Proc. InterSing, 2010
Tokyo, Japan

Para-Linguistic Mechanisms of Production in Human 'Beatboxing': a Real-time Magnetic Resonance Imaging Study

Michael I. Proctor[1,2], *Shrikanth Narayanan*[1,2], *Krishna Nayak*[1]

[1]Viterbi School of Engineering, University of Southern California, USA
[2]Department of Linguistics, University of Southern California, USA
mproctor@usc.edu

Abstract

Real-Time Magnetic Resonance Imaging was used to examine mechanisms of sound production in an American male beatbox artist. The subject's repertoire was found to include percussive elements generated using a wide range of articulatory configurations, and three of the four airstream mechanisms normally observed in human speech production: pulmonic egressive, glottalic egressive, and lingual ingressive. In addition, pulmonic ingressive production were observed, which appears to be used strategically as a means of managing breathing during extended beatbox performance. The data offer insights into the para-linguistic use of articulatory gestures, and the ways in which they are coordinated in musical performance.
Index Terms: human beatbox, percussion, MRI, airstream mechanisms, articulation, coordination

1. Introduction

Human beatboxing is a performance art in which the vocal organs are used to produce a range of percussive sounds, usually an accompaniment to lyrics spoken, rapped or sung at the same time, sometimes by the same artist. Because it is a relatively young vocal artform, beatboxing has not been extensively studied, either in the musical performance or speech science literature.

Acoustic properties of some of the sounds used in beatboxing have been described impressionistally and compared to speech sounds [1]. Tyte has surveyed the range of sounds exploited by beatbox artists [2], and along with Splinter [3], has outlined a system of notation ('Standard Beatbox Notation': SBN) to formally describe beatbox performance. In the only phonetic study of beatboxing to date (to our knowledge), Lederer conducted spectral analyses of three of the most common percussive elements produced by human beatbox artists, and compared these, using twelve acoustic metrics, to equivalent electronically-generated sounds [4].

Although these studies have laid the foundation for a formal analysis of beatboxing performance, the actual mechanisms of production of human-simulated percussion effects are poorly understood, as they have not been examined using articulatory data. Furthermore, it is not well understood how artists are able coordinate linguistic and para-linguistic articulations so as to create the perception of multiple percussive layers, and synchronous speech and accompanying percussion.

2. Goals

The goal of the current study is to describe the articulatory phonetics involved in beatbox performance. Specifically, we make use of dynamic imaging technology to:

i. document the range of percussion sound effects in the repertoire of a beatbox artist
ii. examine the means of production of each of these elements, describing them in phonetic terms where possible
iii. examine the range of airstream mechanisms used in beatbox performance

3. Method

3.1. Participant

The study participant was a 27 year old male professional singer from Los Angeles, a practioner of a wide variety of vocal styles including soul, dancehall and hip-hop, who had been working as an MC in an rap duo since 1995. The subject is a native speaker of African American English, fluent in Spanish, and raps in both English and Spanish.

3.2. Corpus

The participant was asked to demonstrate the range of his beatboxing repertoire by performing in short intervals, as he lay supine in an MRI scanner bore. Forty recordings were made, each lasting between 20 and 40 seconds, of a variety of individual percussion sounds, composite beats, rapped lyrics, sung lyrics, and freestyle combinations of these elements. Each repeatable rhythmic sequence (SBN: *'grove'*) was elicited three times, at slow ($\sim$ 88 beats per minute), medium ($\sim$ 95 b.p.m.) and fast ($\sim$ 104 b.p.m.) rates.

3.3. Image and Audio Acquisition

All data were acquired using a Real-Time Magnetic Resonance Imaging (RT-MRI) protocol developed specifically for the dynamic study of speech production [5]. The subject's upper airway was imaged in the midsagittal plane using a gradient echo pulse sequence on a conventional GE Signa 1.5 T scanner. MR Image data were acquired at a rate of 9 frames per second, and reconstructed into video sequences with a frame rate of 20.8 f.p.s. using a gridding reconstruction method [6].

In-scanner audio recordings were acquired at a sampling rate of 20 kHz, using a custom ceramic noise-canceling microphone system [7], then reintegrated with the reconstructed MR-Imaged video. The resulting data provides dynamic midsagittal audio-visualization of the performer's entire vocal tract, from the upper trachea to the lips, including the nasal cavity.

3.4. Articulatory Analysis

MR image sequences were examined to determine the means of production of each of the percusive elements in the subject's repertoire. The coordination of glottal and supraglottal gestures was examined to provide insights in the airstream mechanisms exploited by the artist to produce different effect.

4. Results and Analysis

Fifteen phonetically distinct percussion effects were observed in this performer's repertoire, summarised in Table 1. For each effect, the performer's description is first listed, along with a description in SBN,[1] the International Phonetic Alphabet (IPA) notation for the closest equivalent sound, and the primary airstream mechanism used to produce it. The articulatory characterization of each of these sounds is described in detail in Sections 4.1–4.4.

EFFECT	SBN	IPA	AIRSTREAM	
Kick drum	|b|	[pʼ]	glottalic	egressive
Kick punchy	|pf|	[p͡fʼ]	glottalic	egressive
Kick 808	|8|	[p̪ʼ]	glottalic	egressive
Snare drum	|k|	[kxː]	pulmonic	egressive
Snare clap	|k|	[kʼ]	glottalic	egressive
Snare meshed	|ksh|	[kʃː]	pulmonic	egressive
Snare click	|tch|	[q‡]	lingual	ingressive
Clave click	|cc|	[k!]	lingual	ingressive
Hi-hat open K	|ks|	[ks]	pulmonic	egressive
Hi-hat open T	|ts|	[tsː]	pulmonic	in/egressive
Hi-hat closed T	|t|	[ts]	pulmonic	in/egressive
Hi-hat kiss	|ˆth|	[k|]	lingual	ingressive
Hi-hat breathy	|h|	[hː]	pulmonic	in/egressive
Cymbal t	|tsh|	[tʃː]	pulmonic	in/egressive
Cymbal h	|h|	[xː]	pulmonic	in/egressive

Table 1: *Classification of beatboxing effects in repertoire of study subject.*

4.1. Articulation of Kick Drum Effects

A variety of kick drum effects were demonstrated by the subject, all of which were produced as bilabial ejectives. The canonical effect, denoted |b| in SBN, was articulated as a bilabial ejective stop [pʼ]. Four frames illustrating the production of this sequence, captured over a 500 msec interval, are shown in Fig. 1.[2]

Laryngeal lowering and lingual retraction commences approximately 380 msec before the acoustic release burst; labial approximation commences 210 msec before the burst. Glottal closure is clearly evident after the larynx achieves the lowest point of its trajectory (frame 233). Rapid upward laryngeal movement after glottal adduction results in motion blurring (frame 235). Mean upward vertical displacement of the glottis during ejective production, measured over four tokens, was 30.75 mm. In the case of the canonical (unreleased) kick-drum effect |b|, the glottis remains adducted until well after the end of the ejective.

Other than |b|, the subject controlled two variant kick drum effects: an '808 kick' |8|, which was produced as an unreleased ejective stop in which the tongue remained retracted ([p̪ʼ]); and a 'punchy kick' |pf|, produced as a bilabial affricate ([p͡fʼ]). Four frames acquired over a 430 msec interval during the production of a |pf| token are shown in Fig. 2. The articulatory sequencing is the same as that used to produce |b|, except that the glottis is openned immediately after the laryngeal raising gesture (frame 103: approximately 160 msec after the beginning of the acoustic release burst).

4.2. Articulation of Snare Drum Effects

Two of the snare drum effects demonstrated by the study subject were realized as pulmonically-generated sequences of a velar stop followed by a sustained fricative. Four frames illustrating the production of the basic snare effect |k|, acquired over a 432 msec interval, are shown in Fig. 3. The data reveal that the effect is produced with a dorsal gesture articulated with varying degrees of constriction against the soft palate, suggesting that this sound is best characterized as a velar africate [kxː].

The meshed snare effect |ksh| was realized with the same initial velar stop, but was followed by a sustained post-alveolar sibilant fricative. Four frames acquired during the production of a token of |ksh| are shown in Fig. 4. As in Fig. 3, glottal abduction is evident throughout the production – a clear airway can be seen extending from the upper trachea into the lower pharynx in all frames – demonstrating that both snare drum effects are produced as pulmonic egressives.

A third snare drum variant – the 'snare clap' – was produced at the same primary place of articulation as the pulmonically-generated snare effects, but was generated as an ejective. As with the kick drum effects, laryngeal lowering preceeds glottal closure (Fig. 5, frame 156), before rapid upward movement of the larynx expels the air in the pharynx out past the velar constriction (frame 158).

4.3. Articulation of Click Effects

A number of percussion effects – claves, woodblocks and cowbells – are simulated in beatboxing performance by using lingual ingressive sounds, or clicks. Three different effects were produced as clicks by the subject in this study: a clave |cc|, a closed hi-hat |ˆth|, and a snare drum variant |tch|.

A 336 msec MRI sequence illustrating the production of a 'clave click' effect |cc| is shown in Fig. 6. The data reveal that the clave is realized as a velar-alveolar click [k!]: a complete lingual constriction is made against the roof of the mouth between the velar and alveolar regions, and lingual release commences with the tongue blade and the anterior part of the dorsum (frame 18), creating an ingressive airstream mechanism. The result of this articulation is a very short sound with rapid attack and decay, which effectively simulates the sound of a struck woodblock.

Another effect realized as a click by this artist was a snare drum variant |tch|. In contrast to the rapid transient of the alveolar clave click (Fig. 7, left), the subject produced a more affricate-like sound of longer duration, to simulate the sustained response of a snare drum (Fig. 7, right).

A 400 msec sequence illustrating the production of a 'snare' click |tch|, is shown in Fig. 8. At the point of release (frames

[1] Standard Beatbox Notation uses square bracket delimiters; however, vertical bars will be used throughout this document to denote effects written in SBN – e.g. |pf| – to avoid confusion with IPA notation – e.g. [pfʼ].

[2] In figures showing MR Image sequences, frame numbers are given in parentheses in the figure caption. For the video reconstruction rate of 20.8 f.p.s. used in this data, frame duration is approximately 48 msec.

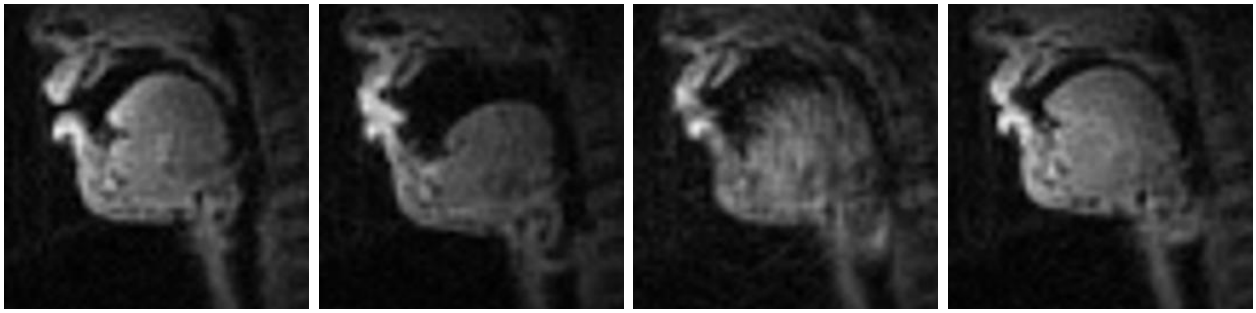

Figure 1: *Articulation of a kick drum effect* |b| *as a bilabial ejective stop* [p']. f228: intial lingual posture before labial closure; f233: lowered larynx and glottalic adduction; f235: rapid laryngeal raising; f238: lingual advancement as glottis remains closed.

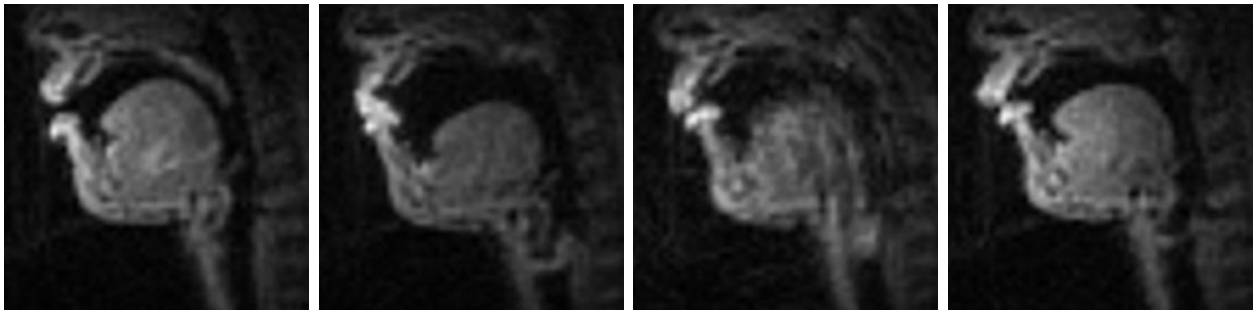

Figure 2: *Articulation of a 'punchy kick drum' effect* |pf| *as an affricated bilabial ejective* [pf']. f94: before labial closure; f98: lowered larynx, glottalic closure; f100: rapid laryngeal raising; f103: glottal abduction.

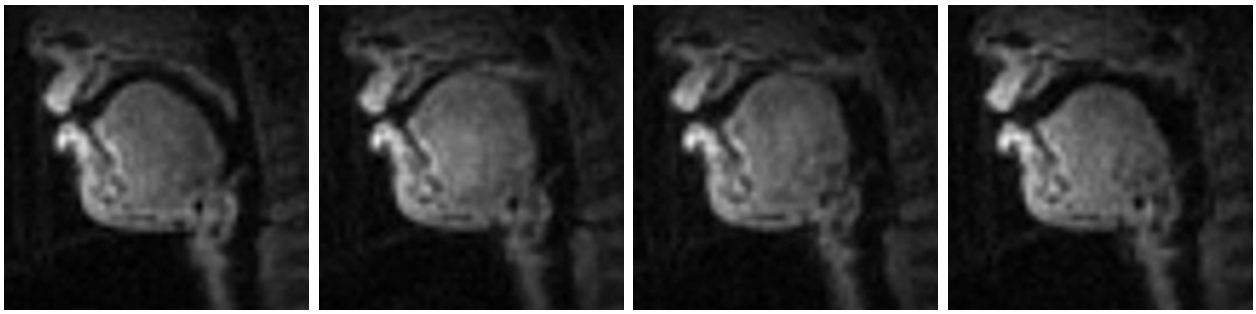

Figure 3: *Articulation of a snare drum effect* |k| *as a pulmonic egressive velar affricate* [kx:]. f86: initial resting posture; f90: dorsal closure; f92: sustained critical dorsal constriction; f95: final resting posture.

86, 88), it can be seen that both anterior ([ǂ]) and posterior ([q]) lingual constrictions are more retracted than those observed in the velar-alveolar click (c.f. Fig. 6), and that a greater area of the rear of the tongue dorsum appears to come into contact with the uvula. The data suggest that the subject is using a palatal-uvular click [qǂ] with a delayed, or possibly lateralized release, to create the acoustic contrast observed in Fig. 7. This analysis is consistent with Ladefoged's characterization of palatal and lateral clicks as having an instrinsically an affricated release [8].

The final type of click demonstrated by the study participant was described as a 'closed hi-hat kiss'. Although the artist perceived this sound to be an implosive |ˆth|, the image sequence in Fig. 9 demonstrates that the effect was articulated as a dental click [kǀ]. At the point of lingual release (frame 380), the anterior constriction [ǀ] can be seen to be more advanced than that observed in the alveolar click (c.f. Fig. 6). As in the case of the palatal click, the release was noticeable more affricated and less abrupt than in the clave effect.

4.4. Articulation of Pulmonically-driven Hi-hats and Cymbals

In addition to the |ˆth| effect just examined, the study participant uses another four sounds to emulate hi-hats, and two different sounds to emulate cymbals. All of these effects were produced as fricatives, affricates, or stop-fricative clusters. All of these pulmonically-generated effects were demonstrated with both egressive and ingressive airstreams.

Two 'open hi-hat' effects were produced as stop-initial sequences terminating in sustained apical-alveolar fricatives: one dorsal-inital (|ks|), and one coronal-initial (|ts|). Three frames acquired during the production of a |ks| token are shown in Fig. 10. The contrastive |t| 'closed hi-hat' effect was articulated as shorter affricate produced at the same place on the alveolar

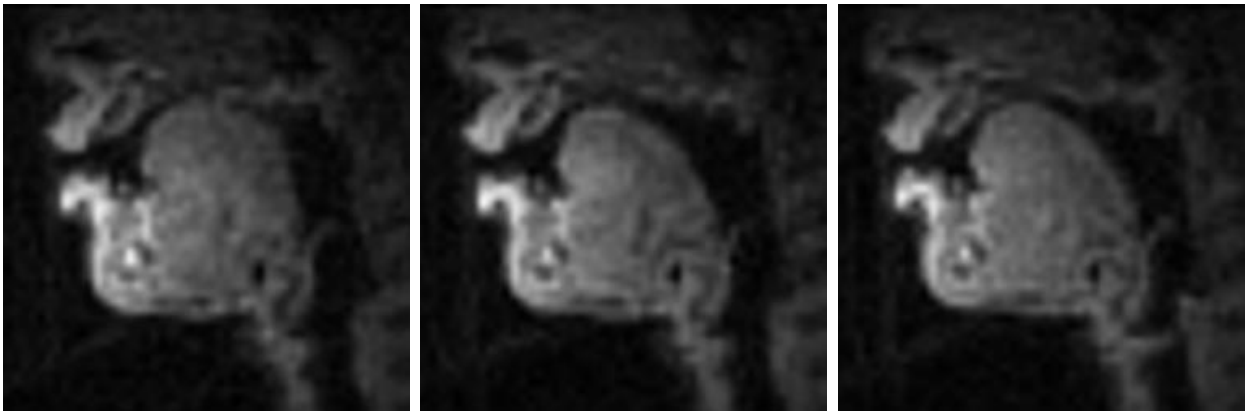

Figure 4: *Articulation of a meshed snare drum effect* |ksh| *as a velar stop-post-alveolar sibilant cluster* [kʃː]. f93: dorsal closure; f95: dorsal-palatal transition; f95: sustained critical post-alveolar constriction.

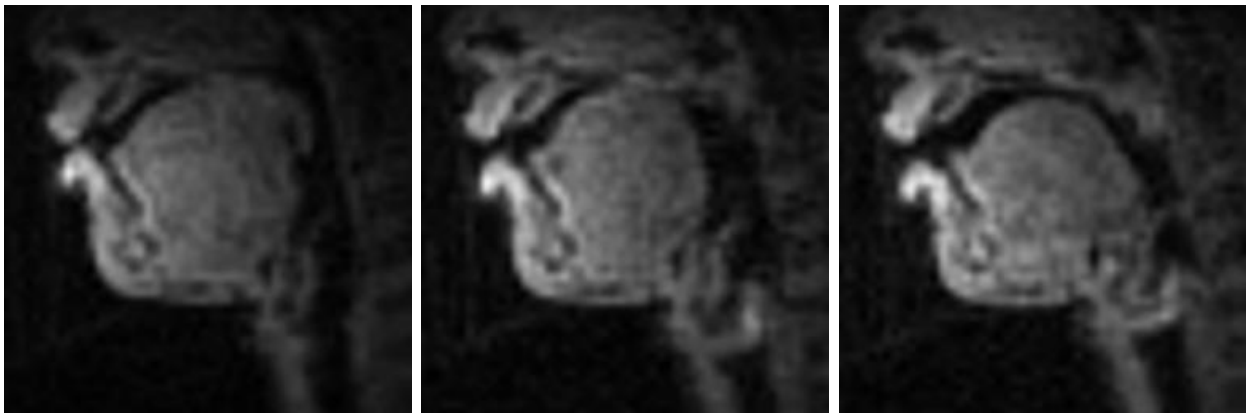

Figure 5: *Articulation of the 'snare clap' effect as a velar ejective* [kʼ]. f152: laryngeal lowering and velar closure; f156: glottal closure; f158: laryngeal raising and dorsal release.

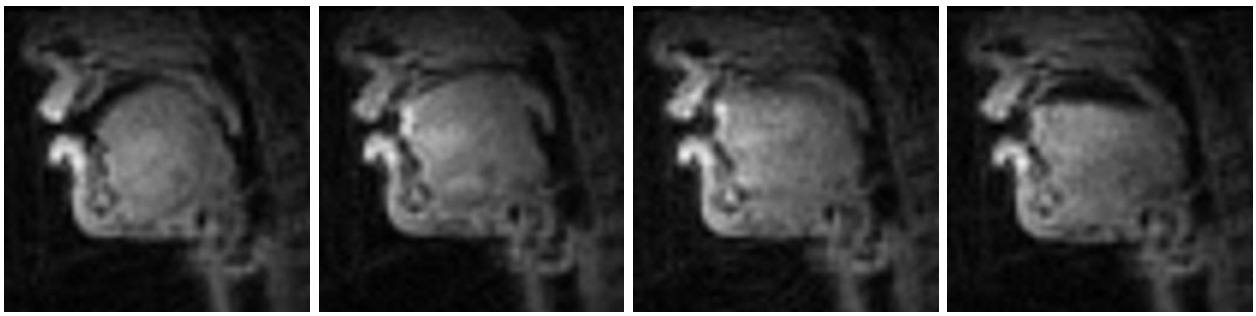

Figure 6: *Articulation of a clave effect* |cc| *as a velar-alveolar click* [k!]. f13: immediately before velar closure; f15: alveolar closure; f18: lingual release; f20: final resting posture.

ridge, and a 'breathy hih-hat' |(ˆ)h| was produced as a glottal fricative [hː], ingressively as well as egressively.

The final two sounds in the percussive repertoire of the experimental subejct were two cymbal effects: a type of crash or ride cymbal |(ˆ)tsh|, articulated as a coronal affricate [tʃː], and a muted sound |(ˆ)h|, described as a 'k cymbal', which was realized as sustained velar fricative [xː].

5. Discussion

This work represent a first step towards the formal study of the paralinguistic articulatory phonetics underlying an emerging vocal performance art. Because beatboxing is a highly individualised artform, examination of the sound effect repertoires of other beatbox artists would be an important step towards a more comprehensive understanding of the articulatory mechanisms involved in producing these sounds.

Highly skilled beatbox artists, such as Rahzel, are capable

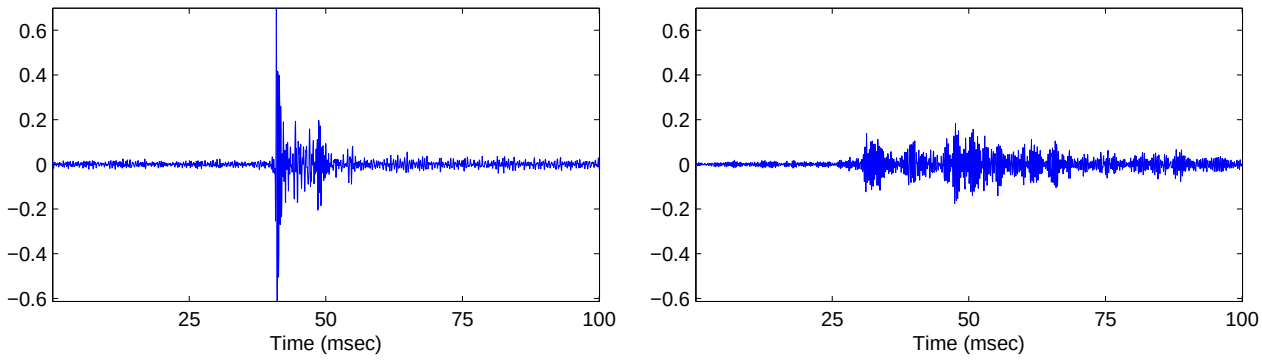

Figure 7: *Acoustic waveforms of two effects articulated as clicks: clave* |cc| *(left), and snare* |tch| *(right).*

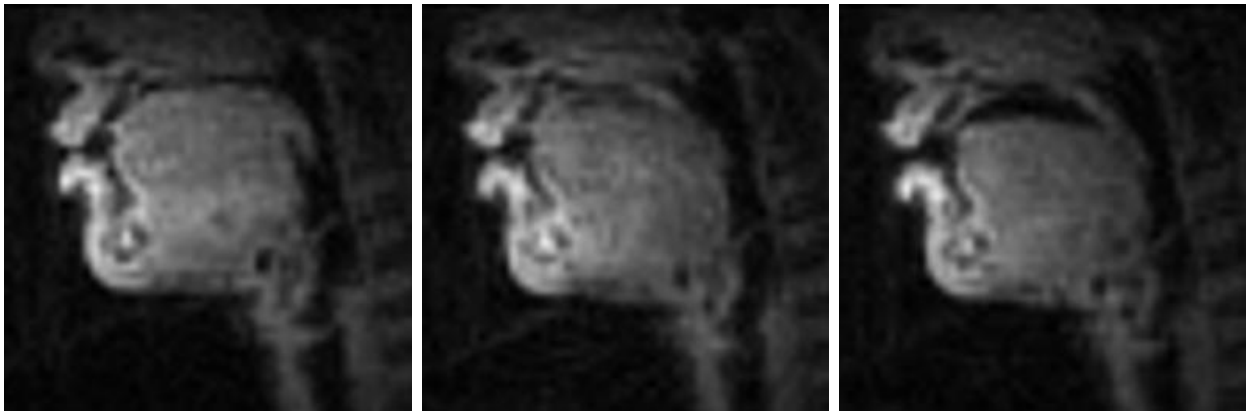

Figure 8: *Articulation of a snare effect* |tch| *as a uvular-palatal click* [q‡]. f84: palatal-uvular closure; f86: lingual release; f88: final resting posture.

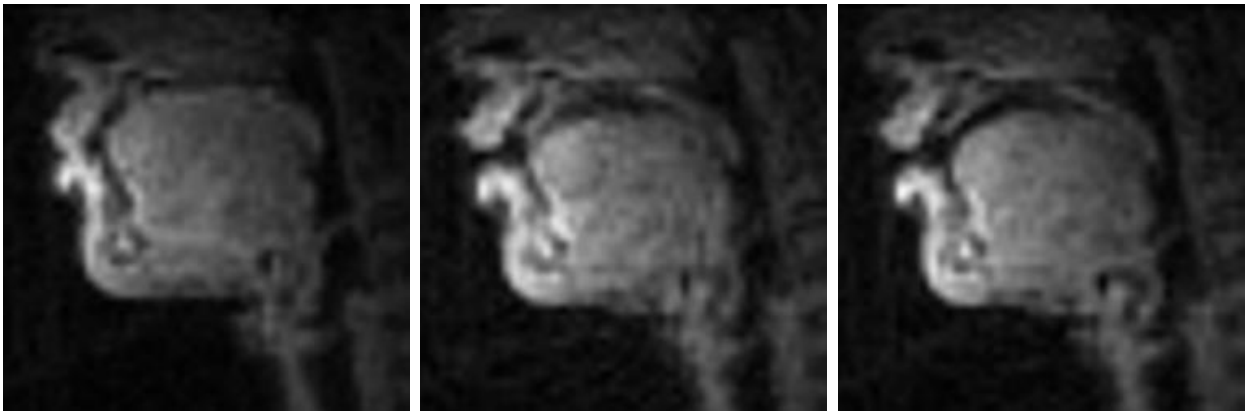

Figure 9: *Articulation of a 'closed hi-hat kiss' as a velar-dental click* [k|]. f377: dental-velar closure; f380: lingual release; f381: final resting posture.

of performing in a way which creates the illusion that the artist is simultaneously singing and providing their own percussion accompaniment, or simultaneous beatboxing while humming [1]. Such illusions raise important questions about the relationship between speech production and perception, and the mechanisms of perception which are engaged when a listener is presented with simultaneous, but incompletely realised, speech and music signals. It would be of great interest to study this type of performance using MR Imaging, to examine the ways in which linguistic and paralinguistic gestures can be coordinated.

5.1. Future Directions

Further insights into the mechanics of human beatboxing would be gained through the use of additional MR Imaging planes. Since many beatbox effects make use of non-pulmonic

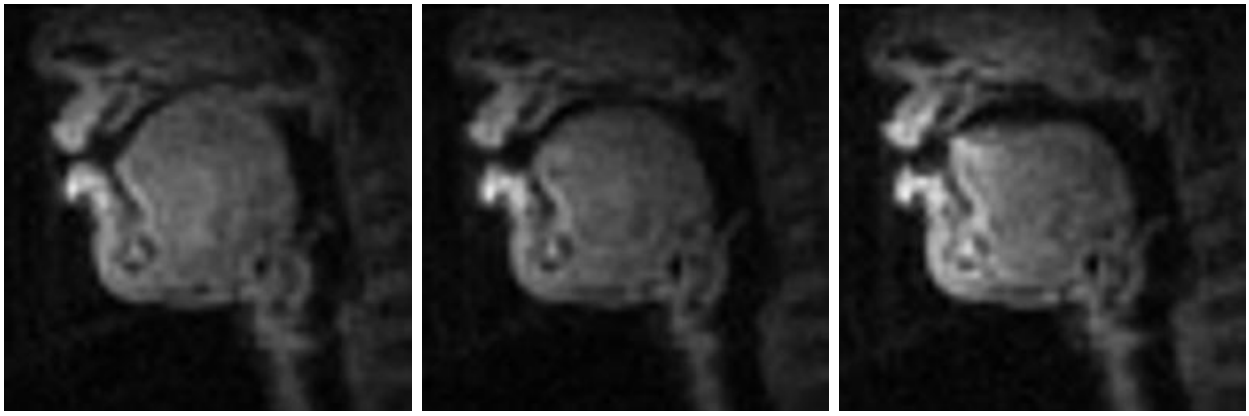

Figure 10: *Articulation of an open hi-hat effect* |ks| *as a dorsal stop-coronal fricative cluster* [ks:]. f69: dorsal closure; f71: dorsal-coronal transition; f73: sustained critical alveolar constriction.

airstream mechanisms, axial imaging could provide additional detail about the articulation of the larynx and glottis during ejective production.

Because clicks carry a high functional load in the repertoire of many beatbox artists, high-speech imaging of the hard palate palate region would be particularly useful. The strategic placement of coronal imaging slices would provide additional phonetic detail about lingual coordination in the mid-oral region. Lateral clicks, which are exploited by many beatbox artists [2] can only be properly examined using coronal or parasagittal slices, since the critical articulation occurs away from the mid-sagittal plane.

6. Conclusion

An approach to studying the phonetics of beatboxing has been outlined. The use of Real-Time Magnetic Resonance Imaging has been shown to be a viable method with which to examine the repertoire of a human beatboxer, affording novel insights into the mechanisms of production of the imitation percussion effects which characterized this performance style. The data reveal that beatboxing performance involves the use of the full range of airstream mechanisms found in human languages, as well as the strategic use of ingressive pulmonic airflow to minimize interruptions to the vocal delivery due to breathing. The study of beatboxing performance has the potential to provide important insights into articulatory coordination in speech production, and mechanisms of perception of simultaneous speech and music.

7. Acknowledgements

Research supported by NIH Grant R01 DC007124-01.

8. References

[1] D. Stowell and M. D. Plumbley, "Characteristics of the beatboxing vocal style," Dept. of Electronic Engineering, Queen Mary, University of London, Technical Report, Centre for Digital Music C4DM-TR-08-01, 2008.

[2] G. Tyte, "Beatboxing techniques," 2010. [Online]. Available: www.humanbeatbox.com

[3] M. Splinter and G. Tyte, "Standard beatbox notation," 2010. [Online]. Available: www.humanbeatbox.com

[4] K. Lederer, "The phonetics of beatboxing," Ph.D. dissertation, 2005. [Online]. Available: http://www.humanbeatbox.com/phonetics

[5] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *JASA*, vol. 115, no. 4, pp. 1771–1776, 2004.

[6] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *Signal Processing Magazine, IEEE*, vol. 25, no. 3, pp. 123–132, May 2008.

[7] E. Bresch, J. Nielsen, K. Nayak, and S. Narayanan, "Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans," *J. Acoust. Soc. Am*, vol. 120, no. 4, pp. 1791–1794, 2006.

[8] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Oxford: Blackwell, 1996.