction to be far back in the mouth. Thomas-Vilakati's EPG study of the post-alveolar click showed that the posterior constriction for [!] was not always visible in EPG, suggesting a farther back constriction than was found in c [|] and x [||]. [9, 10] both showed a uvular posterior constriction in the alveolar click in Khoekhoe and N|uu, as well as a retracted tongue root. Given the slow frame rate of these earlier studies, it was not possible to view the release dynamics of the [!] click. The objective of the current study is to investigate the dynamics of the posterior release in the Xhosa post-alveolar click.

4.2 Methods

One Xhosa speaker was imaged during the production of the [!] click in the IsiXhosa sentence *Ndi qaba isonka.* [ⁿdi !aba isoŋga] *'I'm spreading (butter) on the bread.'* The utterance was repeated 15 times (3 tokens in each of 5 repetitions). The data was collected during one seating of the Ultrasound stabilization headset. The palate was imaged by having the subject drink water, and hold the water in place in the mouth, before swallowing it, as in [11].

Software alignment of the high-speed video with the audio was undertaken in Adobe Premiere Pro 2.0. All six tokens of [!] and [g] over the 3 repetitions were aligned in a single recording, allowing for maximal accuracy.

The head video was mixed with the tongue video in Adobe Premiere Pro. The resulting image was used to trace the tongue edge in Palatoglossatron software described in [5], and to correct for head and probe movement.

4.3 Results

Figure 3 provides a trace of the palate, the tongue in [k], and the closure and release phases of the [!] click. As can be seen, the posterior constriction in the [!] click at the beginning of the closure coincides with the front portion of the velar constriction in the [g] from *isonka*, showing that the closure is definitely velar. The trace of the release of the [!] click shows a much farther back constriction. The soft palate was not visible in the swallow, thus it was not traced. We can surmise that the peak dorsal constriction is making contact with the uvula, suggesting a uvular constriction. These data thus show that there is tongue dorsum retraction during the click cavity dissolution in the Xhosa post-alveolar click. However, tongue root retraction clearly seen in [9] and [10] are not as visible here.

Figure 4 provides traces of the post-alveolar click taken during and after the posterior release, and highlights the presence of tongue tip recoil in the post-alveolar click. Trace 4 is the same as the trace showing the posterior release in Figure 3. The tongue front is still facing upward in this trace (the tongue tip was not clearly imaged during this trace). A subsequent trace (Trace 5) shows a bend still in the tongue front, with the tongue tip laying flat in the mouth. Trace 6 shows the entire tongue front flat and low in the mouth. Trace 7, however, shows the tongue blade raised again to the position it was in just after release. The tongue tip is bent down low in the mouth. I interpret this as tongue tip recoil due to the fast release of the anterior constriction. Every token of this click that we imaged showed this anterior release type.

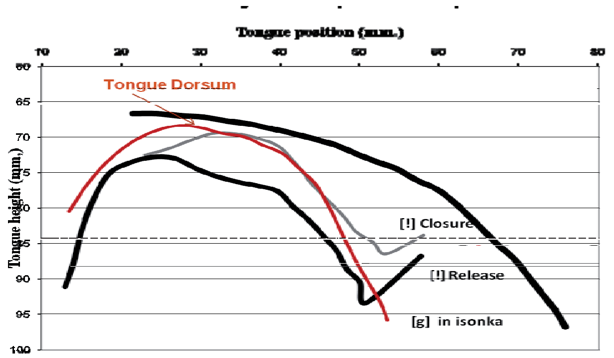

Figure 3: *Tongue traces in the frame prior to release of the [g] in isonka*, and the closure and release phases of the [!] click, along with the palate

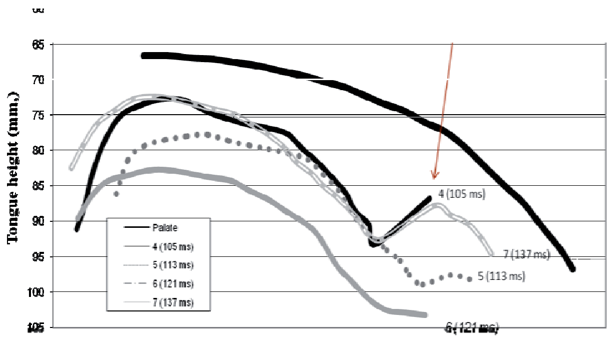

Figure 4: *Tongue traces of the frame prior to release of [!], as well as following traces showing tongue tip recoil, along with the palate*

4.4 Discussion

Retraction of the tongue dorsum during the closure phase of the alveolar click in Xhosa is part of the cavity expansion necessary for rarefaction in clicks. The posterior tongue position at the time of release is uvular, just as it is in the Khoesan languages Khoekhoe [9] and N|uu [10]. TRR seen in the alveolar click in [9] and [10] is not clearly visible here. This suggests that TRR may not be an automatic consequence of tongue dorsum retraction.

5 Conclusion

The CHAUSA architecture and method have enabled viewing of the complete cycle of click cavity formation and dissolution in the IsiXhosa alveolar click. High-speed data has shown tongue dorsum retraction in formation of click cavity that is necessary for rarefaction of air. Tongue root retraction seen in earlier studies is not seen in IsiXhosa clicks, suggesting a possible difference in production in this language compared with Khoesan languages. A new phenomenon of tongue tip recoil has been seen.

References

[1] M. Stone and E. Davis. A head and transducer support system for making ultrasound images of tongue/jaw movement. *J. Acoust. Soc. Am.* ,98, 3107-3112, 1995.

[2] B. Gick, S. Bird and I. Wilson. Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics and Phonetics*, 19, 503-514, 2005.

[3] L. Davidson. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance.*J. Acoust. Soc. Am. 120,1:*407-415, 2006.

[4] D.H. Whalen, K. Iskarous, M.K. Tiede, D.J. Ostry, H. Lehnert-LeHouillier., E. Vatikiotis-Bateson and D.S. Hailey. The Haskins Optically Corrected Ultrasound System (HOCUS), *Journal of Speech, Language, and Hearing Research*, 48, 543-553, 2005.

[5] J. Mielke, A. Baker, D. Archangeli, and S. Racy. Palatron: a technique for aligning ultrasound images of the tongue and palate, In D. Siddiqi and B. V. Tucker, eds.,*,Coyote Papers,* 14, 97-108, 2005.

[6] Articulate Instruments. Ultrasound Stabilisation Headset User's Manual, Revision 1.3. Edinburgh, UK: Articulate Instruments Ltd, 2008.

[7] A. Traill. *Phonetic and phonological studies of* !*Xóõ Bushman* (Quellen zur Khoisan-Forschung 1). Hamburg: Helmut Buske Verlag., 1985.

[8] K. Thomas-Vilakati. *Coproduction and coarticulation in isiZulu clicks*. Ph.D. dissertation, University of California at Los Angeles, 1999.

[9] A. Miller, L. Namaseb and K. Iskarous. Tongue body constriction differences in click types. In J. Cole and J. Hualde eds., *Laboratory phonology 9* (Phonology and phonetics series). New York: Mouton de Gruyter,2007.

[10]A. Miller, J. Brugman, B. Sands, L. Namaseb, M. Exter and C. Collins. Differences in Airstream and Posterior Place of Articulation among Nǀuu Clicks. To appear in *J. Int. Ph. Assoc.*

[11] M. Epstein and M. Stone. The tongue stops here: Ultrasound imaging of the palate. *J. Acoust. Soc. Am.* 118 (4), 2128-2131.