Click Consonants

Edited by

Bonny Sands

BRILL

LEIDEN | BOSTON

Contents

CHAPTER 6

Studying Clicks Using Real-Time MRI

Michael Proctor, Yinghua Zhu, Adam Lammert, Asterios Toutios, Bonny Sands, and Shrikanth Narayanan

1 Introduction

The phonetic properties of clicks have been investigated in a wide range of languages, using a variety of techniques. Aerodynamic and acoustic properties of lingual consonants[1] have been examined in Zulu (Doke 1923a), Naro (Kagaya 1978), Khoekhoe (Ladefoged and Traill 1984), !Xóõ (Traill 1985; Traill 1991; Ladefoged and Traill 1994), Xhosa (Sands 1991), G|ui (Nakagawa 1996; Nakagawa 2006), Hadza (Sands et al. 1996), N|uu (Miller et al. 2007a), and Yeyi (Fulop et al. 2003). X-ray (Doke 1923b; Doke 1925; Traill 1985), palatography and linguography (Doke 1923a; Doke 1925; Beach 1938; Traill 1985; Nakagawa 2006; Sands et al. 1996; Sands et al. 2007), and more recently, high-speed ultrasound studies (Miller et al. 2007b; Miller et al. 2009b; Miller et al. 2009a, etc.), have provided rich detail about the lingual articulation of clicks in these and other languages.

Collectively, these data have provided a good understanding of the principal mechanisms of click consonant production; however, because each these experimental methods focuses on specific aspects of lingual articulation, there are other aspects of click production that are not as well understood. Because ultrasound and palatography do not provide information about the whole of the soft palate, it not clear whether the dorsal constrictions of some clicks are best classified as velar or uvular, and to what extent – if at all – posterior place of articulation is contrastive in clicks. Neither ultrasound nor palatography provide any information about the pharynx or tongue root, so we have little data on the role of these articulators in click production.

More generally, it is not yet clear if lingual ingressive consonants in all languages can be adequately described in terms of small set of basic underlying clicks and combining accompaniments, or if a complete description of these

---

[1] 'Lingual' refers to the airstream mechanism (c.f. 'glottalic' and 'pulmonic'), and is preferred by some authors (Taljaard and Snyman 1989) over the terms 'velaric' (Ladefoged and Traill 1984; Ladefoged and Traill 1994), or 'linguo-velaric', for reasons outlined in Miller et al. (2007b) and Miller (2011).

| DOI:10.1163/9789004424357_007

sounds requires reference to other details of articulation of the tongue body, tongue root, jaw, larynx, and velum. Cross-linguistic surveys (e.g. Köhler et al. 1988; Ladefoged and Maddieson 1996; Vossen 1997; Maddieson 2003) propose that there are five fundamental click contrasts – bilabial, dental, alveolar, palatal, and lateral – yet it is not clear whether this is an adequate inventory, or if clicks sharing the same anterior constriction location are produced in exactly the same way across languages.

Consonants represented as /!/, for example, have been variously described as alveolar (Sands et al. 1996; Vossen 1997; Ladefoged and Maddieson 1996), postalveolar (Maddieson 2003), palatal (Lanham 1964; Snyman 1975; Herbert 1990; Roux 2007), palato-alveolar (Doke 1954), alveo-palatal (Doke and Mofokeng 1957), apicolamino-palatal (Finlayson et al. 1991), and retroflex (Doke 1926; Cruttenden 1992). Some of this descriptive variation clearly reflects differing analyses of the same sounds; in other cases, different sounds, individual speaker variation, and dialect differences are being described. For example, there is considerable variation in the realization of lingual ingressive consonants in Bantu languages which use only a single click place contrast, even when these are considered to be allophones of the same phonological unit (Ziervogel 1952; Doke and Mofokeng 1957; Maddieson 2003).

Our incomplete knowledge of the phonetic properties of clicks has important implications for their phonological representation – a topic of ongoing debate (Kohler et al. 1988; Nakagawa 2006; Brugman 2009; Miller et al. 2009a; Miller 2011; Bradfield 2014). Views differ over whether clicks are best described as a small set of underlying segments which combine in clusters (Traill 1985; Traill 1993; Nakagawa 2006), or as a larger inventory of unitary segments which resist decomposition (Beach 1938; Snyman 1970; Ladefoged and Traill 1994; Miller-Ockhuizen 2003).

Because articulatory data on clicks has largely 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 these consonants remain. A comprehensive understanding of their phonetic and phonological properties will require more data from more languages: spatially-detailed and temporally-rich information about the configuration of the entire vocal tract. While data of this nature has previously been provided by x-ray studies (Doke 1923b; Traill 1985), this modality is no longer generally considered viable because of concerns about the risks of exposing subjects to ionizing radiation.

In this chapter, we report results from some initial investigations into the use of real-time Magnetic Resonance Imaging (rtMRI) to study clicks. The primary goal of this work is to investigate the utility of rtMRI as a method for studying lingual ingressive consonant production. We present data acquired from individual speakers of two different languages that make use of click consonants: Khoekhoegowab and siSwati. We compare these consonants with data acquired from a beatbox artist who makes paralinguistic use of clicks in his vocal percussion repertoire. In each case study, we discuss the advantages and disadvantages of rtMRI compared to other phonetic methods used to study clicks, and report some findings that will inform our understanding of the phonology of the languages under investigation, and our knowledge of click production in general.

2 Real-Time MRI of Speech

All data presented in this chapter were acquired using a real-time Magnetic Resonance Imaging protocol developed specifically for the dynamic study of the upper airway during speech production (Narayanan et al. 2004; Bresch et al. 2008). Video and audio recordings of the target languages were acquired from the participants during a series of scan sessions conducted at LA County Hospital. Stimuli were presented to the participants in the orthography of the study languages as they lay supine in an MRI scanner bore. Participants' vocal tracts were imaged as they produced spontaneous speech, short passages of prose, and lists of words targeting the consonants of interest. Subjects were paid for their participation.

2.1 *Image and Audio Acquisition*

Subjects' upper airways were imaged in the midsagittal plane on a GE Signa Excite HD 1.5T scanner with gradients capable of 40mT/m amplitude and 150$mT/m/ms$ slew rate, using a custom 4-channel head-and-neck receiver coil. Data from two front channels were used for image reconstruction. Image data were acquired from a 5mm midsagittal slice centered on the subject's tongue, extending over a 200mm × 200mm field-of-view. The imaging region was positioned to include the subject's trachea, larynx, velum, hard and soft palates, and the lips and jaw, over the full range of excursion of the articulators during speech. Image data were acquired with a fast gradient echo sequence (T_R = 6.028ms) and an interleaved spiral readout: a new complete image was acquired every 54ms, using information from 9 partial acquisitions captured every 6.028ms. Spatial resolution in the sagittal plane was 68 × 68 pixels (2.9 × 2.9mm^2). 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 while the subject was imaged, using a custom fiber-optic

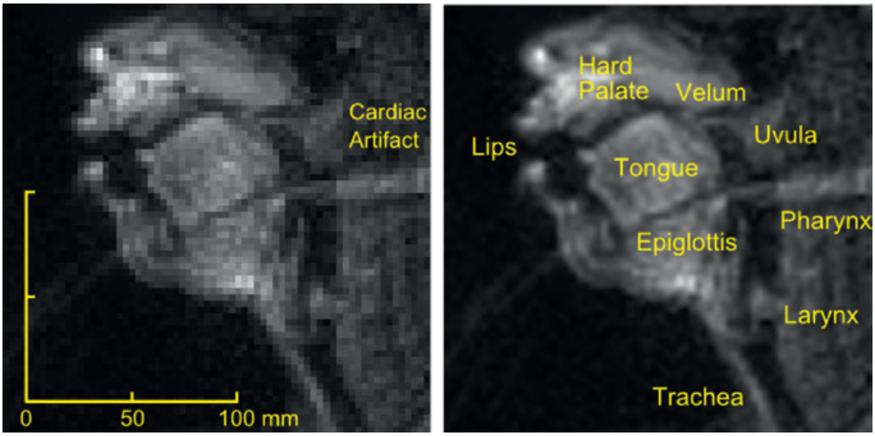

FIGURE 6.1 rtMRI frame showing midsagittal articulation of the upper airway
Notes: Image captured at maximum constriction of the initial dorsal stop in the Nama word *ge* /gè/ 'we (masc.)'. Left: original MR image resolution (68 × 68 px); Right: interpolated image frame (340 × 340 px) with superimposed anatomical labels. Teeth do not image in MRI. An arc-shaped artifact caused by co-planar cardiac activity partially distorts the image through the upper pharynx and tongue root. An additional artifact caused by dental fillings affects resolution of the lower lip, tongue tip, and sublingual cavity.

microphone system. Audio recordings were subsequently noise-canceled and reintegrated with the reconstructed MRI image sequences (Bresch et al. 2006). The resulting video data provide 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. Teeth do not image in MRI, so the location of dentition must be inferred when in contact with soft tissue. Scan planes were located midsagittally through the larynx, so that information about the state of the vocal folds and surrounding structures could also be inferred from pixel intensity in the laryngeal region.

2.2 *Data Analysis*

Images were up-sampled by a factor of five, using bicubic interpolation, from the original image acquisition resolution of 68 × 68 pixels, to enhance resolution of vocal tract structures and to facilitate estimation of distances between articulators (Fig. 6.1). Image sequences were reconstructed post-acquisition 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, constructed from 9 partial acquisitions over 54 ms), so that images could be better time-aligned with acoustic events of interest. Midsagittal palatal outlines for each speaker derived from mean images calculated across multiple frames, and superimposed into image sequences to improve resolution of passive structures around the hard palate, where the relative lack of soft tissue provides a weaker and more variable signal.

Start and end times of each utterance were located by examining time-aligned audio, spectral, and video data in a custom inspection and analysis tool (Proctor et al. 2010a; Narayanan et al. 2014). A typical image sequence, aligned with companion acoustic data, is illustrated in Figure 6.2. 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. Cardiac artifacts – distortions which can arise when the imaging plane intersects the heart – can be seen in Figure 6.2 and other images in these datasets. Because this noise source is localized, it does not usually affect the analysis, although it can compromise study of nasalization when it distorts the region of the image corresponding to the velum and nasopharynx.

3 rtMRI Insights into Nama Clicks

Nama is a variety of Khoekhoegowab, a Khoe language primarily spoken in Namibia (ISO 639–3 code: **naq**), whose other varieties include the ethnolects Damara and Hai||om (Haacke et al. 1997; Güldemann and Vossen 2000). Twenty of the thirty four Nama consonant phonemes use a lingual airstream: each of four different clicks may be produced with five contrastive manners of articulation (Brugman 2009). The four click types are described as dental, alveolar, palatal and lateral (Beach 1938; Westphal 1971; Ladefoged and Traill 1984; Ladefoged and Traill 1994; Güldemann 2001; Haacke and Eiseb 2002), but the phonetic and phonological characterization of the other contrasts are debated. Ladefoged and Traill (1984) and Ladefoged and Traill (1994) describe a system of basic clicks modified by different combinations of laryngeal and nasal accompaniments (Table 6.1). Nama lingual ingressive consonants have also been analyzed as consonant clusters (Traill 1993; Güldemann 2001), unitary segments differentiated by contour airstreams (Miller 2011), and separate series of stops, affricates and nasals (Brugman 2009).

3.1 *Informant and Corpora*

The study participant was a 35 year-old Namibian male trilingual speaker of Afrikaans, Nama, and English. Born and raised in Windhoek, he has lived most of his life in Namibia, and had been living in the United States for a year before

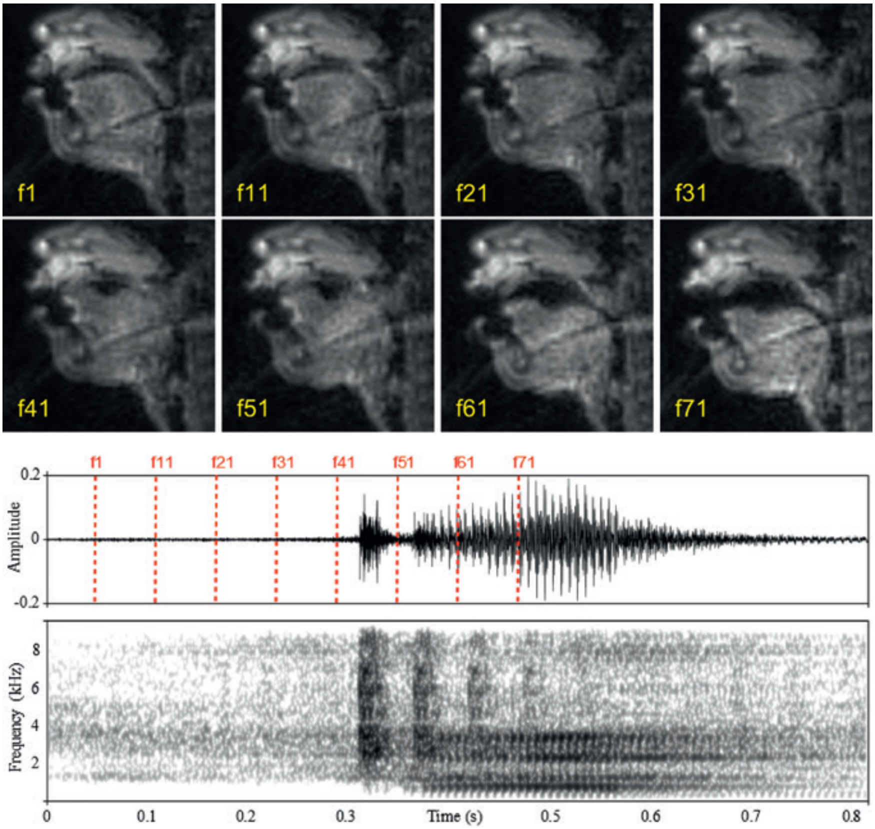

FIGURE 6.2 Time-aligned audio and video data acquired during Khoekhoe lateral click production

Notes: Eight image frames (top two rows) showing midsagittal articulation at key stages of the consonant-vowel sequence beginning the word *//gam* /ǁám̀/ '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 (an artefact of the noise reduction processing) can be seen in the waveform and spectrogram (bottom), following the click release burst at 0.31 s.

participating in this study. His mother speaks Nama as her first language and his father, Afrikaans. The informant lay supine in an MRI scanner and read out words presented in Nama orthography, with accompanying Afrikaans and/or English translations. A list of the subset of words analyzed for this study is provided in Appendix A; most items were elicited twice.

A comprehensive analysis of these data is beyond the scope of this chapter. Initial findings have been presented in Proctor et al. (2016); here we highlight

TABLE 6.1 Nama click consonants

	Glottal closure	**Voiceless unaspirated**	**Voiceless aspirated**	**Delayed aspiration**	**Voiced nasal**
Dental	ǀ	ǀg	ǀkh	ǀh	ǀn
Alveolar	!	!g	!kh	!h	!n
Palatal	ǂ	ǂg	ǂkh	ǂh	ǂn
Lateral	ǁ	ǁg	ǁkh	ǁh	ǁn

Notes: Orthographic representations (Curriculum Committee for Khoekhoegowab 2003) and phonetic descriptions (Ladefoged and Traill 1984; Ladefoged and Traill 1994) of the twenty click phonemes of Nama.

some of the most important ways in which real-time MRI can inform our understanding of the phonetic properties of Nama clicks.

3.2 *Tongue Shaping and Movement in Nama Clicks*

Real-time MRI reveals the shape of the whole tongue, and the way that the tongue is coordinated with the jaw, at each point of an utterance. By examining tongue shape at comparable stages in minimally-different words, we can see how the four basic types of lingual ingressive consonant are formed and released by a speaker of Nama. Midsagittal tongue postures captured at the acoustic onset, and 100 ms into the release of word-initial clicks are illustrated in Figure 6.3.

The data reveal four characteristic midsagittal tongue postures and patterns of release associated with each of the Nama clicks. Dental and alveolar clicks are produced with a more apical coronal gesture, compared to the palatal and lateral clicks, which involve a more extended constriction of the tongue blade across the alveolar ridge and back into the hard palate. The alveolar click is produced by this speaker with a particularly apical, slightly retroflexed tongue posture that is retained throughout release and into the following vowel.

Cavity rarefaction mechanisms and release kinematics associated with each type of click are better understood by tracking changes in lingual posture in greater temporal and spatial detail. Image frames capturing midsagittal vocal tract configurations at the moment of click release, and at 18 millisecond intervals thereafter (every third frame in the sequence) were examined in word-initial clicks produced before the same long, low vowel: *ǀā* /ŋ̊ǁˀáà/ ‘sharp’, *!ā* /ŋ̊!ˀàa̋ / ‘hang’, *ǂā* /ŋ̊ǂˀa̋á/ ‘slaughter’, and *ǁā* /ŋ̊ǁˀa̋á/ ‘wash’. Vocal tract

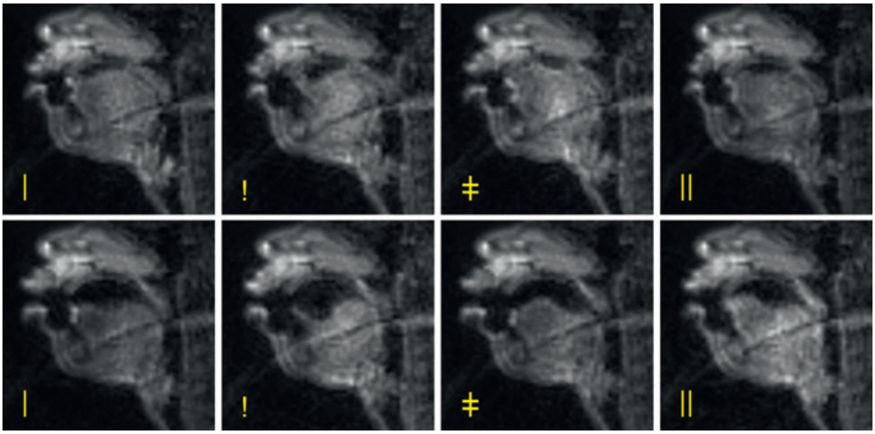

FIGURE 6.3 Comparison of word-initial clicks produced before mid-back vowels
Notes: Top row: midsagittal lingual posture at acoustic onset of click; Bottom row: lingual posture 100 ms after click release. L-to-R: dental: *ǀom* /ŋ̥ǀˀòm̀/ 'breathe', alveolar: *ǃom* /ŋ̥ǃˀóm̀/ 'remove thorn', palatal: *ǂom* /ŋ̥ǂˀóm̀/ 'sew', lateral: *ǁom* /ŋ̥ǁˀòm̋/ 'sleep'.

boundaries were located using the method described in Proctor et al. (2010a), and are superimposed in Fig. 6.4 to show the tongue trajectories associated with each type of release.[2]

The data reveal in more detail the apical nature of the coronal gestures which characterize dental and alveolar click release (Fig. 6.4, top row), and how these differ from the more laminal anterior constrictions used in the palatal and lateral clicks (Fig. 6.4, bottom row). The retroflexed lingual posture of /ǃ/ is even more apparent in this vowel context. The anterior place of articulation of the alveolar click is quite retracted, consistent with some previous descriptions of Khoekhoe /ǃ/ as postalveolar (Miller et al. 2007b), and alveolar/postalveolar Beach (1938: 81).

Greater differences in dorsal articulation are also apparent in the Nama clicks produced before a low vowel. The posterior constriction location at formation and during the initial release phase of the dental click is clearly characterized as velar in these tokens (Fig. 6.4, top left); in the other three clicks produced in this vowel context, part of the dorsal seal is created and released

---

[2] 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. The tongue root was traced beneath the epiglottis in each frame. Soft palate traces are extracted from the first frame in each sequence; the exact location of the velum and uvula varies from frame to frame.

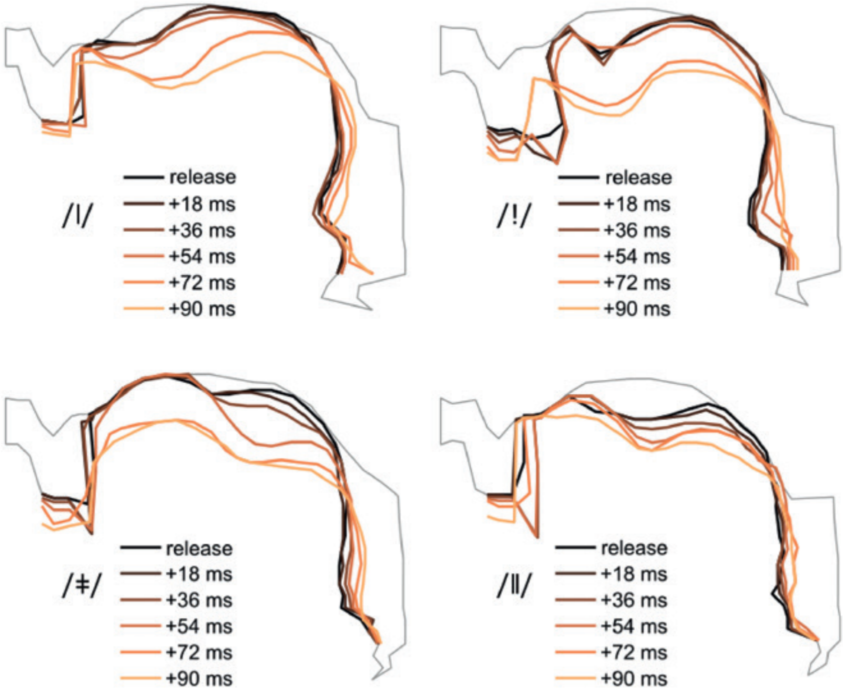

FIGURE 6.4 Evolution of midsagittal lingual posture in four contrastive Nama click releases
Notes: Top left: *ǀā* /ŋ̊ǀˀáà/ ‘sharp’ (dental); Top right: *!ā* /ŋ̊!ˀàa̋/ ‘hang’ (alveolar); Bottom left: *ǂā* /ŋ̊ǂˀa̋á/ ‘slaughter’ (palatal); Bottom right: *ǁā* /ŋ̊ǁˀa̋á/ ‘wash’ (lateral). In each sequence, the first (darkest) outline traces the tongue edge at the moment of click release. Subsequent outlines show tongue posture at 18 ms intervals, as the tongue moves toward the following low vowel target.

against some part of the uvula. Across all vowel contexts, /ǀ/ involves a more advanced posterior constriction location than that observed in the other three clicks produced by this speaker.

These data also show how cavity rarefaction is achieved using different parts of the tongue, and with different lingual kinematics. Dental and alveolar click release involves rapid lowering of the tongue blade and the front of the tongue body in the region immediately behind the anterior constriction. In palatal and lateral clicks, the lingual cavity is rarefied primarily by lowering and retracting the tongue body. In all four clicks, lingual articulation is accompanied by jaw lowering and retraction during release (Proctor et al. 2016).

3.3 *Discussion*

Real-time MRI provides rich information about the location and type of tongue tip and tongue body constriction used to produce each of the four Nama click types, the way that the lingual cavity is rarefied, and lingual motion involved in click release. Not all of these details of production can be obtained from other experimental methods that provide more restricted information about the tongue and the way it interacts with the passive articulators.

For this speaker, /ǀ/ and /ǃ/ are produced with narrow apical coronal gestures in the midsagittal plane, while /ǂ/ and /ǁ/ involve laminal tongue tip constrictions that contact a much wider region of the alveolar ridge and hard palate. The posterior constriction of the dental click produced before the low vowel extends forward into the velar region of the soft palate, but in all four clicks a large part of the tongue body remains constricted against the uvula (Fig. 6.3, top row), increasingly so during release.

Ultrasound provides similar information about the shape of the upper edge of the tongue around the click cavity (see, for example, Miller et al. 2007b); however, because rtMRI reveals the location of the entire mass of the tongue in the midsagittal plane, it allows us to better understand how click constrictions are formed and released, and how parts of the tongue below the oral airway are articulated and coordinated with other activity in the vocal tract.

Figs. 6.3 and 6.4 reveal that different parts of the tongue are lowered – in different ways and with different timing – during cavity rarefaction. The dental clicks examined here, for example, are released with more tongue body retraction than alveolar clicks, which are formed with a more retracted tongue body posture to begin with, and are therefore characterized primarily by tongue lowering. The apex of the tongue dorsum in the dental click produced before a low vowel, for example (Fig. 6.4, top left) is located 10 mm further forward along the soft palate compared to the dorsal apex in the alveolar click, and then retracts 11 mm towards the pharynx during release, while the dorsum retracts less than 5 mm during /ǃ/ release. The lowest point of the lingual cavity retracts over 15 mm during the same interval in /ǀ/, compared to 4 mm retraction in the alveolar click release. Asymmetries can also be seen in the trajectories of the anterior and posterior constrictions: this speaker's palatal clicks are rarefied and released primarily through dorsal lowering and retraction, while the tongue blade remains constricted against the hard palate (Fig. 6.4, bottom left: 18…54 ms).

Greater pharyngeal apertures are observed for this speaker during /ǀ/ and /ǃ/ formation and release, compared to /ǂ/ and /ǁ/. In clicks produced before a low vowel, for example, the tongue root (measured at the base of the epiglottis) is approximately 11 mm more advanced at /ǃ/ release (Fig. 6.4, top row, early frames) than at the same point during /ǂ/ release (Fig. 6.4, bottom row, early frames). These data suggest that dental and alveolar clicks may be produced in Khoekhoegowab with a more advanced tongue root than palatal and lateral clicks.

Looking beyond the tongue, these data illustrate another major advantage of rtMRI for studying non-pulmonic consonants: the global midsagittal view of the vocal tract reveals how other articulators are coordinated with lingual activity. The images in Fig. 6.3 show the state of the velum at any point in time, which remains lowered throughout the entire period of click production in these words. Although it is not always possible to determine the exact state of the glottis from rtMRI data, the images in Fig. 6.3 also show that the glottis is at least partially constricted during click release (c.f. the completely open glottises shown in Fig. 6.13). On the basis of these observations, the rtMRI data provide support for previous characterizations of the 'plain' Nama click series as voiceless nasal lingual ingressive consonants with a glottalized component (Brugman 2009; Miller 2011); however, in palatal and lateral clicks, the posterior constriction location may be better characterized as uvular: /ŋ̥ǀˀ/–/ŋ̥ǃˀ/–/ɴ̥ǂˀ/–/ɴ̥ǁˀ/.

4 rtMRI Insights into siSwati Clicks

SiSwati is a Southern Bantu language, primarily spoken in Swaziland and South Africa. siSwati (ISO 639–3 code: SSW) is a Tekela variety of the Nguni language group (S43: Guthrie 1971), along with Ndebele, Phuthi, and Lala (Gowlett 2003).

There is no consensus in the literature over the number and nature of click contrasts in siSwati. Most accounts argue for a single click place, most commonly characterized as dental (Nussbaum 1969; Rycroft 1981; Taljaard et al. 1991; Kockaert 1996), but also described as alveolar (Gowlett 2003). Doke (1954), and Chen and Malambe (1998) list /ǀ/ and /ǃ/ as distinct phonemes. The picture is complicated by multilingualism and the influence of loan words, as many neighbouring languages use more extensive click inventories. Even in the native lexicon, siSwati clicks appear to be realized with considerable variation in place (Ziervogel 1952; Nussbaum 1969; Corum 1980). There is also disagreement over the laryngeal and nasal features that differentiate clicks in siSwati (Ziervogel 1952; Doke 1954; Nussbaum 1969; Kockaert 1996), although again, this may reflect different approaches to transcription and phonological description.

One reason for the incomplete understanding of these segments is a lack of phonetic data. Most descriptions of the sound system of siSwati appear to rely solely on impressionistic transcription. To the best of our knowledge, no instrumental studies of siSwati click consonants have previously been published. Here, we present some results from an initial rtMRI investigation into the articulatory characteristics of clicks produced by a single speaker of siSwati, to shed more light on some of these issues.

4.1 *Informant and Corpora*

The study participant was a 47 year-old female, born and raised in Swaziland to first language speakers of siSwati. The informant also speaks English and Xitsonga but identifies siSwati has her primary language. She has lived most of her life in Swaziland, and has also lived in South Africa and Arizona as an adult. The participant reported normal hearing abilities and no speech pathologies. She has three missing teeth: left and right lower second molars and upper right second pre-molar.

The participant read out short passages of prose and lists of words targeting siSwati phonological contrasts of interest. In addition to the study corpus, some spontaneous speech was recorded. Twenty one recordings were made in total, each lasting between 19 and 34 seconds. The subset of the corpus analyzed in this study is listed in Appendix B.

4.2 *Articulatory Characterization of siSwati Clicks*

A real-time MRI sequence of a ‘plain’ siSwati word-initial click produced before a high-back vowel is illustrated in Figure 6.5, with time-aligned acoustic data. Every 5th frame is shown, spaced at 30 ms intervals, with corresponding landmarks indicated on the speech waveform. SiSwati click production before low and high front vowels is illustrated in the frame sequences in Figures 6.6 and 6.7, for comparison.

Only minor influences of vowel context on anterior constriction formation are seen in these data. In each vowel context, the target of the initial tongue-tip gesture is dental (Fig. 6.5, frames 1–6; Fig. 6.6, frames 78–84), and the anterior seal is formed with an apico-laminal constriction extending from the back of the upper teeth over the entire alveolar ridge. As observed in the Nama dental clicks, the laminal component of the tongue tip gesture becomes more prominent at the point of release of the siSwati click before a low vowel (Fig. 6.6: 90–102).

Before the low vowel /ɑ/, the posterior seal is created with a broad dorsal gesture extending across the lower velum and uvula (Fig. 6.6: 90–108). As in Nama /ǀ/, rarefaction appears to be achieved primarily through lowering of the tongue blade and the centre of the tongue in the region immediately behind the anterior constriction, below the center of the hard palate (Fig. 6.6: 102–114). Click release in this vowel context was characterized by a large amount of jaw lowering and retraction (–13 mm vertical, and –13 mm horizontal displacement, from click release to vowel target).

Articulation of the nasalized click in the word *ncisha* /ᴺǀiʃɑ/ ‘delete it’ is shown in Fig. 6.7. The coarticulatory influence of the high front vowel is evident throughout the sequence: at each stage of production, the tongue body is articulated in a higher, more fronted position, compared to the posture captured

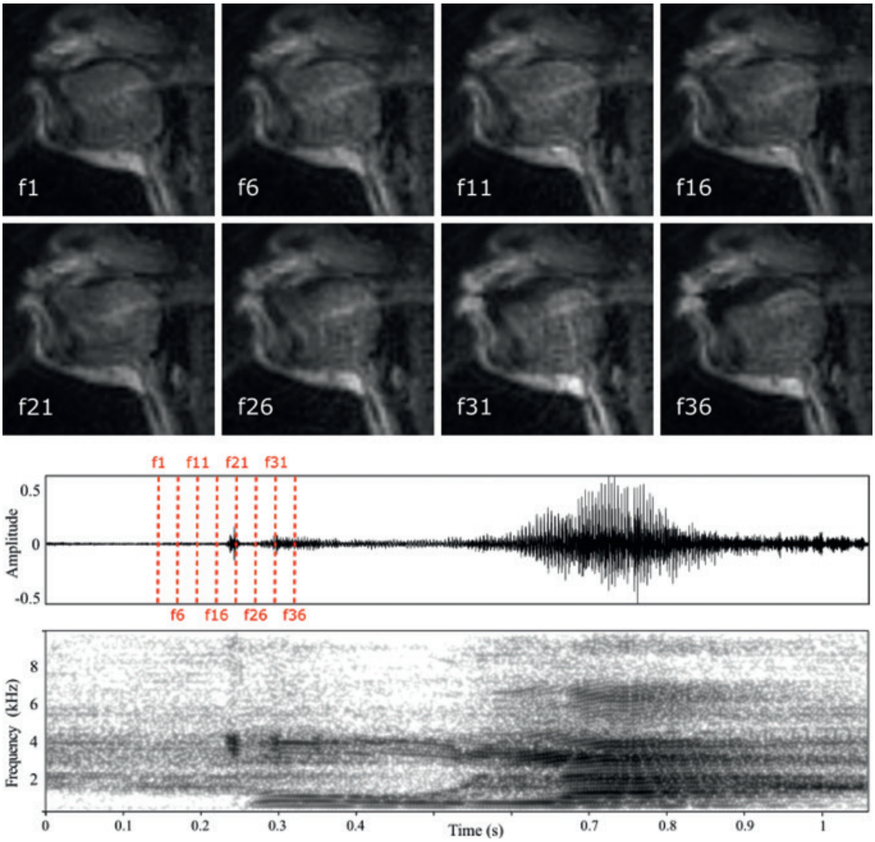

FIGURE 6.5 siSwati 'plain' click production before a high-back vowel
Notes: Eight image frames (top two rows) showing midsagittal articulation at key stages of the consonant-vowel sequence beginning the word *cula* /ǀúlɑ/ 'knock down'. Broken vertical lines superimposed on the waveform of the companion acoustic recording indicate the location in time of each frame.

at equivalent points in time in Fig. 6.6. The dental place of articulation of the anterior seal remains the same in this vowel context (Fig. 6.7, frames 754–772). The posterior seal is still achieved with a broad dorsal gesture, but with a more advanced closure target. The dorsal constriction extends forward into the velar region and becomes progressively more uvular during rarefaction and release (Fig. 6.7, frames 772–787). Tongue lowering ceases approximately 50 ms after click release, after which the tongue raises and advances towards the vowel target (Fig. 6.7, frames 793–805).

Another difference between siSwati clicks in these two vowel contexts is the articulation of the pharynx: before the high front vowel, the same click

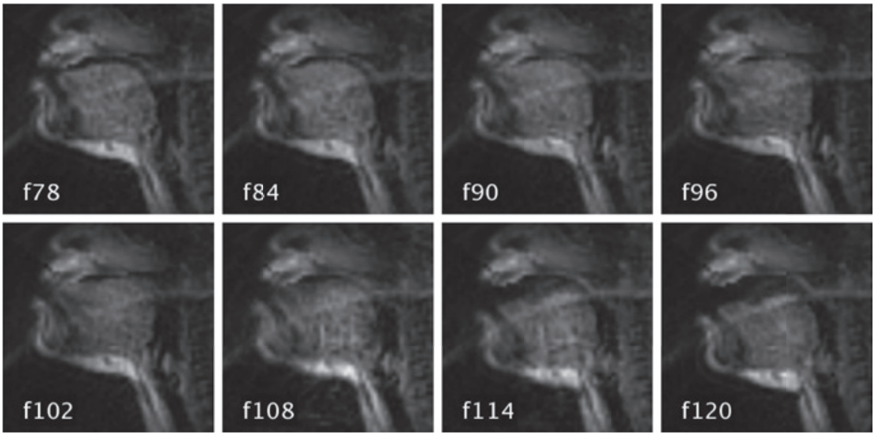

FIGURE 6.6 siSwati nasal click production in a low back vowel context: mid-sagittal articulation of the onset consonant in *ncata* /ᴺǀɑtɑ/ 'sound made when irritated'
Notes: MRI frames shown at 36 ms intervals. Frame 78: initial posture (140 ms before release); Frames 84–96: coronal closure followed by dorsal constriction; Frame 102: acoustic onset of click; Frame 120: lingual posture at acoustic target of post-consonantal vowel.

is produced with an advanced tongue root (evident in the relative displacement of the epiglottis), and greater overall pharyngeal aperture (Fig. 6.7, frames 90–108). Vowel context also influences the kinematics of the jaw and tongue dorsum: click release into a following i vowel is characterized by less dorsal lowering and retraction, and much less jaw motion (−4 mm vertical, and −1 mm horizontal displacement, click release to vowel target) than was observed during click release into a following /ɑ/ vowel.

Differences in clicks arising from vocalic coarticulatory effects are better observed in a side-by-side comparison; lingual postures of voiceless aspirated siSwati clicks produced before three different vowels are compared in the enlarged image frames juxtaposed in Figure 6.8. Differences in the size and geometry of the midsagittal cavity can be observed. These images also show the influence of vowel context on labial posture: the onset consonant of *choba* /ǀʰoɓɑ/ 'crush' is realized with a smaller labial aperture (20 mm) and more lip protrusion (tongue tip to lips = 6 mm) at the moment of click release, compared to clicks produced before non-rounded vowels (/ǀʰi/: labial aperture = 6 mm; tongue tip to lips = 17 mm).

We can characterize these clicks more accurately by comparing them to other consonants used in siSwati. Tongue postures captured at the articulatory centers of velar stops produced in three different vowel contexts are compared in Fig. 6.9. As expected, the images reveal large differences in dorsal posture

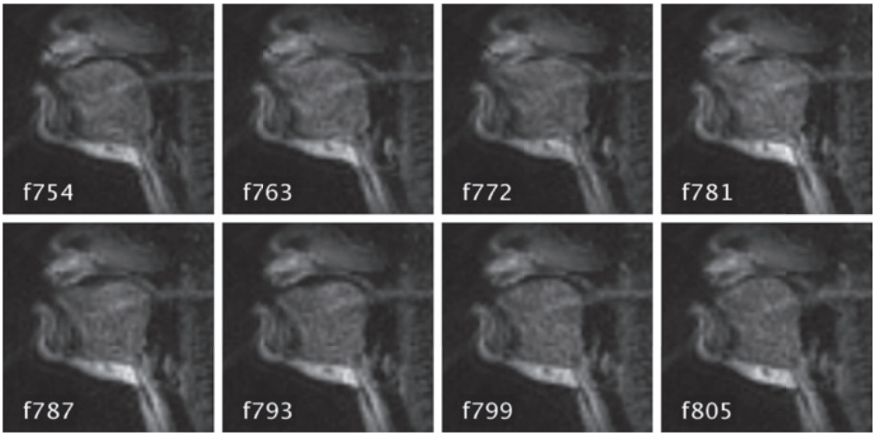

FIGURE 6.7 siSwati nasal click production in a high front vowel context: mid-sagittal articulation of the onset consonant in *ncisha* /ᴺǀiʃɑ/ 'delete it'
Notes: MRI frames shown at 36 ms intervals. Frame 754: initial posture (140 ms before release); Frames 763–781: coronal closure followed by dorsal constriction; Frame 787: acoustic onset of click; Frame 805: lingual posture at acoustic target of post-consonantal vowel.

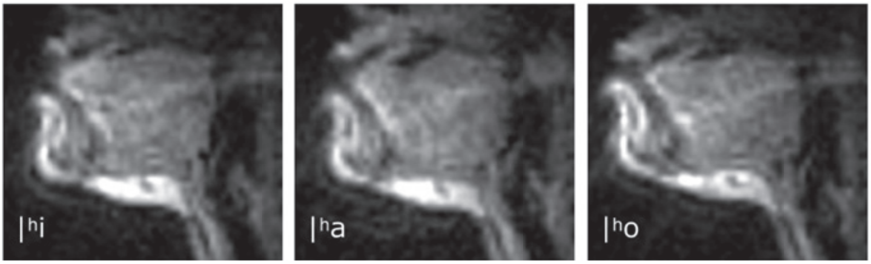

FIGURE 6.8 Influence of vowel context on siSwati clicks: Lingual posture at acoustic onset of voiceless aspirated clicks
Notes: Left: *lichibi* /liǀʰiɓi/ 'pond, puddle'; Centre: *chacha* /ǀʰɑǀʰɑ/ 'undo'; Right: *choba* /ǀʰoɓɑ/ 'crush'.

due to the coarticulatory influence of the vowels: the apex of the dorsum is 28 mm more advanced in the velar stop in **kimi** /k'imi/ [k̟ˀimi], compared to **liduku** /lí̤d̤uk'ú/ [lí̤d̤uqˀu]. Comparison with the images in Figs. 6.6–6.7 reveals that the clicks use a generally more retracted and extended dorsal constriction: the anterior limit of the posterior click seal is always more retracted than the most advanced velar allophones, and most closely resembles that of the uvular allophones. The dorsal constriction of the click typically appears to extend further back than the equivalent stop constriction, making contact with a greater

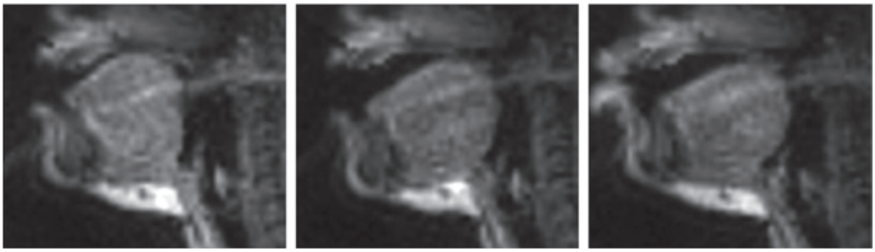

FIGURE 6.9 Place of articulation of siSwati velar stops
Notes: Left to right: lingual posture at point of maximum constriction of velar stop in three vowel contexts: *kimi* /k’ími/ ‘in me’ ; *kabi* /k’áɓi/ ‘badly, severely’ ; *liduku* /líd̤uk’ú/ ‘piece of cloth’.

length of the uvula, because the tongue root appears be to less advanced and the lower pharynx more constricted during lingual ingressive consonant production; further investigation will be required to see if this observation holds across a wider range of utterances.

5 Paralinguistic Click Production

Vocal percussion – the use of the vocal organs to generate percussion sounds, or imitate percussion instruments – has a long history in many cultures and musical performance genres, such as *konnakol* recitation in southern Indian musical traditions (Nelson 2008). Recently, vocal percussion has come to find new expression in hip-hop performance, where ‘human beatboxers’ imitate synthetic drum sounds, typically to create an accompanying percussion track for another rapper.

Many vocal percussion sounds can be described using the International Phonetic Alphabet because they are produced in the same ways, and share many phonetic properties with sounds exploited phonologically in languages (Lederer 2005; Proctor et al. 2013). Of particular interest are the clicks used by beatbox artists to imitate percussion elements such as wood blocks, snare drums, and rimshots (Stowell and Plumbley 2008). These data present an interesting case study for two reasons: the clicks are produced with musical, rather than linguistic goals, and they are typically produced by speakers of languages that do not use clicks phonemically.

In this section, we review rtMRI data showing click articulation by a third informant, a beatbox artist. These data (originally described in Proctor et al. 2010b; Proctor et al. 2013) allow us to examine mechanisms of click production in a speaker uninfluenced by phonological constraints. By comparing these sounds with equivalents in Nama and siSwati, we can consider which aspects of click production might be language-specific, and which aspects arise from more general mechanisms of sound production.

5.1 *Participant and Corpus*

The participant is a male professional singer, working in a wide variety of vocal performance styles including hip-hop, pop, and R&B. At the time of the study, the subject was 27 years old, and had been working for 10 years as an emcee in a hip-hop duo, and as a session vocalist with other groups. The informant was born in Orange County, California, to Panamanian parents, and lives and works in Los Angeles. He is a native speaker of American English, and a heritage speaker of Panamanian Spanish.

The participant demonstrated all of the percussion effects in his repertoire, and several beatboxing sequences, performing in short intervals as he lay supine in the scanner bore. Each target effect was demonstrated at least five times, separated by short pauses of approximately two seconds. Image data were acquired with a different pulse sequence to that used for the other click studies (for details see Proctor et al. 2013), but the same methods were used for audio acquisition, image processing, and analysis.

5.2 *Articulation of Clicks in Vocal Percussion*

Of the seventeen phonetically-distinct percussion effects in the informant's repertoire, four made use of clicks. These were described by the subject as a snare drum, a high-hat, and two different rimshot effects. Before examining how these sounds were produced, we first identified the lingual postures associated with English consonants produced by the informant, to establish articulatory landmarks which would allow us to characterize his clicks. Typical dental, alveolar, and velar constriction targets (captured at consonant midpoints in rapidly-rapped English words) are illustrated in Fig. 6.10.

Articulation of the effect described as a 'side K rimshot' is illustrated in the image sequence in Fig. 6.11, acquired over a 480 ms interval. The posterior constriction is formed first, followed by the coronal constriction, completed 119 ms before the acoustic onset of the click. The seal formed between the body of the tongue and the soft palate initially extends from the velar region to the bottom of the uvula (frame 283); after click release, the tongue remains in contact with the bottom of the uvula (f293). Place of articulation cannot be characterized in terms of a single target for the anterior lingual seal at formation, as the blade of the tongue forms a continuous constriction along the entire midline of the palate (f289). After click release, the midsagittal anterior constriction location can be characterised as alveolar (Fig. 6.11: f293; c.f. Fig. 6.10: /t/).

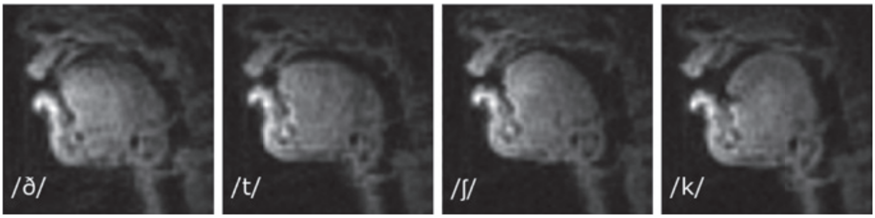

FIGURE 6.10 Places of articulation in English consonants produced by the beatbox artist
Notes: Mid-consonantal posture captured in dental, alveolar and two velar obstruents, in rapid speech. Left-to-right: /ð/ ('the'); /t/ ('master'); /ʃ/ ('she'); /k/ ('authentic').

At the point of maximum lingual constriction, no midsagittal cavity can be seen between the tongue and the palate (Fig. 6.11: f289). Rarefaction and cavity formation occur in the mid palatal region while anterior and posterior lingual seals are maintained (frames 290–291), suggesting that the consonantal influx is lateralized, consistent with the subject's description of the click as being produced at 'the side of the mouth'. The velum remains lowered throughout. The same articulatory sequence was observed in all seven examples of this sound effect demonstrated by the subject, which may be described as a voiceless nasalized lateral click: [ᴺ̥ǁ].

An additional rimshot effect demonstrated by the subject is illustrated in Fig. 6.12: five image frames acquired over a 384 ms interval. Unlike the Nama dental and alveolar clicks, which were initiated with an apico-(post)alveolar gesture, the anterior seal in this sound effect is formed with a laminal coronal alveolar constriction. The posterior closure is spread across the velar/uvular regions (Fig. 6.12, frames 17–21), but initially at a more anterior location to that observed during constriction formation in Fig. 6.11. In terms of initial constriction location, and the way that the seal formed between the tongue body and the velum/uvula, articulation of the posterior constriction in this effect closely resembles that of the Nama /ǀ/. As in the lateral rimshot, posterior constriction formation precedes anterior closure in most cases, although some variation in the relative timing of these gestures was observed across eight repetitions of this effect.

Rarefaction occurs in the palatal/velar region, as a result of jaw and tongue lowering. The velo-pharyngeal port remains open throughout, and the velum is lowered onto the rear part of the tongue dorsum as the click is released, so that the final location of the posterior lingual seal is against the lowest part of the uvula. This effect may be described as a type of voiceless nasalized alveolar click: [ᴺ̥!]. The informant used the additional description 'sucking in' to differentiate this sound from the previous rimshot effect, which suggests that he

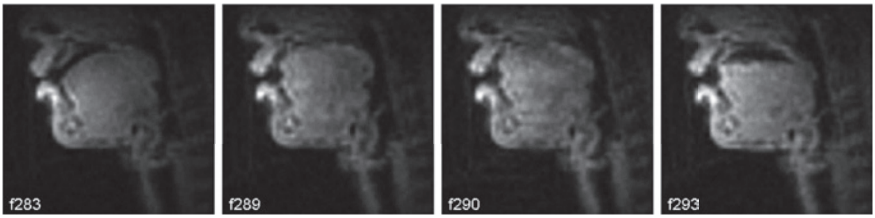

FIGURE 6.11 Articulation of a 'side K' rim shot effect as a lateral click [ᶰ̥ǁ]
Notes: Frame 283: posterior constriction formed between tongue dorsum and soft palate/uvula; f289: alveolar anterior constriction completes lingual seal; f290: rarefaction of palatal cavity; f293: final lingual posture after lateral influx: lowered tongue body retaining anterior & posterior lingual seals. Velum remains lowered throughout click production.

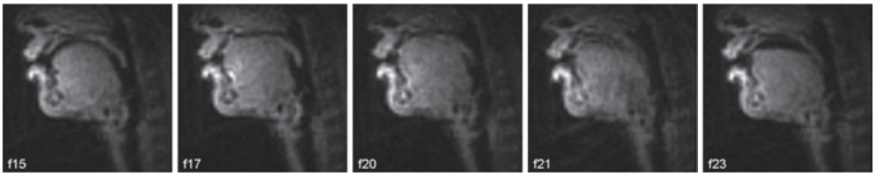

FIGURE 6.12 Articulation of a rimshot effect as an alveolar click [ᶰ̥!]
Notes: Frame 15: lingual raising and advancement towards palate; f17: lingual seal extending from alveolar ridge to soft palate; f20–21: rarefaction of lingual cavity beneath palate; f23: final lingual posture after click release. Velum remains lowered throughout click production.

may have accompanied the click release with some ingressive pulmonic activity. The glottis appears to remain largely open throughout this effect, but it is not possible to establish whether additional airstream mechanisms were used from these data alone.

The image sequence in Fig. 6.13 – four frames acquired over a 340 ms interval – illustrates articulation of a 'clap snare' sound effect. As in most of the clicks demonstrated by this subject, an extended lingual seal is first created along the the entire midline of the hard and soft palates (frame 387). The front constriction in this click is more anterior than that observed in the lateral (Fig. 6.11) and alveolar clicks (Fig. 6.12), with the point of influx closer to the subject's teeth (Fig. 6.13, frames 393–4; c.f. Fig. 6.10: /ð/). The dorsal seal extends along the entire length of the uvula, and the place of articulation of the posterior constriction is clearly uvular at release (frames 393–394). Labial approximation begins when the tongue and jaw are raised to form the lingual seal; the lips are completely closed during the rarefaction stage (frame 391),

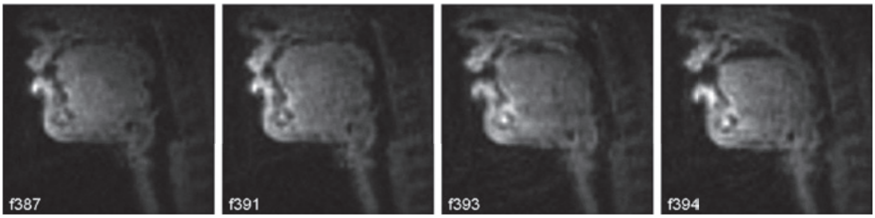

FIGURE 6.13 Articulation of a ‘clap’ snare drum effect as a labialized dental click [ᵑ̊ǀʷ]
Notes: Frame 387: tongue pressed into palate; f391–393: rarefaction of palatal cavity: lowering of jaw and tongue centre; f394: click release reveals dental-alveolar anterior constriction and extended uvular posterior lingual seal; Velum remains lowered throughout.

and reopened as the click is released (frame 393). The velum remains lowered and the glottis is clearly open throughout the entire sequence, which may be described as a labialized voiceless nasalized dental click: [ᵑ̊ǀʷ].

One final sound effect in the repertoire was produced as a click, described by the subject as a ‘hi-hat closed: kiss teeth’. Six examples of this sound were demonstrated; lingual articulation in each case very closely resembled that observed in the ‘clap snare’ (Fig. 6.13). Some minor variation was observed in the timing of labialization and the degree of constriction of the glottis; however, the effect may also be described as a labialized voiceless nasalized dental click [ᵑ̊ǀʷ].

5.3 *Insights from Beatboxing*

These data demonstrate that speakers of languages lacking lingual ingressive consonants can produce different types of clicks for paralinguistic purposes, and that, at least in the case of this informant, these sounds are generated in very similar ways to the click phonemes produced by native speakers. The sound effects demonstrated by this beatboxer can be described using standard IPA because they are characterized by the same broad places and patterns of articulation as the /ǀ/, /ǃ/ and /ǁ/ consonants used in Bantu and Khoisan languages, although we do not yet know if all clicks produced for paralinguistic purposes have equivalents in phonological systems.

The posterior constriction location in all of these sound effects is best characterized as uvular. All these clicks are intrinsically voiceless, and pervasively nasalized (produced with the velo-pharyngeal port open), suggesting that when nasalization does not need to be controlled for phonological reasons, the dorsal lingual seal may be more naturally formed against a lowered velum and uvula. More research is needed to establish how different airstream mechanisms combine with nasalized clicks (cf. Miller et al. 2009a), and to examine the acoustic consequences of nasal coupling at different stages in the production of non-pulmonic consonants in more detail.

A notable difference between the clicks produced by the beatboxer and those of the two native speakers of click languages involves the way the anterior constriction is formed. While some separation between posterior and anterior seals can be observed during closure in all the clicks illustrated in Figs. 6.3 to 6.8, the beatboxer's clicks were formed with what appears to be a continuous seal along the entire palate. It is not clear if this reflects the paralinguistic nature of the task, anatomical differences between speakers, or language differences. The clicks produced by this siSwati speaker in high front vowel contexts, for example, are formed with a much smaller initial cavity than comparable clicks produced by the Nama speaker. Further research might examine whether native speakers of click languages ever articulate clicks without an initial cavity within the lingual-palatal seal.

Extensive pre-labialization was used to shape the sound of two of the four types of clicks demonstrated – a secondary articulation not commonly exploited in click phonologies, even in the most extended lingual ingressive consonant inventories (Maddieson 1984; Ladefoged and Maddieson 1996).

6 Discussion

These data illustrate the utility of real-time MRI as a method for investigating the phonetic and phonological properties of clicks. By providing a means of visualizing the dynamic configuration of the entire vocal tract, rtMRI offers some important advantages over other methods used to study clicks.

rtMRI provides rich information about place of articulation in click consonants. Unlike ultrasound, in which the location of the palate must be reconstructed, rtMRI directly images the alveolar ridge, hard palate, velum and uvula, and reveals exactly how and where the tongue comes into contact with it. Unlike palatography, the field of view rtMRI includes the entire soft palate, including the uvula. In practice, the accuracy and amount of detail that rtMRI can provide about linguo-palatal contact will depend on the spatio-temporal resolution of the imaging sequences and other factors (see Sections 6.1–6.3). These data demonstrate the importance of this information for accurate characterization of the phonetic properties and differences between clicks. Nama dental clicks are characterized by an apico-laminal coronal constriction extending from the back of the upper teeth over the entire alveolar ridge; siSwati clicks are initiated as apico-dentals and released as lamino-alveolars; while the beatboxer initiates all clicks with an extended lingual constriction across the whole palate.

Because it reveals the location and posture of the whole mass of the tongue in a given imaging plane, rtMRI can provide richer insights into lingual articulation and coarticulation than modalities that only provide information from part of the vocal tract. Dental and alveolar clicks were found to be produced by this speaker of Nama with a greater pharyngeal aperture and more advanced tongue root than his palatal and lateral clicks. Palatal clicks are produced by this speaker with a more bunched tongue posture than /!/, which shows greater independence of coronal and dorsal tongue gestures. Clicks produced by this speaker of siSwati before back rounded vowels have been shown to involve greater labial constriction and protrusion than her clicks produced in other vowel contexts. More research is required to test these characterizations for other speakers of these languages, and for other speakers' paralinguistic click productions.

Neither palatography nor ultrasound is able to track the formation and release of the posterior click constriction with the same level of detail as rtMRI, as neither method provides reliable information about the articulation of the velum and uvula against the tongue dorsum. rtMRI data from these three subjects reveal that click production typically involves an extended dorsal seal which may begin as far forward as the velar place of articulation and extend all the way to the bottom of the uvula. Characterizing the posterior 'place of articulation' of click consonants is not straightforward, as the anterior point of contact between the tongue dorsum and the soft palate often retracts throughout click release, may not be revealed in the midsagittal plane until after acoustic release of the click, and is influenced by vowel coarticulation. Nevertheless, these data suggest that the posterior constriction is typically best characterized as uvular, and velar in the case of Nama dental clicks.

The main phonetic properties of the clicks examined in these data are summarized in Table 6.2.

X-ray offers many of the same key advantages as rtMRI: the ability to view the whole vocal tract, and observe the coordination of articulators beyond the tongue during click production. However, because of the risks associated with exposing subjects to ionizing radiation, X-ray is no longer considered to be a suitable method for general phonetic investigation. rtMRI therefore offers the prospect of safely examining some aspects of click production that have not been attempted since Doke's (1923; 1923) and Traill's (1985) pioneering X-ray studies of click consonants.

6.1 *Limitations*

We present these data as an initial exploration of the potential of rtMRI for studying clicks. The selection of languages is not intended to represent a typologically-balanced sample of phonological systems that use click

TABLE 6.2 Summary of phonetic properties of click consonants

		Anterior constriction	**Coronal posture**	**Posterior constriction**	**Cavity location**	**Release mechanism**
/ǀ/	Nama	dental-alveolar	apical	velar	palatal	dorsal lowering
	Swati	dental-alveolar	laminal	velar	palatal	dorsal retraction
	BBox	dental-alveolar	apical	uvular	palatal	dorsal retraction
/ǃ/	Nama	postalveolar	apical/retroflex	uvular	palato-velar	tongue lowering
	BBox	alveolar	laminal	uvular	palatal	dorsal retraction
/ǁ/	Nama	alveolar	laminal	uvular	velar	dorsal lowering
	BBox	alveolar	laminal	uvular	palatal	dorsal retraction
/ǂ/	Nama	palatal	laminal	uvular	velar	tongue lowering

Notes: L-to-R: anterior constriction location at click release; coronal posture at release; posterior constriction location at release; location of lingual cavity at greatest aperture; and primary lingual motion characterizing click release.

consonants, but was determined in part by the availability of participants. The use of a single informant in each of the target languages is another obvious limitation; it is not clear which of the patterns of articulation described here are speaker-specific, nor how click production might be influenced by factors including vocal tract morphology (Lammert et al. 2013) and prosody (Brugman 2009).

The audio component of these data is not ideal for studying lingual ingressive consonants. Acoustic recordings in an MRI scanner are characterized by limited bandwidth and echo artifacts that are particularly problematic for highly transient sounds such as clicks. Speech recordings acquired during an MRI scan additionally require extensive post-processing to attenuate the pervasive scanner noise (Bresch et al. 2006; Vaz et al. 2013). Although the resulting recordings are of sufficient quality for general acoustic analysis (e.g. Lammert et al. 2013) and automatic phonetic transcription (Katsamanis et al. 2011), the audio signal is degraded in ways that currently prevent more extensive analysis.

These data also demonstrate some inconsistencies in image quality which can affect MRI studies of the upper airway. Signal-to-noise ratios may vary throughout the image depending on the type and configuration of receiver coil, which can affect resolution of the velum, pharynx and larynx. Cardiac artifacts can mask or distort key anatomical regions, in particular the tongue root and velum. In some cases, implanted medical devices prevent some subjects from participating in any kind of magnetic resonance imaging study, due to safety issues. More commonly, metallic substances used in dental work introduce imaging artifacts which distort parts of the signal beyond the point where accurate information about vocal tract configuration can be recovered. Although the minor dental artifacts observed in these Nama data were not problematic, the much larger artifacts that occasionally degrade other subjects' MRI data can prove to be prohibitive for some types of analysis.

6.2 *Recommendations*

Although not yet as widely used as static structural imaging, real-time MRI is an increasing viable technology for speech studies. Lingala et al. (2016) outline the requirements, considerations, and expectations for rtMRI, and suggest protocols for acquiring speech data with different spatio-temporal resolutions. Although click consonants are especially demanding objects of study, different aspects of click production can be fruitfully studied using a wide variety of different sequences on commonly available scanners.

Global and consistent imaging of the vocal tract midline is not always possible: due to misorientation of the participant's head and/or anatomical asymmetries, different regions of the vocal tract may be more closely aligned with the midsagittal imaging plane. A study focusing on airstream mechanisms and laryngeal setting should verify the mid-glottal alignment of the imaging plane by eliciting voiced, voiceless, and glottally-stopped segments during localization. There is a trade-off between imaging slice thickness and signal-to-noise ratio: slices wider than 5mm offer good temporal and spatial resolution at the high frame rates needed to image clicks (Lingala et al. 2016), but may be too wide to resolve some anatomical details.

It is important to immobilize the participant's head so that they are as comfortable and stable as possible throughout the scan session. Nevertheless, small movements can and do occur, shifting the anatomical alignment across images acquired over the course of an utterance or a scan session. Where necessary, misalignment due to head movement can be corrected by calculating displacement between frames (using correlation of static anatomical features), and applying compensating image transformations.

6.3 *Future Directions*

The data presented here were all acquired using a midsagittal imaging plane, so they provide no information about articulation beyond the midline of the vocal tract. A comprehensive understanding of click production will need to consider activity in other parts of the oral cavity. MRI can be acquired using parasagittal, coronal and axial imaging planes, planes of arbitrary orientation that intersect any part of the upper airway of interest, and with interleaved acquisition from multiple imaging planes (Proctor et al. 2008; Proctor et al. 2010c; Kim et al. 2012; Zhu et al. 2013). Some of these advanced imaging techniques would provide additional information about tongue shaping and vocal tract configuration, which would be especially helpful in the study of lateral clicks. Image sequences offering greater SNR, and improved spatial and temporal resolution (e.g. Niebergall et al. 2013; Lingala et al. 2017) will be important to better resolve the hard palate and allow more accurate analysis of laryngeal and velic activity.

Acknowledgments

This work was supported by National Institutes of Health grant R01 DC007124–01 and Australian Research Council Award DE150100318. Special thanks to our informants for their participation, patience, talents, and insights into their languages.

Appendix A: Nama Elicitation Items

TABLE 6.3 Nama orthography and broad phonetic transcriptions, with Afrikaans disambiguators (where used during presentation), and English glosses

Nama	IPA	Afrikaans	English
ǀā	/ᵑ̊ǀˀáà/	*skerp*	‘sharp’
ǁā	/ᵑ̊ǁˀa̋á/	*was*	‘wash’
ǃā	/ᵑ̊ǃˀàa̋/	*ophang*	‘hang out’
ǂā	/ᵑ̊ǂˀa̋á/	*slag*	‘slaughter’
ǀom	/ᵑ̊ǀˀòm̀/	*asemhaal*	‘breathe’
ǁom	/ᵑ̊ǁˀòm̋/	*slaap*	‘sleep’
ǃom	/ᵑ̊ǃˀóm̀/	*doring uithaal*	‘remove thorn’
ǂom	/ᵑ̊ǂˀóm̀/	*werk met naald*	‘sew’
ǀgam	/ǀa̋ḿ/	*warm*	‘warm’
ǁgam	/ǁám̀/	*gesels*	‘talk’
ǃgam	/ǃȁḿ/	*diep*	‘deep’
ǂgam	/ǂȁḿ/	*van blydskap rondspring*	‘jumping for joy’
ga	/kȁá/		‘fool, trick’
ge	/kè/		‘we’ (masc.)
gu	/kù/		‘they’ (masc.)
ega	/e̋kȁ/		‘a bit later’

Appendix B: siSwati Elicitation Items

TABLE 6.4 siSwati orthography and broad phonetic transcriptions, with English glosses

siSwati	IPA	English
caba	/ǀáɓɑ/	‘cut down’
coba	/ǀoɓɑ/	‘break, snap’
cula	/ǀúlɑ/	‘knock down’
ncata	/ᴺǀɑtɑ/	‘sound made when irritated’
ncusa	/ᴺǀusɑ/	‘ask a favor, send with a message’
ncisha	/ᴺǀiʃɑ/	‘delete it (isiZulu)’
ncesi	/ᴺǀêsi/	‘sorry!’
choba	/ǀʰoɓɑ/	‘steal (isiZulu)’
lichubu	/liǀʰuɓu/	‘hump, hunchback’
lichibi	/líǀʰiɓí/	‘lake, pond’
chefe	/ǀʰefe/	‘irritating’
chacha	/ǀʰɑǀʰɑ/	‘undo’
kimi	/k’imi /	‘in me’
kabi	/k’ɑɓí/	‘badly, severely’
liduku	/líd̤úk’u/	‘piece of cloth’

References

Beach, Douglas M. 1938. *The Phonetics of the Hottentot Language*. Cambridge: W. Heffer & Sons.

Bradfield, Julian. 2014. “Clicks, concurrency and Khoisan.” *Phonology* 31(1): 1–49.

Bresch, Erik, Yoon-Chul Kim, Krishna Nayak, Dani Byrd, and Shrikanth Narayanan. 2008. “Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP].” *IEEE Signal Processing Magazine* 25(3): 123–132.

Bresch, Erik, Jon Nielsen, Krishna Nayak, and Shrikanth Narayanan. 2006. “Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans.” *Journal of the Acoustical Society of America* 120(4): 1791–1794.

Brugman, Johanna. 2009. “Segments, tones and distribution in Khoekhoe prosody.” Ph.D. thesis, Cornell University.

Chen, Su-I, and Gloria B. Malambe. 1998. “Palatalization in siSwati: an optimality theoretical approach.” *Language History and Linguistic Description in Africa*, ed. Ian Maddieson and Thomas J. Hinnebusch. Trenton, NJ: Africa World Press, pp. 137–146.

Corum, Claudia W. 1980. *SiSwati: Special Skills Handbook*. Battleboro VT: Peace Corps.

Cruttenden, Alan. 1992. “Clicks and syllables in the phonology of Dama.” *Lingua* 86(2–3): 101–117.

Curriculum Committee for Khoekhoegowab. 2003. *Khoekhoegowab: 3IIî xoaigaub* (*Orthography 3*). Windhoek: Gamsberg Macmillan Publishers.

Doke, Clement M. 1923a. “A Dissertation on the Phonetics of the Zulu Language.” *Bulletin of the School of Oriental and African Studies* 2(4): 687–729.

Doke, Clement M. 1923b. “Notes on a problem in the mechanism of the Zulu clicks.” *Bantu Studies* 2(1): 43–45.

Doke, Clement M. 1925. “The phonetics of ʗhũ̬ː Bushman.” *Bantu Studies* 2(3): 129–165.

Doke, Clement M. 1926. “The Phonetics of the Zulu Language.” *Bantu Studies* 2. Johannesburg: Wiwatersrand University.

Doke, Clement M. 1954. *The Southern Bantu languages: Handbook of African Languages*. London: Oxford University Press for the International African Institute.

Doke, Clement M., and S. Machabe Mofokeng. 1957. *Textbook of Southern Sotho Grammar*. Cape Town: Longman.

Finlayson, Rosalie, J. Jones, K. Podile, and Jan Snyman. 1991. *An Introduction to Xhosa Phonetics*. Hout Bay: Marius Lubbe.

Fulop, Sean A., Peter Ladefoged, Fang Liu, and Rainer Vossen. 2003. “Yeyi Clicks: Acoustic Description and Analysis.” *Phonetica* 60(4): 231–260.

Gowlett, Derek F. 2003. “Zone S.” *The Bantu Languages*, ed. Derek Nurse and Gérard Philippson. 4. London, New York: Routledge, pp. 609–638.

Güldemann, Tom. 2001. “Phonological regularities of consonant systems across Khoisan lineages.” *University of Leipzig Papers on Africa* 16: 1–50.

Güldemann, Tom, and Rainer Vossen. 2000. “Khoisan.” *African Languages: An Introduction*, ed. Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press, pp. 99–122.

Guthrie, Malcolm. 1971. *Comparative Bantu: an Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.

Haacke, Wilfrid H.G., and Eliphas Eiseb. 2002. *A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index*. Windhoek: Gamsberg Macmillan.

Haacke, Wilfrid H.G., Eliphas Eiseb, and Levi Namaseb. 1997. “Internal and External Relations of Khoekhoe Dialects: a Preliminary Survey.” *Namibian Languages: Reports and Papers*, ed. Wilfrid H.G. Haacke and Edward D. Elderkin. Köln: Rüdiger Köppe Verlag; University of Namibia (UNAM), pp. 125–209.

Herbert, Robert K. 1990. “The relative markedness of click sounds: Evidence from language change, acquisition, and avoidance.” *Anthropological Linguistics* 32(1–2): 120–138.

Kagaya, Ryohei. 1978. “A phonetic sketch of Naron around Ghanzi.” *Journal of Asian and African studies/Ajia Afuriku gengo bunka kenkyu* 16: 29–63.

Katsamanis, Athanasios, Matthew Black, Panayiotis Georgiou, Louis Goldstein, and Shrikanth Narayanan. 2011. “SailAlign: Robust long speech-text alignment.” *Proceedings of the Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*. University of Pennsylvania. Philadelphia, pp. 44–47.

Kim, Yoon-Chul, Michael I. Proctor, Shrikanth S. Narayanan, and Krishna S. Nayak. 2012. “Improved imaging of lingual articulation using real-time multislice MRI.” *Journal of Magnetic Resonance Imaging* 35(4): 943–948.

Kockaert, Hendrik J. 1996. *Corrective Strategies for the Pronunciation of French as a Foreign Language among Swazi Learners*. Johannesburg: Randse Afrikaanse Universiteit.

Köhler, Oswin, Peter Ladefoged, Jan Snyman, Anthony Traill, and Rainer Vossen. 1988. “The symbols for clicks.” *Journal of the International Phonetic Association* 18(2): 140–142.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World’s Languages*. Oxford; Cambridge, MA: Blackwell.

Ladefoged, Peter, and Anthony Traill. 1984. “Linguistic phonetic description of clicks.” *Language* 60(1): 1–120.

Ladefoged, Peter, and Anthony Traill. 1994. “Clicks and their accompaniments.” *Journal of Phonetics* 22: 33–64.

Lammert, Adam, Michael Proctor, and Shrikanth Narayanan. 2013. “Interspeaker Variability in Hard Palate Morphology and Vowel Production.” *Journal of Speech, Language, and Hearing Research* 56(6): S1924–S1933.

Lanham, Leonard W. 1964. “The proliferation and extension of Bantu phonemic systems influenced by Bushman and Hottentot.” *Proceedings of the 9th International Congress of Linguists, Cambridge 1962*, ed. Horace Lunt. The Hague: Mouton & Co., pp. 382–391.

Lederer, Karen. 2005. “The phonetics of beatboxing.” B.A. dissertation. Leeds University.

Lingala, Sajan G., Brad P. Sutton, Marc E. Miquel, and Krishna S. Nayak. 2016. “Recommendations for real-time speech MRI.” *Journal of Magnetic Resonance Imaging* 43(1): 28–44.

Lingala, Sajan G., Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan, and Krishna S. Nayak. 2017. “A fast and flexible MRI system for the study of dynamic vocal tract shaping.” *Magnetic Resonance in Medicine* 77(1): 112–125.

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

Maddieson, Ian. 2003. “The Sounds of Bantu Languages.” *The Bantu Languages*, ed. Derek Nurse and Gérard Phillipson. London: Routledge, pp. 15–41.

Miller-Ockhuizen, Amanda L. 2003. *Phonetics & Phonology of Gutturals: A Case Study from Ju|’hoansi*. Outstanding Dissertations in Linguistics. New York: Routledge.

Miller, Amanda L. 2011. “The Representation of Clicks.” *The Blackwell Companion to Phonology*, vol. 1, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice. Oxford: Blackwell, pp. 416–439.

Miller, Amanda L., Johanna Brugman, and Bonny Sands. 2007a. “Acoustic and auditory analysis of N|uu lingual and linguo-pulmonic stop bursts.” *Proceedings of the 16th International Congress of Phonetic Sciences*, ed. Jürgen Trouvain and William J. Barry. Saarbrücken: ICPhS, pp. 769–772.

Miller, Amanda L., Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, and Chris Collins. 2009a. “Differences in airstream and posterior place of articulation among N|uu clicks.” *Journal of the International Phonetic Association* 39(02): 129–161.

Miller, Amanda L., Levi Namaseb, and Khalil Iskarous. 2007b. “Tongue body constriction differences in click types.” Proc. *Laboratory Phonology 9*, ed. Jennifer S. Cole and José Ignacio Hualde. Berlin: Mouton de Gruyter, pp. 643–656.

Miller, Amanda L., Abigail Scott, Bonny E. Sands, and S. Shah. 2009b. “Rarefaction gestures and coarticulation in Mangetti Dune !Xung clicks.” *Proceedings of the International Conference on Speech Communication and Technology*. Brighton, pp. 2279–2282.

Nakagawa, Hirosi. 1996. “A first report on the click accompaniments of G|ui.” *Journal of the International Phonetic Association* 26(1): 41–54.

Nakagawa, Hirosi. 2006. “Aspects of the phonetic and phonological structure of the G|ui language.” Ph.D. thesis, University of the Witwatersrand.

Narayanan, Shrikanth, Krishna Nayak, Sungbok Lee, Abhinav Sethy, and Dani 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, Shrikanth et al. 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.

Nelson, David P. 2008. *Solkattu Manual: An Introduction to the Rhythmic Language of South Indian Music*. Middleton, CT: Wesleyan University Press.

Niebergall, Aaron, Shuo Zhang, Esther Kunay, Götz Keydana, Michael Job, Martin Uecker, and Jens 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.

Nussbaum, Loren V. 1969. *Understanding and Speaking siSwati*. Washington, DC: Center for Applied Linguistics.

Proctor, Michael I., Daniel Bone, and Shrikanth S. Narayanan. 2010a. “Rapid semi-automatic segmentation of real-time Magnetic Resonance Images for parametric vocal tract analysis.” *Proceedings of the International Conference on Speech Communication and Technology*. Makuhari, Japan, pp. 1576–1579.

Proctor, Michael I., Krishna S. Nayak, and Shrikanth S. Narayanan. 2010b. "Linguistic and para-linguistic mechanisms of production in human "beatboxing": a rtMRI study." *Proceedings of Intersinging*. Univeristy of Tokyo.

Proctor, Michael I., Christine H. Shadle, and Khalil Iskarous. 2010c. "Pharyngeal articulation differences in voiced and voiceless fricatives." *Journal of the Acoustical Society of America* 127(3): 1507–1518.

Proctor, Michael I., Christine Shadle, and Khalil Iskarous. 2008. "A method of co-registering multiple Magnetic Resonance Imaged vocal tract volumes for fricatives." *Journal of the Acoustical Society of America* 123(5): 3735–3735.

Proctor, Michael, Erik Bresch, Dani Byrd, Krishna Nayak, and Shrikanth Narayanan. 2013. "Paralinguistic mechanisms of production in human "beatboxing": a real-time magnetic resonance imaging study." *Journal of the Acoustical Society of America* 133(2): 1043–1054.

Proctor, Michael, Yinghua Zhu, Adam Lammert, Asterios Toutios, Bonny Sands, Ulrich Hummel, and Shrikanth Narayanan. 2016. "Click Consonant Production in Khoekhoe: a Real-time MRI Study." *Khoisan Languages and Linguistics – Proceedings of the 5th International Symposium, July 13–17, 2014, Riezlern/Kleinwalsertal*, ed. S. Shah and M. Brenzinger. *Research in Khoisan Studies*, vol. 34. Köln: Rüdiger Köppe Verlag, pp. 337–366.

Roux, Justus. 2007. "Unresolved issues in the representation and phonetic description of click articulation in Xhosa and Zulu." *Language Matters: Studies in the Languages of Africa* 38(1): 8–25.

Rycroft, David K. 1981. *Concise SiSwati Dictionary*. Pretoria: J.L. van Schaik.

Sands, Bonny. 1991. "Evidence for click features: Acoustic characteristics of Xhosa clicks." *UCLA Working Papers in Phonetics* 80: 6–37.

Sands, Bonny, Johanna Brugman, Mats Exter, Levi Namaseb, and Amanda Miller. 2007. "Articulatory characteristics of anterior click closures in N|uu." *Proceedings of the 16th International Congress of Phonetic Sciences*, ed. Jürgen Trouvain and William J. Barry. Saarbrücken: ICPhS, pp. 401–404.

Sands, Bonny, Ian Maddieson, and Peter Ladefoged. 1996. "The phonetic structures of Hadza." *Studies in African Linguistics* 25(2): 171–204.

Snyman, Jan W. 1970. *Introduction to !Xũ (!Kung) Language*. Cape Town: A.A. Balkema.

Snyman, Jan W. 1975. *Žul'hõasi Fonologie & Woordeboek*. Communication 37 of the University of Cape Town School of African Studies. Cape Town; Rotterdam: A.A. Balkema.

Stowell, Dan, and Mark D. Plumbley. 2008. "Characteristics of the beatboxing vocal style." *Technical Report*, Centre for Digital Music, Dept. of Electronic Engineering, University of London, C4DM-TR-08–01: 1–4.

Taljaard, P.C., J.N. Khumalo, and Sonja E. Bosch. 1991. *Handbook of siSwati*. Pretoria: J.L. van Schaik.

Taljaard, Petrus C., and Jannie W. Snyman. 1989. *An Introduction to Zulu Phonetics*. Constantia: M. Lubbe.

Traill, Anthony. 1985. *Phonetic and Phonological Studies of !Xóõ Bushman*. Hamburg: Helmut Buske.

Traill, Anthony. 1991. "Pulmonic control, nasal venting, and aspiration in Khoisan languages." *Journal of the International Phonetic Association* 21(1): 13–18.

Traill, Anthony. 1993. "The feature geometry of clicks." *Linguistica: Festschrift E. B. van Wyk, 'n Huldeblyk*, ed. Paul Michael Siegfried von Standen. Pretoria: J.L. van Schaik, pp. 134–140.

Vaz, Colin, Vikram Ramanarayanan, and Shrikanth S. Narayanan. 2013. "A two-step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis." *Proceedings of the 14th Annual Conference of the International Speech Communication Association*. Lyon, France.

Vossen, Rainer. 1997. "What click sounds got to do in Bantu." *Human Contact through Language and Linguistics*, ed. Birgit Smeja and Meike Tasch. Frankfurt: Peter Lang, pp. 353–366.

Westphal, Ernst O.J. 1971. "The click languages of Southern and Eastern Africa." *Linguistics in Sub-Saharan Africa*, ed. Jack Berry and Joseph H. Greenberg. Berlin: Mouton de Gruyter, pp. 367–420.

Zhu, Yinghua, Yoon-Chul Kim, Michael I. Proctor, Shrikanth Narayanan, and Krishna S. Nayak. 2013. "Dynamic 3D Visualization of Vocal Tract Shaping during Speech." *IEEE Transactions on Medical Imaging* 32(5): 838–848.

Ziervogel, Dirk. 1952. *A Grammar of Swazi* (*siSwati*). Bantu grammatical archives, vol. 3. Johannesburg: Witwatersrand University Press.