

# **Click consonant production in Khoekhoe: A real-time MRI study**

Michael Proctor (Macquarie University)

Yinghua Zhu (University of Southern California)

Adam Lammert (Swarthmore College)

Asterios Toutios (University of Southern California)

Bonny Sands (Northern Arizona University)

Ulrich Hummel (Polytechnic of Namibia)

Shrikanth Narayanan (University of Southern California)

## **Abstract**

Production of click consonants by a single speaker of Khoekhoegowab was examined using real-time Magnetic Resonance Imaging (rtMRI). This technology provides high frame-rate dynamic information about the configuration of the entire upper airway, including the lips, jaw, velum, sublingual cavity, tongue-root, and glottis. We demonstrate the utility of rtMRI in the study of clicks, as an unparalleled method for tracking the global coordination of speech articulators. Patterns of tongue shaping and lingual kinematics of click consonants in the Nama dialect of Khoekhoegowab are described, and the implications of this data for phonetic characterization and phonological models of linguo-velaric consonants are discussed.

## **1. Introduction**

The characteristic phonetic properties of click consonants have been well described in a variety of languages. Click production has been examined using X-ray (Doke 1923b, 1925; Traill 1985), palatography and linguography (Doke 1923a, 1925; Beach 1938; Traill 1985; Nakagawa 2006; Sands et al. 2007), and ultrasound (Miller et al. 2007b, 2009a, 2009b, etc.). The aerodynamic (Doke 1923a; Ladefoged & Traill 1984; Traill 1985, 1991; Nakagawa 2006), and acoustic properties of linguo-velaric consonants (Sands 1991; Ladefoged & Traill 1994; Nakagawa 1996, 2006; Miller et al. 2007a) have also been

analyzed in detail.

Although we have a good understanding of the principal mechanisms of click production as a result of this extensive body of work, the phonological characterization of linguo-velaric consonants is still a matter of debate (Köhler et al. 1988; Nakagawa 2006; Brugman 2009; Miller et al. 2009a; Miller 2011; Bradfield 2014). In particular, there is no clear consensus on the constituency of click consonants: whether they are best described as a small set of underlying segments which combine in clusters (Traill 1985, 1993; Nakagawa 2006), or as a larger inventory of unitary segments which resist decomposition (Beach 1938; Snyman 1970; Ladefoged & Traill 1994; Miller-Ockhuizen 2003). Also at issue is whether contrastive airstream mechanisms should be considered an intrinsic characteristic of clicks or not (Miller et al. 2009a).

Further insights into each of these issues will require more data on click production across a variety of languages; however, an intrinsic limitation of ultrasound and palatography – the main methods which have been used to study click consonants to date – is that they do not provide information about the configuration of the vocal tract beyond the tongue and palate. In particular, ultrasound and palatography cannot track the state of the velum, jaw, larynx and glottis, nor reliably image the posterior regions of the tongue, including the lower pharynx and tongue root.

Because articulatory data on click production in most languages has been restricted to the region of the vocal tract associated with the primary mechanisms of sound production, it is not surprising that fundamental questions about the phonetics of these consonants remain. Considering that much of the debate about clicks relates to ‘accompanying’ articulation and details of airstream mechanisms, a comprehensive understanding of their phonetic and phonological properties will remain elusive while we lack sufficiently detailed information about the *whole* tract. X-ray can image the entire upper airway, and x-ray studies have provided some of the most comprehensive information about the phonetics of click consonants to date (Doke 1923b; Traill 1985);

however, because of the health risks arising from exposure to ionizing radiation, there are limits to the viability of this modality.

This paper reports on an initial investigation into click production, designed to address some of these issues. The goals of this study are twofold:

- to investigate the utility of rtMRI as a new method for studying linguo-velaric consonant production
- to investigate the global configuration of the vocal tract throughout click production by a single speaker, to better understand the nature of click consonant contrasts

### 1.1 Khoekhoe click consonants

Khoekhoegowab is a Khoe dialect chain, primarily spoken in Namibia, whose major varieties include the ethnolects Nama, Damara, and Hai||om (Haacke et al. 1997; Güldemann & Vossen 2000). It is the Nama variety of Khoekhoegowab which is the subject of this study (ISO 639-3 code: naq).

Nama uses an extensive segmental inventory, distributed over six contrastive places of articulation and produced with seven broad manners of articulation (Brugman 2009). Twenty of thirty-four Nama consonants (59%) are produced with a lingual airstream: five series of clicks, produced at four contrastive places of articulation. Click place contrasts in Nama are uncontroversially described as dental, alveolar, palatal, and lateral (Beach 1938; Westphal 1971; Ladefoged & Traill 1984, 1994; Güldemann 2001; Haacke & Eiseb 2002; Brugman 2009; Fredericks 2013), but the exact nature of manner contrasts amongst Nama clicks is less clear. Ladefoged and Traill (1984, 1994) describe a system of twenty click segments differing only in terms of aspiration, nasalization, and glottalization (Table 1), while Brugman (2009) re-analyzes the Nama linguo-velaric consonant system into separate series of stops, affricates, and nasals. Traill (1993) and Güldemann (2001) propose that these contrasts should be modeled as consonant clusters, and Miller (2011) characterizes Nama clicks as a set of unitary segments differentiated by contour airstreams.

	Glottal closure	Voiceless unaspirated	Voiceless aspirated	Delayed aspiration	Voiced nasal
Dental		g	kh	h	n
Alveolar	!	!g	!kh	!h	!n
Palatal	‡	‡g	‡kh	‡h	‡n
Lateral		g	kh	h	n

Table 1. Nama click consonant inventory: Orthographic representations (Curriculum Committee for Khoekhoegowab 2003) for four place and five manner contrasts (descriptions following Ladefoged & Traill 1984, 1994).

## 2. Corpora and data acquisition

The study participant (author UH) is a 35-year-old Namibian male trilingual speaker of Afrikaans, Nama, and English. He was born and raised in Windhoek and has lived most of his life in Namibia; at the time of the study he had been living in the United States for a year. His mother speaks Nama as her first language and his father, Afrikaans.

The informant read out wordlists and prose eliciting Nama consonant contrasts, as he lay supine in an MRI scanner. Stimuli were presented to the participant in Nama orthography, with accompanying Afrikaans/English translations to clarify homophones and uncommon lexical items. In addition to the study corpus, some spontaneous speech was recorded. Most items were elicited twice. Thirty-five recordings were made in total, each lasting between 7 and 61 seconds. A list of the subset of words analyzed for this study, along with their transcriptions, is provided in the Appendix.

### 2.1 Image and audio acquisition

Data was acquired using an rtMRI protocol developed specifically for the dynamic study of upper airway movements, especially during speech production (Narayanan et al. 2004). The subject's upper airway was imaged in the midsagittal plane with a fast gradient echo sequence ( $T_R = 6.028$  ms) on a conventional GE Signa 1.5 T scanner, using a custom 4-channel head-

and-neck receiver coil. Data from two front channels was used for image reconstruction.

The scan acquisition region was a 5 mm thick midsagittal slice centered on the subject's tongue, extending over a  $200 \times 200 \text{ mm}^2$  field-of-view. The imaging region was positioned to include the subject's upper trachea, larynx, velum, hard and soft palates, and the lips and jaw, over the full range of excursion of the articulators during speech. Spatial resolution in the sagittal plane was  $68 \times 68$  pixels ( $2.9 \times 2.9 \text{ mm}^2$ ). Image data was acquired with a 9 interleaved spiral readout: a new complete image was acquired every 54 ms, using information from 9 partial acquisitions captured every 6.028 ms ( $T_R$ ). Further details of image acquisition and reconstruction are provided in Bresch et al. (2008).

Audio was simultaneously recorded at a sampling frequency of 20 kHz inside the MRI scanner, using a custom fiber-optic microphone system. Audio recordings were subsequently noise-canceled and reintegrated with the reconstructed MRI video (Bresch et al. 2006). The data provides dynamic visualization, with synchronous audio, of the informant's entire midsagittal vocal tract, from the upper trachea to the lips, including the oropharynx, velum, and nasal cavity. The scan plane was located in the midsagittal plane of the glottis, so that information about abduction and adduction of the vocal folds could also be inferred from pixel intensity in the laryngeal region.

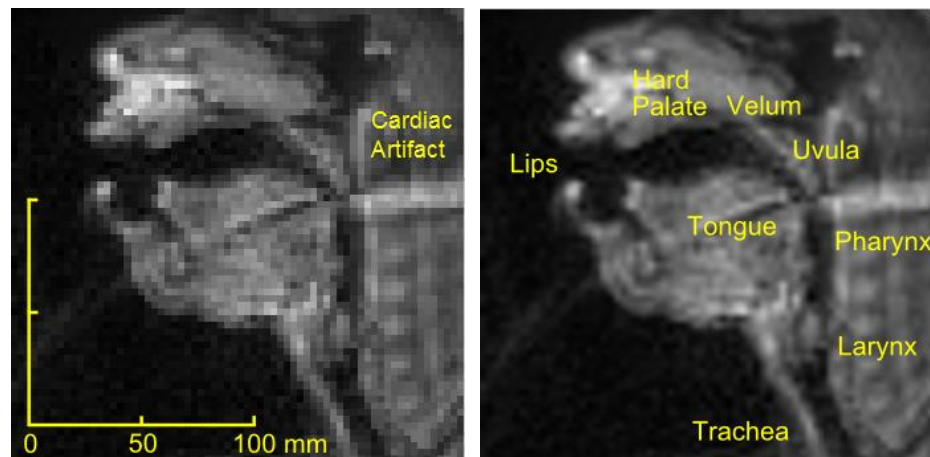


Figure 1. rtMRI frame showing midsagittal articulation of the Nama speaker's upper airway. Left: Original MR image resolution (68 x 68 px); Right: Interpolated image frame (340 x 340 px) with superimposed anatomical labels. Teeth do not image in MRI. An arc-shaped cardiac artefact can be observed passing over the jaw and tongue body, through the upper pharynx. An additional artefact caused by dental reconstruction has affected the resolution of the lower lip, tongue tip, and sublingual cavity.

## 2.2 Data analysis

Images were up-sampled by a factor of five, using bicubic interpolation, from the original image acquisition resolution of 68 x 68 px, to enhance resolution of vocal tract structures and to facilitate estimation of distances between articulators (Figure 1). Image sequences of interest were reconstructed using a sliding-window technique to produce oversampled high-speed video with an effective rate of 165.9 frames per second (one frame every 6.028 ms, including information acquired over 54 ms). Because the image of the upper oral vocal tract was weaker than surrounding areas, the signal from the hard palate was enhanced by superimposing a mean palatal image constructed from all frames in each sequence.

Companion audio and video recordings were synchronized and loaded into a custom graphic user interface for inspection and analysis (Proctor et al. 2010; Narayanan et al. 2014). Start and end times of each utterance were identified by examining the audio signal, spectrogram, and time-aligned video

frames, and the corresponding intervals of each recording were labelled. A typical image sequence, aligned with companion acoustic data, is illustrated in Figure 2. Every 10th frame, spaced at 60.2 ms intervals, selected from the high-speed video reconstruction of the rtMRI data is shown, with corresponding landmarks indicated on the speech waveform.

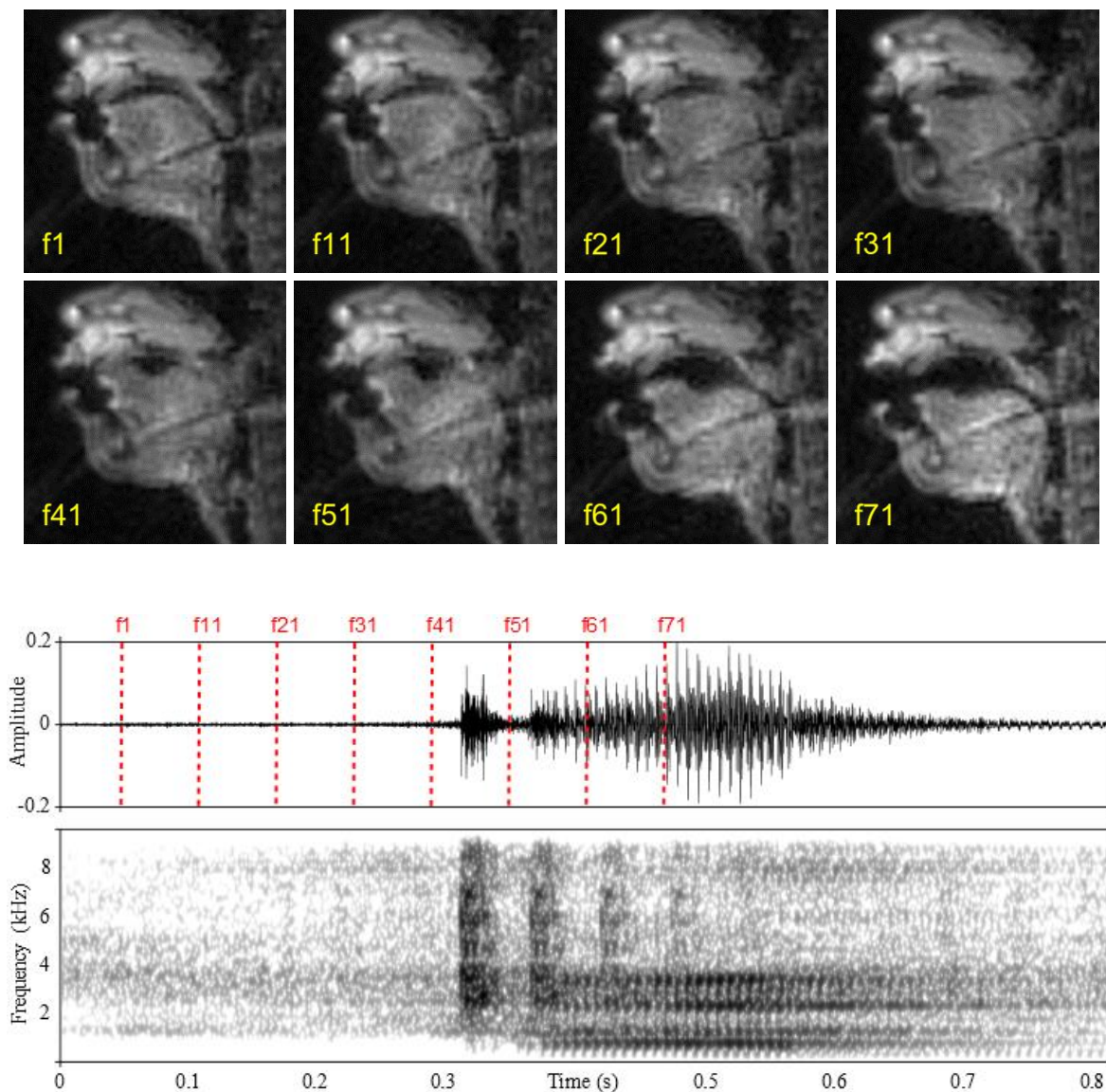


Figure 2. Time-aligned audio and video data acquired during lateral click production. Eight image frames (top two rows) showing midsagittal articulation at key stages of the consonant-vowel sequence beginning the word *//gam /gãm/* ‘talk’. Broken vertical lines superimposed on the waveform of the companion acoustic recording indicate the location in time of each frame. A decaying series of echoes, spaced at 53 ms intervals, can be seen in the waveform and spectrogram (bottom), following the click release burst (at 0.31 s) – an unavoidable artefact of producing a highly transient sound in a cylindrical metallic MRI scanner bore.



### 3. Lingual articulation of clicks

#### 3.1 Dental click production

A sequence of frames captured during articulation of the initial consonant in the word / $\bar{a}$  / $\text{ʔ}^{\text{á}}\text{à}$ / ‘sharp’ is shown in Figure 3.<sup>1</sup> The same characteristic mechanisms of production were observed in all dental click consonants produced by the informant (see Appendix), and are described below with reference to the frame numbers indicated in Figure 3.

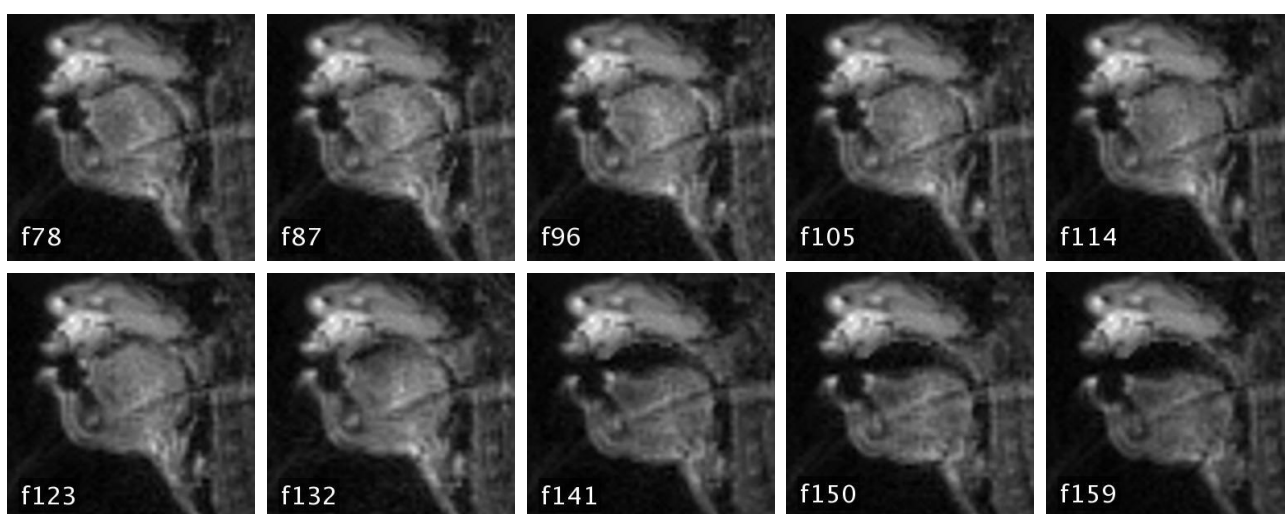


Figure 3. Dental click production: Midsagittal articulation of the onset consonant in / $\bar{a}$  / $\text{ʔ}^{\text{á}}\text{à}$ / ‘sharp’. MRI frames shown at 54 ms intervals. Frame 78: Initial posture (265 ms before release); Frame 123: Lingual posture at acoustic onset of click; Frame 159: Lingual posture at acoustic target of post-consonantal vowel (223 ms after click release).

The data shows that the anterior constriction is primarily apical – formed between the tip of the tongue and the back of the upper teeth – but also involves a seal between the front of the tongue blade and the alveolar ridge (Figure 3, frame 123). Rapid lowering of the tongue blade and the front of the tongue body in the region immediately behind the anterior constriction can be seen as the click is released (frames 114–132). The tongue tip and dorsum

<sup>1</sup> Macrons indicate long vowels, no mid tone, in Khoekhoegowab orthography (Curriculum Committee for Khoekhoegowab 2003).

remain coordinatively raised toward the teeth and uvular region to maintain the lingual seals, before releasing as the body of the tongue lowers into the vowel gesture (frames 141–159). Lingual articulation of the dental click is accompanied by jaw lowering and retraction ( $-8.2$  mm vertical and  $-7.8$  mm horizontal displacement from click release to vowel target).

### 3.2 Alveolar click production

The characteristic patterns of lingual articulation observed in all alveolar clicks produced by the informant are exemplified in Figure 4: A selection of MRI frames captured at key points during production of the initial consonant in the word *!ā /ŋ!ʔàǎ/* ‘hang out’.

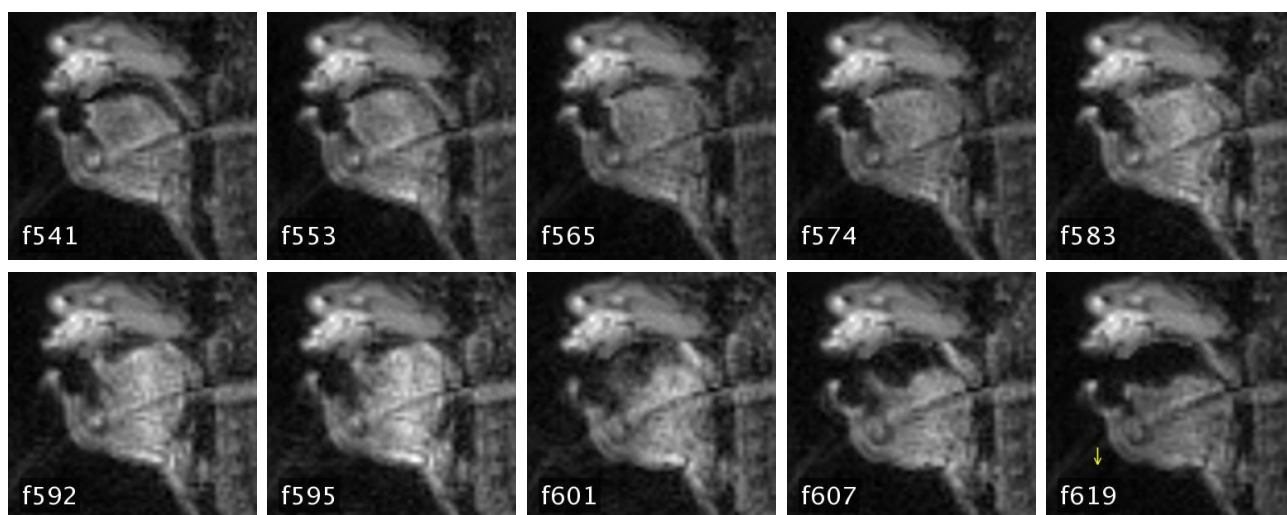


Figure 4. Alveolar click production: Midsagittal articulation of the onset consonant in *!ā /ŋ!ʔàǎ/* ‘hang out’. Frame 541: Initial posture (301 ms before release); Frame 592: Lingual posture at acoustic onset of click; Frame 619: Lingual posture of post-consonantal vowel (169 ms after click release). Arrow indicates total vertical jaw displacement.

The anterior constriction of the alveolar click is formed with a more apical and more retracted tongue tip gesture than that observed in the dental click. The tongue tip first contacts the apex of the alveolar ridge (Figure 4, frame 565) as the tongue body forms a broad constriction extending from the back

of the hard palate to the middle of the velum.<sup>2</sup> The midsagittal area of the cavity formed between the tongue and the hard palate is at a minimum around 170 ms before click release. Rarefaction of the cavity is primarily achieved by lowering and retracting the front part of the tongue (frames 556–607), which assumes a retroflexed posture (Cruttenden 1992) due to the perseverance of both tongue tip and tongue body gestures into the release phase (frames 583–607). Lingual articulation of the alveolar click is accompanied by more jaw lowering (–12.2 mm vertical displacement), but less jaw retraction (–5.8 mm horizontal displacement), from click release to vowel target, than observed in the dental click produced in the same vowel context.

### 3.3 Palatal click production

Palatal click production is illustrated in Figure 5: A sequence of frames captured at key points during articulation of the initial consonant in the word  $\#ā / \text{ᵛᵗ}^{\text{?}}\text{áá} /$  ‘slaughter’. The same general patterns of lingual articulation were observed for all palatal clicks produced by the informant.

---

<sup>2</sup> Resolution of the tongue tip, and possibly the initial phase of coronal closure, has been partially compromised by the dental artefact in this image sequence (Figure 4, frames 541–565).

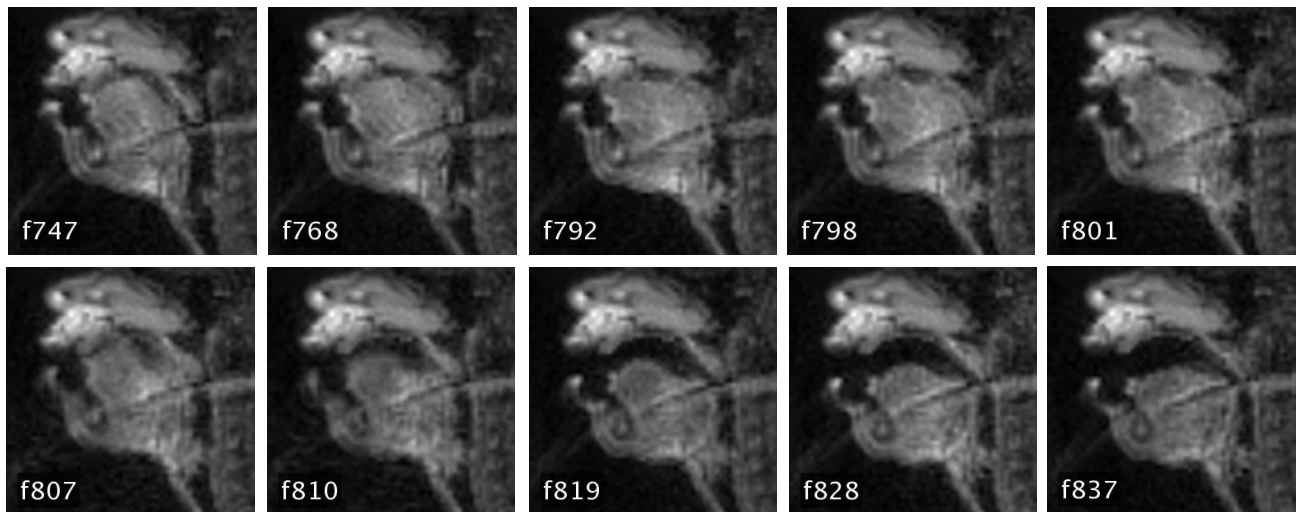


Figure 5. Palatal click production: Midsagittal articulation of the onset consonant in  $\#ā /ʔᵀʰáá/$  ‘slaughter’. Frame 747: Initial posture (326 ms before release); Frame 801: Lingual posture at acoustic onset of click; Frame 837: Lingual posture of post-consonantal vowel (217 ms after click release).

Palatal clicks are initiated with the body of the tongue raising towards the mid-palate. When constrictions are first formed, lingual contact appears to extend across the entire midline of the palate, from the back of the upper incisors to the bottom of the uvula (Figure 5, frame 768). The lingual cavity emerges in the midsagittal plane at the back of the hard palate – the highest point of the oral cavity for this speaker (frame 798) – approximately 43 ms before click release. The lingual cavity is rapidly rarefied by lowering and retracting the tongue body, while the entire tongue tip and blade maintain the anterior constriction, and the tongue body maintains a broad posterior constriction against the velar and uvular regions (frames 798–810). Lingual articulation of the palatal click is accompanied by a similar degree of jaw lowering (–11.6 mm vertical displacement) and jaw retraction (–5.9 mm horizontal displacement), from click release to vowel target, as observed in the alveolar click.

### 3.4 Lateral click production

The characteristic patterns of lingual articulation observed in all lateral clicks produced by the informant (see Appendix) are exemplified in Figure 6: A selection of frames captured at key points during production of the initial consonant in the word  $\|\bar{a} / \text{ʔ}\| \text{ʔ} \text{á} /$  ‘wash’.

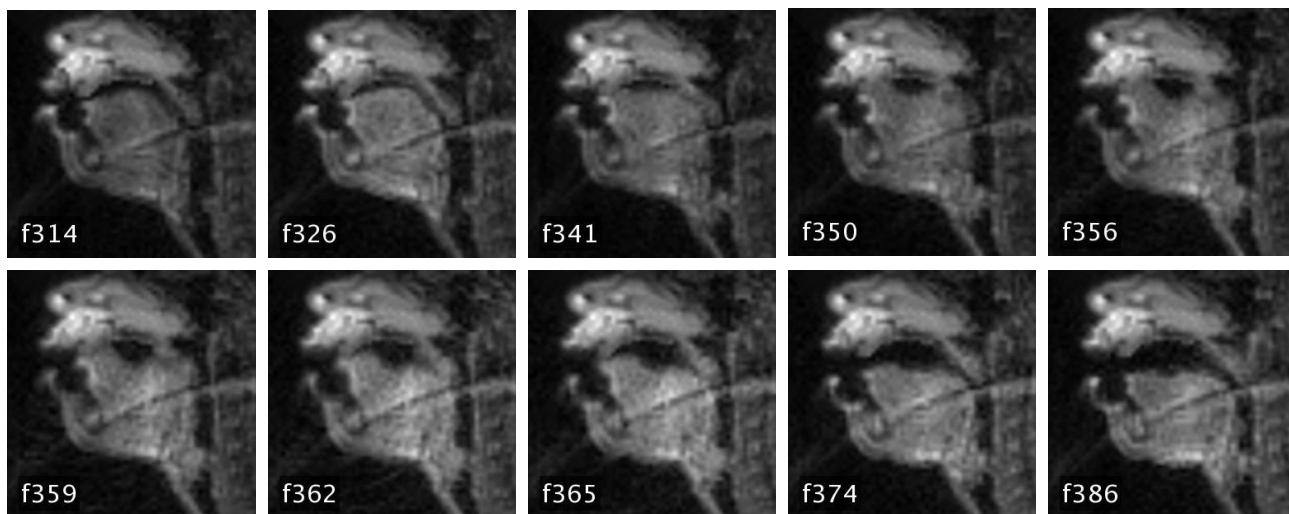


Figure 6. Lateral click production: Midsagittal articulation of the onset consonant in  $\|\bar{a} / \text{ʔ}\| \text{ʔ} \text{á} /$  ‘wash’. Frame 314: Initial posture (217 ms before release); Frame 350: Lingual posture at acoustic onset of click; Frame 386: Lingual posture of post-consonantal vowel (217 ms after click release).

Lateral clicks are initiated with tongue raising, and lingual elongation in the midsagittal plane, as anterior and posterior constrictions are simultaneously formed (Figure 6, frames 314–350). The anterior constriction is formed with a laminal coronal gesture with a post-alveolar target (frame 326); at maximum compression, the seal formed by the tongue-tip extends across the entire alveolar ridge, from the upper incisors to the post-alveolar region (frame 356). The posterior constriction results from a dorsal gesture centered on the uvular region, extending over the full length of the velum (frames 350–362).

A lingual cavity is evident throughout the entire interval of lateral click production, but the location of the cavity in the midsagittal plane shifts

backwards as the click evolves. At initial formation, the cavity is centered on the mid-palate and extends from the alveolar ridge to the back of the hard palate (frames 326–341). During click release, the lingual cavity retracts to a more posterior location, ultimately extending across both hard and soft palates (frames 350–362).

It is not clear from this data whether the click is released to the left or right or bi-laterally, although this could be revealed if additional data was acquired using a coronal imaging plane. Lingual articulation of the lateral click is accompanied by the least amount of jaw retraction of the four places of articulation (–4.6 mm horizontal displacement) and a similar degree of jaw lowering to that observed in dental clicks produced by this speaker (–9.3 mm vertical displacement from click release to low vowel target).

### **3.5 Comparison of lingual articulation**

Differences which characterize clicks produced at each of the four contrastive places of articulation can be better understood by comparing tongue shapes directly: tongue postures captured at and after release of word-initial clicks before the mid-back vowel /o/ are illustrated in Figure 7.

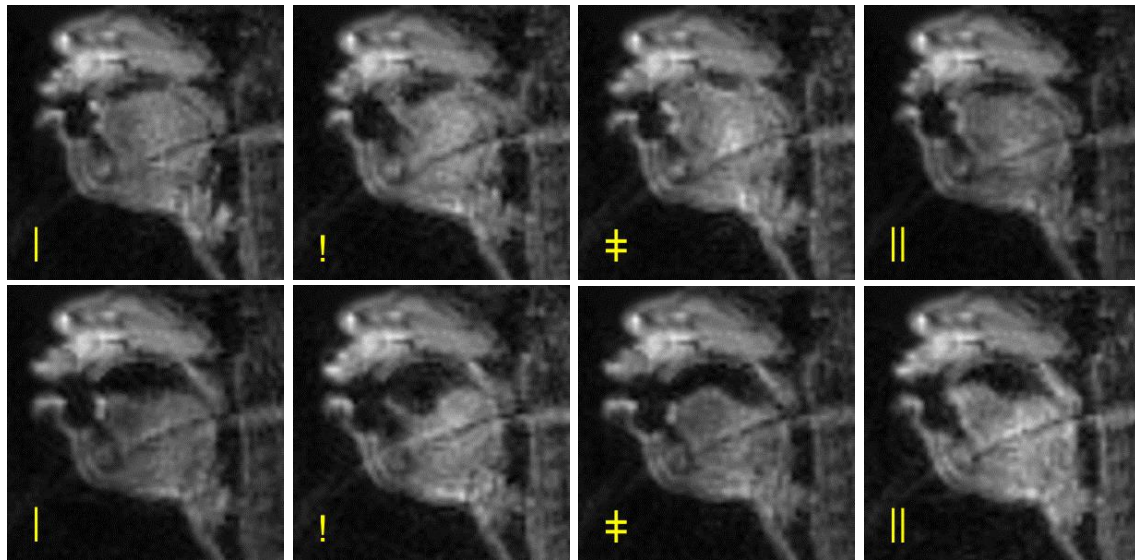


Figure 7. Comparison of word-initial clicks produced before mid-back vowels. Top row: Midsagittal lingual posture at acoustic onset of click; Bottom row: Lingual posture 100 ms after click release. L-to-R: Dental:  $|om / \text{ǀ}^? \text{òm} /$  ‘breathe’, alveolar:  $!om / \text{ǃ}^? \text{ó} \text{m} /$  ‘remove thorn’, palatal:  $\neq om / \text{ǁ}^? \text{ó} \text{m} /$  ‘sew’, lateral:  $\|om / \text{ǁ}^? \text{ò} \text{m} /$  ‘sleep’.

To facilitate direct comparison of click place contrasts, MRI frames capturing lingual posture at key stages of click production were identified and vocal tract boundaries were located using the analysis technique described in Proctor et al. (2010). Tongue outlines at the moment of click release, 100 ms after release, and at the acoustic target of the following vowel were traced, and are superimposed in Figure 8 for comparison. The tongue root was traced beneath the epiglottis in each frame. A palate and pharynx trace captured at a single point in time is superimposed on each image to locate the tongue edges with respect to the passive structures.<sup>3</sup>

To minimize coarticulatory and manner influences, each consonant was produced as a word-initial click, with glottal closure, before the same low vowel:  $|\bar{a} / \text{ǀ}^? \text{á} \text{à} /$  ‘sharp’,  $!\bar{a} / \text{ǃ}^? \text{à} \text{á} /$  ‘hang out’,  $\neq \bar{a} / \text{ǁ}^? \text{á} \text{á} /$  ‘slaughter’, and  $\|\bar{a}$

<sup>3</sup> Because the articulation of the velum varies from frame to frame, the soft palate trace intersects the outline of the tongue dorsum in some figures.

/ʘ||ʔǎǎ/ ‘wash’.<sup>4</sup>



Figure 8. Comparison of midsagittal lingual posture at key stages of click production. Top left: 100 ms before click release; Top right: Moment of click release; Bottom left: 100 ms after click release; Bottom right: Lingual posture at acoustic target of following vowel. Vocal tract outlines extracted from four frames each of utterances of /ā /ʘ||ʔǎǎ/ ‘sharp’ (dental), /ā /ʘ||ʔǎǎ/ ‘hang out’ (alveolar), /ā /ʘ||ʔǎǎ/ ‘slaughter’ (palatal), and /ā /ʘ||ʔǎǎ/ ‘wash’ (lateral).

### 3.6 Articulatory coordination

A major advantage of rtMRI is the information it provides about the behavior of articulators beyond the tongue. A topic of particular interest in Khoekhoe consonant phonology is the way that the velum and glottis are

<sup>4</sup> There may be additional influences of lexical tone on lingual posture. It is beyond the scope of this paper to analyze the interaction of tone and supralaryngeal articulation, but a preliminary analysis suggests that any such effects are largely confined to the laryngeal and lower pharyngeal regions, and therefore do not have a major impact on the patterns of tongue shaping being compared here.



coordinated in clicks differentiated by accompaniment. Vocal tract configurations at the moment of release in four of the five contrastive alveolar click types are compared in Figure 9. Each click appears word-initially, before the same mid-back vowel /o/ carrying high tone.

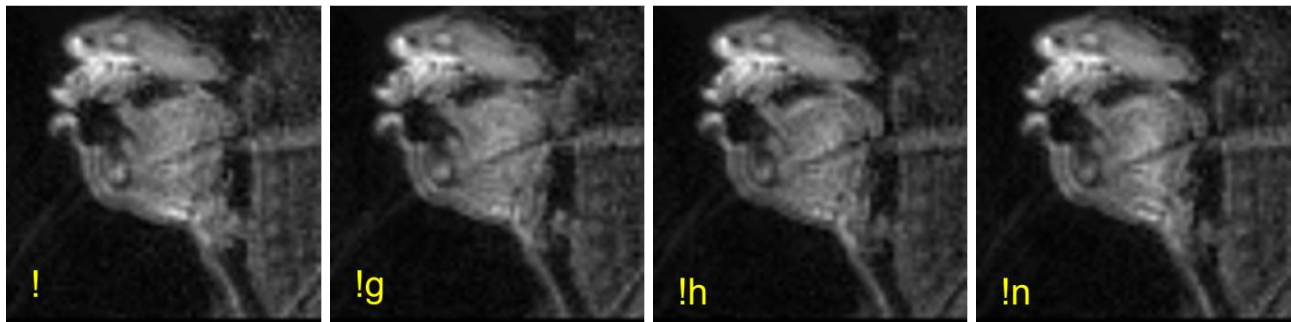


Figure 9. Comparison of word-initial alveolar clicks contrasting in manner. Midsagittal lingual posture captured at acoustic onset of click. L-to-R: Glottalized: *!oas* /ʔ!²òás/ ‘meeting’; voiceless unaspirated: *!goas* /!óás/ ‘hollow’; delayed aspirated: *!hoas* /ʔ!ʰóás/ ‘narrating’; and voiced nasal: *!noras* /ʔ!òrás/ ‘freedom’.

Differences in laryngeal activity and the state of the velum can be observed in the images in Figure 9. At the point of release of the initial consonant in *!oas* ‘meeting’, tissue can be seen in the glottal region, indicating a constriction of the vocal folds or surrounding structures, consistent with previous descriptions of this consonant as glottalized (Ladefoged & Traill 1984, 1994; Güldemann 2001).<sup>5</sup> The data also reveals a partially open nasopharynx, consistent with the phonetic description of this click as [ʔ!²] (Miller et al. 2009a; Brugman 2009).

In contrast to the glottalized click, the initial consonant of *!goas* ‘hollow’ is produced with a raised velum and evidence of glottal abduction – a gap appears immediately below and to the right of the epiglottis, indicating an absence of tissue that is present during voicing or glottalization. Similar

<sup>5</sup> Because the informant was unfamiliar with this word in his variety of Nama (*nūs* is his preferred term for ‘meeting’), it was realized with a different tone contour ([ʔ!²òás]) to the citation Khoekhoegowab form [ʔ!²óás] (Haacke & Eiseb 2002).

glottal and nasopharyngeal states can be observed in the consonant illustrated in Figure 2 (*//gam //ám/* ‘talk’), consistent with the characterization of these clicks as voiceless unaspirated (Ladefoged & Traill 1984; Brugman 2009); however, considerable variability in voice onset time was observed amongst the clicks produced by this speaker with this manner of articulation.

The initial clicks of *!hoas* ‘narrating’ and *!noras* ‘freedom’ both show a lowered velum, consistent with the airflow data presented in Ladefoged and Traill (1984), and the phonetic descriptions of these consonants as aspirated voiceless nasal [ʔ<sup>h</sup>] and nasalized [ʔ!], respectively (Ladefoged & Traill 1994; Miller et al. 2009a; Brugman 2009). Delayed aspirated clicks – such as the initial consonant of *!hoas* ‘narrating’ – have also been described as glottalized (Snyman 1970; Cruttenden 1992), yet the laryngeal constrictions in the delayed aspirated clicks produced by this speaker differ from those observed in the glottalized equivalents (Figures 7 and 9). Further analysis is required to properly understand the phonetic basis of these contrasts, yet this data is consistent with the characterization of delayed aspirated manner of production as [–glottal] (Traill 1992; Ladefoged & Traill 1984).

rtMRI also reveals phonetic detail about the fifth contrastive manner of linguo-velaric consonant production in Nama: voiceless aspirated (‘kh’) clicks. At the moment of release, lingual articulation of these clicks closely resembles that observed in the other four click types (Figure 9), but articulatory differences – especially in the posture and motion of the tongue dorsum – can later be observed, during the release phase. Vocal tract configurations of glottalized and voiceless aspirated clicks produced at each place of articulation, captured 100 ms after release, are compared in Figure 10. Each click appears before the same rime /-om/.

In each of the clicks compared in Figure 10, a different lingual posture can be observed in the voiceless aspirated variant produced at the same place: the tongue dorsum remains in a higher position during click release (bottom row) than for the comparable glottalized click (top row). These differences can be quantified by measuring the coordinates of the pixels defining the apex of the

tongue dorsum in different image frames: the mean location of dorsal apices captured 100 ms after click release is 3.5 mm higher and 1.35 mm more fronted in the voiceless aspirated variants. As a result, the tongue body remains raised towards the center of the uvula during voiceless aspirated click release, while a narrow airway is still evident between the tongue and the uvula. The data suggests that, for this speaker, this manner of production involves a uvular fricative release, consistent with description of these clicks as  $[|\chi]$ – $[!χ]$ – $[≠χ]$ – $[||χ]$  (Miller et al. 2009a; Brugman 2009).

The images in Figure 10 also demonstrate that the velum remains raised throughout the release of the voiceless aspirated clicks (except for the palatal click in  $≠khom$  ‘shave hair’), consistent with airflow data presented in Ladefoged and Traill (1984, 1994) and the description of this manner of articulation as non-nasalized (Güldemann 2001; Miller 2011).

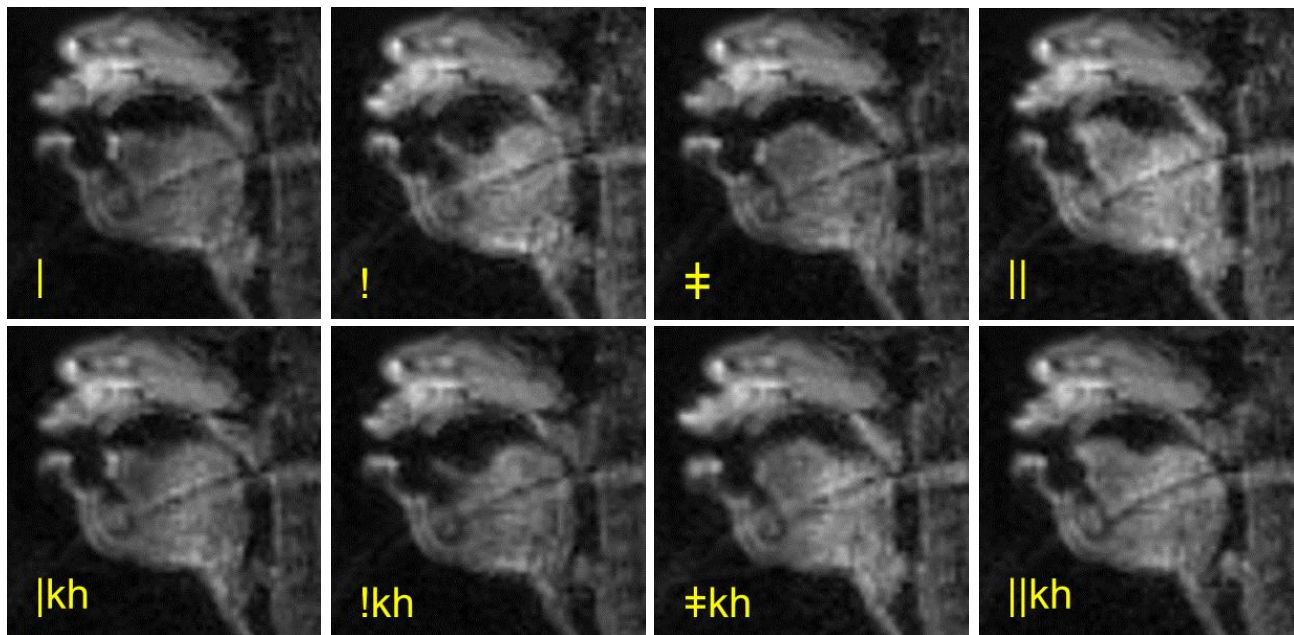


Figure 10. Comparison of glottalized and voiceless aspirated click releases. Top row: Glottalized clicks, 100 ms after release; L-to-R:  $|om$  /  $ʔ|òṁ$ / ‘breathe’,  $!om$  /  $ʔ!òṁ$ / ‘remove thorn’,  $≠om$  /  $ʔ≠òṁ$ / ‘sew’,  $||om$  /  $ʔ||òṁ$ / ‘sleep’. Bottom row: Voiceless aspirated clicks, 100 ms after release. L-to-R:  $|khom$  /  $|χóṁ$ / ‘pity’,  $!khom$  /  $!χòṁ$ / ‘building collapse’,  $≠khom$  /  $≠χòṁ$ / ‘shave hair’,  $||khom$  /  $||χóṁ$ / ‘abdomen’.

#### 4. Discussion

The data introduced here demonstrates the relevance of rtMRI as a new technique for investigating linguo-velaric consonant production. For the first time since Doke's (1923a, 1923b) and Trill's (1985) groundbreaking X-ray studies of clicks, we have a method for visualizing and tracking the dynamic configuration of the *whole* vocal tract. Unlike X-ray imaging, rtMRI allows us to do so for extended periods of time, with no risk to participants.

rtMRI has revealed details about the location and type of the anterior and posterior constrictions used in linguo-velaric consonant production, and the ways that these differ between clicks. Some of this information is readily obtained using existing phonetic methods such as ultrasound and palatography. Other details of click production cannot be obtained using these approaches. In particular, rtMRI provides rich information about global lingual posture and kinematics, and tongue shaping beyond the field of view of ultrasound and EPG.

The rtMRI data for this speaker, for example, reveals that different articulatory actions are used to rarefy the mid-oral cavity in different clicks. Dental clicks are produced with tongue body lowering while a relatively stable laminal coronal constriction is maintained. Alveolar clicks are characterized by a more apical lingual posture: they are released with a rapid coronal articulation while the tongue dorsum remains raised at the posterior constriction. The location of the lingual cavity also varies between clicks; it is centered in the mid-palatal region when lateral clicks are released and at a more posterior location at the same stage of production in palatal clicks.

These findings are broadly consistent with previous observations of rarefaction mechanisms in Mangetti Dune !Xung (Miller et al. 2009b). rtMRI data may contribute to a better understanding of these articulatory actions, because it can track motion of the whole mass of the tongue – including the tongue root – and the way it operates in concert with the jaw. This data, for example, reveals that the dental click is released with a more monolithic lingual posture and coordinated jaw action than the alveolar click, which

involves greater independence of the tongue tip from the tongue body and jaw in the realization of the partially retroflexed release gesture for this speaker.

The exact location and extent of the posterior linguo-velaric seal varied across click types and vowel contexts, and especially in the case of the palatal and lateral clicks, the seal between the dorsum and the velum was found to release at a more posterior place to the location at which it initially forms, as Miller (2008) has observed in post-alveolar clicks in isiXhosa. Overall this data suggests that, for this speaker, the dorsal closure in Khoekhoe is best characterized as uvular, consistent with previous findings from ultrasound studies (Miller et al. 2007b), as well as data on constriction location in N|uu (Miller et al. 2009a) and Mangetti Dune !Xung (Miller et al. 2009b).

Most importantly, because rtMRI tracks the state of the velum and provides some information about laryngeal state, it can reveal phonetic details of manner contrasts between clicks and help to clarify issues concerning their phonological characterization. The glottal and nasal contrasts observed in this data provides further evidence for conclusions previously drawn from airflow data (Ladefoged & Traill 1984, 1994) and are consistent with the phonetic descriptions of manner contrasts proposed in Miller et al. (2009a). Further analysis is required to properly interpret the rich temporal information about velic articulation provided by rtMRI.

#### **4.1 Limitations**

The most obvious limitation of this study is the use of a single speaker. It is not clear to what extent some of the patterns of articulation we describe are speaker-specific. More data will be needed to establish how click production is influenced by vocal tract morphology (Lammert et al. 2013) and prosodic factors (Brugman 2009). This data is also restricted to the mid-sagittal plane. MRI acquisition from coronal, axial, and parasagittal planes, in conjunction with midsagittal, would provide additional information about tongue shaping and vocal tract configuration, which would be especially helpful in the study

of lateral clicks.

The audio signal provided by current rtMRI technology is not ideal for studying clicks. Noise-cancelled in-scanner speech recordings (Bresch et al. 2006) are of sufficient quality for general acoustic analysis (e.g. Lammert et al. 2013) and automatic phonetic transcription (Katsamanis et al. 2011), but due to the nature of the recording environment and the need for extensive signal processing, this data is degraded in ways that currently prevent more extensive analysis. Acoustic recordings in an MRI scanner are characterized by limited bandwidth and echo artefacts that are particularly problematic for highly transient sounds such as clicks, and may need to be supplemented by companion recordings made outside the scanner.

The most practical restriction on rtMRI as a tool for studying clicks is its limited availability. The technology used to acquire the data presented in this study has been especially developed for speech research and is unique to the Magnetic Resonance Engineering Laboratory at the University of Southern California. Currently, only a few research facilities world-wide are equipped with similar capabilities – none, as far as we are aware, in southern Africa. Nevertheless, the increasing availability of new MRI technologies (e.g. Niebergall et al. 2013, Lingala et al. 2015) will create further opportunities to acquire dynamic data from speakers of click languages.

## 4.2 Future directions

More sophisticated methods will be needed to quantify articulation in MRI data in ways that allow for meaningful comparison with data obtained using other modalities. States of the velum and the nasopharynx can be tracked in more detail, using both manual (Proctor et al. 2013) and automatic (Lammert et al. 2010) methods, to shed more light on click nasalization. Velic activity should also be observed in a greater range of phonological environments to examine exactly how it is timed before and after nasalized and non-nasalized clicks. These insights will inform larger debates about phonological compositionality.

More detailed analysis is needed to properly interpret articulation in and around the larynx during click production. The midsagittal plane is carefully aligned during scan localization, so that it intersects the subject's larynx as close to the midline as possible. Despite this, it is not always clear exactly which anatomical structures are involved when tissue appears in the glottal region of an MR image, due to anatomical differences between subjects and the limits of scan resolution. Further analysis may help determine the extent to which structures other than the vocal folds – such as aryepiglottic folds and ventricular folds (Esling & Harris 2005; Edmondson & Esling 2006) – are involved in glottal contrasts in linguo-velaric consonants. Yet again, there may be different implications for phonological models, depending on the exact nature of the constrictions formed (Esling 2005; Moisik & Esling 2011).

It remains to be seen what role rtMRI might play in further investigations of Khoekhoe and related languages, but we are optimistic that it may continue to shed new light on mechanisms of production of lingual consonants, the phonology of Khoisan languages, and our understanding of non-pulmonic consonants in general.

## **5. Acknowledgments**

This work was supported by National Institutes of Health grant R01 DC007124. Special thanks to Doug Whalen and an anonymous reviewer for comments on an earlier draft of this manuscript, and for feedback from participants at the 5th International Symposium on Khoisan Languages and Linguistics.

## Appendix

Nama	IPA	Afrikaans	English
/ā	/ʔ <sup>2</sup> áà/	skerp	sharp
//ā	/ʔ <sup>2</sup> ǃá/	was	wash
!ā	/ʔ <sup>2</sup> ǎá/	ophang	hang out
ǂā	/ʔ <sup>2</sup> ǃá/	slag	slaughter
/om	/ʔ <sup>2</sup> òṁ/	asemhaal	breathe
//om	/ʔ <sup>2</sup> òṁ/	slaap	sleep
!om	/ʔ <sup>2</sup> óṁ/	doring uithaal	remove thorn
ǂom	/ʔ <sup>2</sup> óṁ/	werk met naald	sew
/gam	/ǃám/	warm	warm
//gam	/ǃám/	gesels	talk
!gam	/ǎám/	diep	deep
ǂgam	/ǃám/	van blydschap rondspring	jumping for joy
/khom	/ǃóṁ/	jammer kry	pity
//khom	/ǃóṁ/	maag of pens	abdomen, belly
!khom	/ǎóṁ/	gebou omval	building collapse
ǂkhom	/ǃóṁ/	hare sker	shave hair
/gôas	/ǃóás/	dogter	daughter
//gôab	/ǃóáb/	brul padda	bullfrog
!gôab	/ǎóáb/	getal	number
ǂgôas	/ǃóás/	bok/skaap wat nie kan lam nie	goat/sheep that can't lamb
/hôas	/ʔ <sup>h</sup> ǃóás/	kat	cat
//hôas	/ʔ <sup>h</sup> ǃóás/	hoek	corner
!hôas	/ʔ <sup>h</sup> ǃóás/	kierie	walking stick
ǂhôas	/ʔ <sup>h</sup> ǃóás/	nuus of tyding	news or tiding
!oas	/ʔ <sup>2</sup> óás/	(unfamiliar to informant)	meeting
!goas	/ǃóás/		hollow
!hoas	/ʔ <sup>h</sup> ǃóás/		narrating
!noras	/ʔ <sup>h</sup> ǃóás/		freedom

Table 2. Nama elicitation items: Nama orthography and phonological transcriptions, with Afrikaans disambiguators (where used during presentation), and English glosses.



## References

- Beach, D. M. (1938) *The Phonetics of the Hottentot Language*. Cambridge: Heffer.
- Bradfield, J. (2014) Clicks, concurrency and Khoisan. *Phonology* 31: 1–49.
- Bresch, E., Y.-C. Kim, K. Nayak, D. Byrd & S. Narayanan (2008) Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]. *Signal Processing Magazine, IEEE* 25,3: 123–132.
- Bresch, E., J. Nielsen, K. Nayak & S. Narayanan (2006) Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *The Journal of the Acoustical Society of America* 120,4: 1791–1794.
- Brugman, J. (2009) Segments, tones and distribution in Khoekhoe prosody. Ph.D. dissertation: Cornell University.
- Cruttenden, A. (1992) Clicks and syllables in the phonology of Dama. *Lingua* 86,2-3: 101–117.
- Curriculum Committee for Khoekhoegowab (2003) *Khoekhoegowab Orthography 3 / 3//t Xoa/gaub*. Windhoek: Gamsberg Macmillan.
- Doke, C. M. (1923a) A dissertation on the phonetics of the Zulu language. *Bulletin of the School of Oriental and African Studies* 2,4: 685–729.
- \_\_\_ (1923b) Notes on a problem in the mechanism of the Zulu clicks. *Bantu Studies* 2,1: 43–45.
- \_\_\_ (1925) The phonetics of !Xũ Bushman. *Bantu Studies* 2,3: 129–165.
- Edmondson, J. A. & J. H. Esling (2006) The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology* 23: 157–191.
- Esling, J. H. (2005) There are no back vowels: The laryngeal articulator model. *The Canadian Journal of Linguistics/La Revue Canadienne de Linguistique* 50,1: 13–44.

- Esling, J. H. & J. G. Harris (2005) States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In: W. J. Hardcastle & J. M. Beck (eds.) *A Figure of Speech: A Festschrift for John Laver*. Mahwah (NJ): Lawrence Erlbaum, 347–383.
- Fredericks, N. J. (2013) A study of dialectal and inter-linguistic variations of Khoekhoegowab: Towards the determination of the standard orthography. Ph.D. dissertation: University of the Western Cape.
- Güldemann, T. (2001) Phonological regularities of consonant systems across Khoisan lineages. (University of Leipzig Papers on Africa, Languages and Literatures 16). Leipzig: Institute for African Language Studies, University of Leipzig.
- Güldemann, T. & R. Vossen (2000) Khoisan. In: B. Heine & D. Nurse (eds.) *African Languages: An Introduction*. Cambridge: Cambridge University Press, 99–122.
- Haacke, W. H. G. & E. Eiseb (2002) *A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index*. Windhoek: Gamsberg Macmillan.
- Haacke, W. H. G., E. Eiseb & L. Namaseb (1997) Internal and external relations of Khoekhoe dialects: A preliminary survey. In: W. H. G. Haacke & E. D. Elderkin (eds.) *Namibian Languages: Reports and Papers*. (Namibian African Studies 4). Cologne: Rüdiger Köppe, 125–209.
- Katsamanis, A., M. Black, P. Georgiou, L. Goldstein & S. Narayanan (2011) SailAlign: Robust long speech-text alignment. *Proceedings of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*. Pennsylvania: University of Pennsylvania, 44–47.
- Köhler, O., P. Ladefoged, J. Snyman, A. Traill & R. Vossen (1988) The symbols for clicks. *Journal of the International Phonetic Association* 18,2: 140–142.
- Ladefoged, P. & A. Traill (1984) Linguistic phonetic description of clicks. *Language* 60,1: 1–20.

- \_\_\_ (1994) Clicks and their accompaniments. *Journal of Phonetics* 22: 33–64.
- Lammert, A., M. I. Proctor & S. S. Narayanan (2010) Data-driven analysis of realtime vocal tract MRI using correlated image regions. In: T. Kobayashi, K. Hirose & S. Nakamura (eds.) *Proceedings of the 11<sup>th</sup> Annual Conference of the International Speech Communication Association (Interspeech 2010)*. Makuhari (Japan), ISCA: 1572–1575.
- Lammert, A., M. Proctor & S. Narayanan (2013) Interspeaker variability in hard palate morphology and vowel production. *Journal of Speech, Language, and Hearing Research* 56,6: S1924–S1933.
- Lingala, S. G., Y. Zhu, Y.-C. Kim, A. Toutios, S. S. Narayanan & K. S. Nayak (2015) High-frame-rate real-time imaging of speech production. SPIE News Room. Available online: <http://spie.org/x113691.xml>
- Miller, A. L. (2008) Click cavity formation and dissolution in IsiXhosa: Viewing clicks with high-speed ultrasound. In: R. Sock, S. Fuchs & Y. Laprie (eds.) *Proceedings of the 8<sup>th</sup> International Seminar on Speech Production*. Strasbourg: INRIA, 137–140.
- \_\_\_ (2011) The representation of clicks. In: M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.) *The Blackwell Companion to Phonology*, Volume 1. Malden (MA): Wiley-Blackwell, 416–439.
- Miller, A. L., J. Brugman & B. E. Sands (2007a) Acoustic and auditory analysis of N|uu lingual and linguo-pulmonic stop bursts. In: J. Trouvain & W. J. Barry (eds.) *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences*. Saarbrücken, ICPHS: 769–772.
- Miller, A. L., L. Namaseb & K. Iskarous (2007b) Tongue body constriction differences in click types. In: J. S. Cole & J. I. Hualde (eds.) *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 643–656.
- Miller, A. L., J. Brugman, B. Sands, L. Namaseb, M. Exter & C. Collins (2009a) Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association* 39,2: 129–161.

- Miller, A., A. Scott, B. E. Sands & S. Shah (2009b) Rarefaction gestures and coarticulation in Mangetti Dune !Xung clicks. In: M. Uther, R. Moore & S. Cox (eds.) *Proceedings of the 10<sup>th</sup> Annual Conference of the International Speech Communication Association (Interspeech 2009)*. Brighton: ISCA, 2279–2282.
- Miller-Ockhuizen, A. (2003) *The Phonetics and Phonology of Gutturals: A Case Study from Ju/'hoansi*. New York: Routledge.
- Moisik, S. & J. H. Esling (2011) The 'whole larynx' approach to laryngeal features. In: W.-S. Lee & E. Zee (eds.) *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 1406–1409.
- Nakagawa, H. (1996) A first report on the click accompaniments of |Gui. *Journal of the International Phonetic Association* 26,1: 41–54.
- \_\_\_ (2006) Aspects of the phonetic and phonological structure of the G|ui language. Ph.D. dissertation: University of the Witwatersrand.
- Narayanan, S., K. Nayak, S. Lee, A. Sethy & D. Byrd (2004) An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America* 115,4: 1771–1776.
- Narayanan, S., A. Toutios, V. Ramanarayanan, A. Lammert, J. Kim, S. Lee, K. Nayak, Y.-C. Kim, Y. Zhu, L. Goldstein, D. Byrd, E. Bresch, P. Ghosh, A. Katsamanis & M. Proctor (2014) Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *Journal of the Acoustical Society of America* 136,3: 1307–1311.
- Niebergall, A., S. Zhang, E. Kunay, G. Keydana, M. Job, M. Uecker & J. Frahm (2013) Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. *Magnetic Resonance in Medicine* 69,2: 477–485.
- Proctor, M. I., D. Bone & S. S. Narayanan (2010) Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal

- tract analysis. In: T. Kobayashi, K. Hirose & S. Nakamura (eds.) *Proceedings of the 11<sup>th</sup> Annual Conference of the International Speech Communication Association (Interspeech 2010)*. Makuhari (Japan): ISCA, 1576–1579.
- Proctor, M., L. Goldstein, A. Lammert, D. Byrd, A. Toutios & S. S. Narayanan (2013) Velic coordination in French nasals: A real-time magnetic resonance imaging study. In: F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino & P. Perrier (eds.) *Proceedings of the International Conference on Speech Communication and Technology*. Lyon: International Speech Communication Association, 577–581.
- Sands, B. (1991) Evidence for click features: Acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 80: 6–37.
- Sands, B. E., J. Brugman, M. Exter, L. Namaseb & A. L. Miller (2007) Articulatory characteristics of anterior click closures in N|uu. In: J. Trouvain & W. J. Barry (eds.) *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences*. Saarbrücken: ICPhS, 401–404.
- Snyman, J. W. (1970) *An Introduction to the !Xū (!Kung) Language*. Cape Town: A. A. Balkema.
- Truill, A. (1985) *Phonetic and Ohonological Studies of !Xõo Bushman*. (Research in Khoisan Studies 1). Hamburg: Helmut Buske.
- \_\_\_ (1991) Pulmonic control, nasal venting, and aspiration in Khoisan languages. *Journal of the International Phonetic Association* 21: 13–18.
- \_\_\_ (1992) A confusion of sounds: The phonetic description of !Xū clicks. In: D. F. Gowlett (ed.) *African Linguistic Contributions*. Pretoria: Via Afrika, 345–362.
- \_\_\_ (1993) The feature geometry of clicks. In: P. M. S. von Steuden (ed.) *Linguistica: Festschrift E.B. van Wyk: 'n Huldeblyk*. Pretoria: J. L. van Schaik, 134–140.

Westphal, E. O. J. (1971) The click languages of southern and eastern Africa. In: J. Berry & J. H. Greenberg (eds.) *Linguistics in Sub-Saharan Africa*. (Current Trends in Linguistics 7). Berlin: Mouton de Gruyter, 367–420.