NON-PULMONIC INITIATION IN HUMAN BEATBOXING: A REAL-TIME MRI STUDY

Alexis Dehais-Underdown[1], Paul Vignes[1], Lise Crevier-Buchman[1,2], Didier Demolin[1], Pierre-André Vuissoz[3], Karyna Isaieva[3], Marc Fauvel[4], Yves Laprie[5], Jacques Felblinger[3,4]

[1]Laboratoire de Phonétique et Phonologie (UMR 7018 - CNRS/Sorbonne-Nouvelle), [2]Unité Voix, Parole Déglutition, Service ORL et de Chirurgie de la Face et du Cou, Hôpital Foch, [3] IADI, INSERM-U947, Université de Lorraine, Nancy, France, [4] CIC-IT 1433, INSERM, Université de Lorraine & CHRU Nancy, Nancy, France, [5] LORIA (CNRS/Inria), Université de Lorraine, Nancy, France

alexis.dehais-underdown@sorbonne-nouvelle.fr

ABSTRACT

Human Beatboxing (HBB) is a musical technique produced by vocal tract movements. Beatboxers seem to exploit vocal tract capacities to their full extent. Thus, HBB is a good paradigm to study production mechanisms. This study uses aerodynamic and laryngoscopic along with 2D MRI data to analyze non-pulmonic initiation mechanisms during HBB production. The production of 4 professional beatboxers was analyzed while producing glottalic and lingual mechanisms. Results shows that glottalic mechanisms result from a sequence of laryngeal vertical movement, tongue root maneuvers and, sometimes, additional pharyngeal gestures to increase or decrease pressure. Lingual mechanisms were produced by front-back tongue movements to generate egressive or ingressive airflows. A reformulation of Catford model of initiation mechanisms, in terms of initiatory gestures rather than aerodynamic parameters, is proposed.

Keywords: Human Beatboxing, Laryngoscopy, Initiation, rt-MRI, Aerodynamics

1. INTRODUCTION

This study aims to describe Human Beatboxing (HBB) production in a similar way to Proctor *et al.* [1] and Blaylock *et al.* [2]. This study focused on the production of non-pulmonic initiation mechanisms. Contrary to Proctor's and Blaylock's study, this research on HBB production was not originally based on rt-MRI data but on aerodynamic, laryngoscopic and acoustic data. MRI was acquired to confirm our findings in the physiological and acoustic data of 4 professional beatboxers [6]. Both [1] and [2] offer a formal description of HBB production based on MRI. While both studies report a wide range of production mechanisms, Proctor's study focuses on the "paralinguistic use" of production mechanisms found in the world's languages (e.g. [p'] [ts'] [k']) and Blaylock points out that beatboxers not only make use of mechanisms found in phonetic inventories but also mechanisms that are unknown to phonetic inventories (e.g. [↓ʈː] [↓ʁ̥ˡ]). Thus, HBB constitutes a good paradigm to investigate the extent of the human vocal tract capacities.

Initiation is the phase that sets air in motion. Airflow is generated by increasing or decreasing of the volume between the place of initiation and the place of articulation. Table 1 shows Catford's aerodynamic model of initiation [3]. He defines initiation in terms of location (i.e. lungs, larynx, mouth) and directionality of airflow (i.e. egressive and ingressive). Compressive gestures produce positive pressure and generate egressive airflow: *pulmonic pressure*, *glottalic pressure* and *velaric pressure*. Rarefactive gestures produce negative pressure and ingressive airflow: *pulmonic suction*, *glottalic suction* and *velaric suction*. Velaric initiation is more commonly referred as *lingual*. Hereafter, *lingual* will be used over *velaric*.

Table 1: Catford's initiation parameters

	Direction	
Location	**Compressive**	**Rarefactive**
Lungs	pulmonic pressure	pulmonic suction
Larynx	glottalic pressure	glottalic suction
Mouth	velaric pressure	velaric suction

Non-pulmonic mechanisms (i.e. *glottalic* and *lingual*) have been extensively documented in the world's languages. Concerning the production of HBB, it was found that glottalic initiation is composed of a glottal closure, vertical laryngeal movements and tongue root maneuvers to increase or decrease the volume of the lower vocal tract [1, 4, 2, 5]. These studies also report an expansion of the lower vocal tract (i.e. tongue root protraction, laryngeal lowering, pharyngeal expansion) preceding the glottal closure and the laryngeal raising of ejectives. Lingual mechanisms in beatboxing can be produced either by increasing the volume between the anterior constriction and posterior constriction (i.e. click) or by decreasing the volume between the anterior constriction and posterior constriction [1, 2]. Furthermore, an aerodynamic study of [6] reported pressure values between $40hPa$, $100hPa$ for the production of glottalic beatboxed sounds. In comparison, during speech production, intraoral pressure of ejectives ranges from $20hPa$ for labial stops up to $40hPa$ for velar stops [7]. It suggests that beatboxers use a greater degree of compression and expansion to produce glottalic sounds. To further understand initiation and possible differences between beatboxing and speaking, we propose to investigate the production of glottalic and lingual production mechanisms with laryngoscopic and aerodynamic recordings of 4 professional beatboxers and additionnal 2D MRI data of one subject.

2. METHODS

2.1. Protocol and Corpus

The data presented here comes from a larger database of 4 professional beatboxers (3 males, 1 females). Fibroscopic and aerodynamic data were acquired first. To confirm observations from the physiological data (see [6]), a pilot study using real-time Magnetic Resonance (rt-MRI) with one subject was carried out. The subset of the corpus containing only non-pulmonic sounds is given in Table 2. Hereafter, transcriptions with curved brackets indicates sounds produced with lingual mechanism and down arrow indicates ingressive airflow (e.g. {↓ʙ̥}). Sounds were produced in isolation and in *Beatboxing Patterns*. Sounds in isolation were produced 8 times each. This study only report sounds produced in isolation. The analysis focuses on gestures implicated in raising and lowering the pressure.

2.2. Data Acquisition

Laryngoscopic data was acquired by the team's MD with a Xion video-stroboscopic system using a flexible fiberscope with a rate of 25 frames per second. The acoustic signal was acquired through the fiberscope microphone. A 2% Xylocaine anesthesia was administered before the endoscope was inserted through the nose into the pharyngeal cavity above the supraglottic plan. In this session, subjects were not asked to produced lingual mechanisms. Aerodynamic and acoustic signals were acquired with an EVA2 Workstation [8] allowing simultaneous recording of acoustic signal, intraoral pressure (measured in hPa, $1hPa = 1.02cmH_2O$), oral airflow and nasal airflow (measured in dm^3/s). Intraoral pressure (Po) was obtained by inserting a small tube into the pharynx through the nasal cavity. Pressure for clicks is very difficult to measure because one would need to measure it between the two constrictions. Thus, the measured pressure corresponds to the pharyngeal pressure. In the experiment, beatboxers were asked to produce continuous humming while producing lingual mechanisms, so the pressure is 0 and nasal airflow is positive.

The 2D MRI data was recorded at Nancy Central Regional University Hospital with a Siemens Prisma 3T scanner, Erlangen, Germany. The speaker was in supine position and a Siemens Head/Neck 64 coils was used. For the 2D real-time we used radial RF-spoiled FLASH sequence with TR = 2.22 ms, TE = 1.47 ms, FoV 192x192mm, flip angle = 5°, and slice thickness was 8 mm. Pixel bandwidth was 1670 Hz/pixel. Image size was 136x136, in-plane resolution was 1.6 mm, recorded at 50 fps and reconstructed with a nonlinear inverse technique presented in [9]. Audio was recorded at a sampling frequency of 16 kHz inside the MRI scanner with a FOMRI III optoacoustics fibre-optic microphone (FOMRI III, Optoacoustics Ltd., Mazor, Israel).

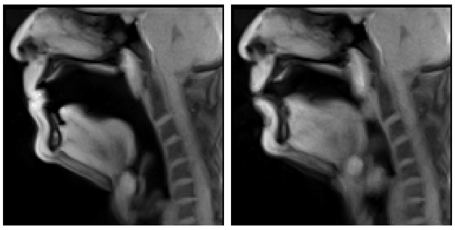
Table 2: Corpus of non-pulmonic sounds of our database.

Glottalic	*Stop*	p'ɓ̚ʙ̥		t' ɗ̚	
	Affricate		pf'	ts' tʃ'	ʛʟ
Lingual	*Stop*	p			
	Affricate		pf	ts	↓kʟ
	Trill	↓ʙ̥			

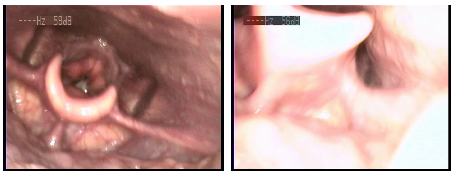
3. RESULTS

3.1. Glottalic mechanisms

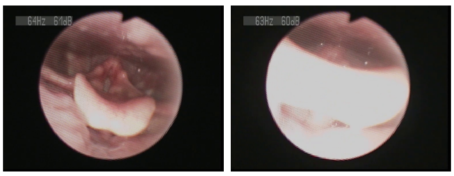
The Figure 1 illustrates the production of [p'] with MRI frames as well as laryngoscopic images at the beginning and at the end of the initiatory phase. Laryngeal behavior in the laryngoscopic data for glottalic egressive sounds were highly similar across subjects and sounds. The mean pressure for glottalic sounds ranges from 20hPa to 85hPa depending on the subject and the sound. Fibroscopic images show a systematic behavior across subjects and sounds: laryngeal lowering and pharyngo-laryngeal expansion preceding a glottal closure and, simultaneously, laryngeal elevation and tongue root retraction. Two subjects added additional aryepiglottic constriction (i.e. approximation of the arytenoids on top of the root of the epiglottis) and one subject added pharyngeal compression following laryngeal elevation and tongue root retraction. MRI images confirms the systematic tongue root maneuver and laryngeal elevation observed for all subjects. Furthermore, the MRI data suggests that tongue root retraction is voluntarily recruited to decrease the volume of air between the glottis and the lips.



(a) Subject VP



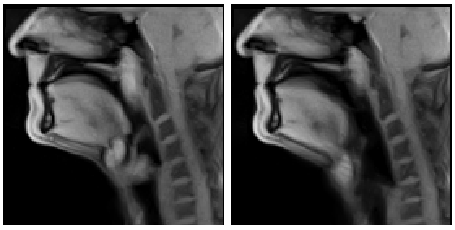
(b) Subject VP



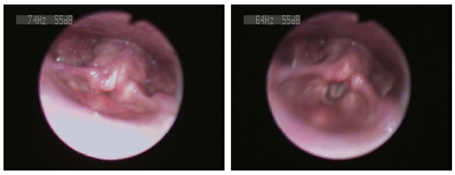
(c) Subject AI

Figure 1: MRI frames (top) and laryngoscopic images (middle and bottom) of the glottalic mechanism produced for [pʼ].

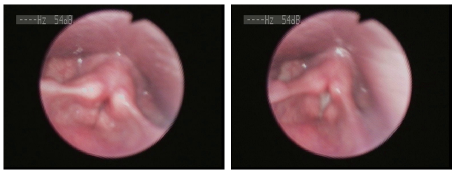
Figure 2 illustrates the production of [ɓ̚]. Once again, individual behaviors were very similar. The mean pressure for [ɓ̚] ranges −5hPa to −25hPa and −50hPa −75hPa for [ɓ̥] (produced only by 2 subjects). Concerning [ɗ̚] and [ʛʟ] the pressure signal clipped. The fibroscopic frames show laryngeal raising and pharyngo-laryngeal narrowing preceding a ventricular closure (i.e. a closure of both vocal and ventricular folds) and laryngeal lowering and tongue root protraction. One subject added additional aryepiglottic constriction. 2 subjects added pharyngeal expansion following laryngeal lowering. MRI data shows the same phases confirming once more the active role of tongue root during the initiation.



(a) Subject VP



(b) Subject CJ



(c) Subject GA

Figure 2: MRI frames (top) and laryngoscopic images (middle and bottom) of the glottalic mechanism produced for [b].

3.2. Lingual mechanism

Figure 3 illustrates the egressive lingual mechanism for {pf}. MRI frames shows the formation of two constrictions in the vocal tract. We found anterior closure to be bilabial, labiodental, dental and alveolar. The posterior constriction is formed in the velar or uvular region. To generate airflow for the production of lingual egressive sounds, the tongue dorsum make a frontward movement to compress the volume of air between the posterior and the anterior constriction. For all sounds, except {↓kʟ}, the front constriction was released first. For {↓kʟ} the velar/uvular constriction releases first and was considered as lateral velar based on MRI, acoustic data and the subject's description.

Figure 4 illustrates the the production of {↓ʙ̥ˡ}. Lingual ingressive initiation is well-known for the production of clicks. Here, a similar mechanism occurs: a labial and a velar constriction are created

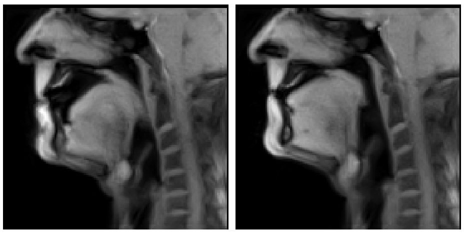


Figure 3: MRI frames of the lingual mechanism produced for {pf} - Subject VP

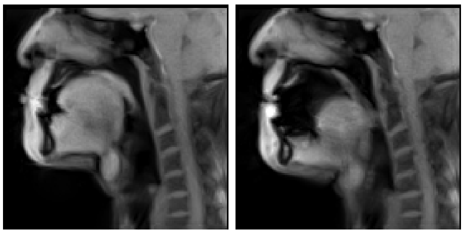


Figure 4: MRI frames of the lingual mechanism produced for {↓ʙ̥^{l}} - Subject VP

and the tongue retracts to expand the volume which results in an ingressive airflow. Ingressive bilabial trills in HBB are lateralized. Beatboxers were unable to produce central ingressive bilabial trilling.

4. DISCUSSION

The data presented here concurs with previous studies of initiation mechanisms during HBB production. For the ejective production, our data support the findings reported in [1, 4, 5, 2], that is expansion of the lower vocal tract previous to compression of the volume of air between the glottis and the place of articulation by laryngeal elevation and tongue root retraction. Also, lingual mechanisms support the findings of [1, 2] concerning tongue maneuver to produce either egressive or ingressive airflow. Lingual trills are not attested in any phonological system but are attested in beatboxing [2, 6]. Additional tongue root maneuvers, in coordination with vertical movement of the larynx, to raise or lower the pressure during the production of glottalic mechanisms has been hypothesized by Kingston [10]. It has been observed on MRI data for ejectives in Tigrinya [11] and implosives in Hendo [12]. Our data confirms that tongue root maneuvers play a key role in raising or lowering the pressure. Pharyngeal expansion or compression may also participate in decreasing or increasing the pressure.

What does it mean for a sound to be "*glottalic*"? Originally, the term was used to differentiate ejectives and implosives from *glottalized consonants* (see Catford's discussion [3], p.247-8, Ch. 5 notes 1, 2). Both "*glottalic*" and "*glottalized*" referred to glottal or laryngeal maneuvers implicated in consonant production. A *glottalic consonant* is now understood (and taught) as a consonant produced with a closed glottis and vertical movements of the larynx. Though, both linguistic and HBB data on ejectives and implosives suggests that this initiation mechanism cannot merely be reduced to vertical movements to raise or lower the pressure. Therefore, we propose to use the term "*laryngeal*" instead of "*glottalic*". Catford's model of initiation is based on aerodynamic parameters (i.e. pressure and airflow direction) to distinguish between initiatory mechanisms. Although the model is motivated, the aerodynamic phase is the result of vocal tract maneuvers to change the pressure. We propose to reformulate Catford's model in terms of initiatory gestures. By analogy to places and modes of articulation, Table 2 proposes a revision of the model with the 3 places of initiation from the original model and 2 modes of initiation. The mode of initiation refers to gestures that either compress or expand the volume between the place of initiation and the place of articulation. The phonetic description takes into account the place of initiation and the direction of the airflow resulting from compression (i.e. egressive) or expansion (i.e. ingressive).

Table 3: Reformulated initiation parameters

Location	Mode	Phonetic description
Lungs	Compression	Pulmonic Egressive
	Expansion	Pulmonic Ingressive
Larynx	Compression	Laryngeal Egressive
	Expansion	Laryngeal Ingressive
Mouth	Compression	Lingual Egressive
	Expansion	Lingual Ingressive

5. CONCLUSION

Human Beatboxing constitutes an original contribution to discuss production mechanisms with non-linguistic data. The reformulation of Catford's model in terms of initiatory gestures is motivated as it allows to make a direct link with constraints on sound patterns (e.g. clicks and the "*back constraint vowel*"). However, in terms of phonetic description, the reformulated model needs to be supported by an acoustic analysis to confirm the relevance of the phonetic features used to differentiate initiation mechanisms.

6. ACKNOWLEDGMENT

This physiological protocol was approved by an ethical committee (RCB-ID n° 2020-A00246-33). The MRI protocol was approved ethical protocol "METHODO" (ClinicalTrials.gov Identifier: NCT02887053, RCB-ID n°CPP EST-III, 08.10.01). The research is supported in France by the Delegation of Research and Innovation of Hôpital Foch and LABEX EFL (Empirical Fundation of Linguistics, ANR-10-LABX-0083) and the ArtSpeech project (ANR-15-CE23-0024) and supported by European founds CPER "IT2MP", "LCHN" and "FEDER".

7. REFERENCES

[1] M. Proctor, E. Bresch, D. Byrd, K. Nayak, and S. Narayanan, "Paralinguistic mechanisms of production in human "beatboxing": A real-time magnetic resonance imaging study," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 1043–1054, 2013.

[2] R. Blaylock, N. Patil, T. Greer, and S. S. Narayanan, "Sounds of the human vocal tract." in *INTERSPEECH*, 2017, pp. 2287–2291.

[3] J. C. Catford, *Fundamental problems in phonetics*. Bloomington And London: Indiana University Press, 1977.

[4] T. De Torcy, A. Clouet, C. Pillot-Loiseau, J. Vaissiere, D. Brasnu, and L. Crevier-Buchman, "A video-fiberscopic study of laryngopharyngeal behaviour in the human beatbox," *Logopedics Phoniatrics Vocology*, vol. 39, no. 1, pp. 38–48, 2014.

[5] N. Patil, T. Greer, R. Blaylock, and S. S. Narayanan, "Comparison of basic beatboxing articulations between expert and novice artists using real-time magnetic resonance imaging." in *Interspeech*, 2017, pp. 2277–2281.

[6] A. Dehais-Underdown, P. Vignes, L. Crevier-Buchman, and D. Demolin, "In and out: production mechanisms in human beatboxing," *Proceedings of Meetings on Acoustics, Vol. 45, 060005*, 2022.

[7] D. Demolin, "Acoustic and aerodynamic characteristics of ejectives in amharic," *The Journal of the Acoustical Society of America*, vol. 115, no. 5, pp. 2610–2610, 2004.

[8] A. Ghio and B. Teston, "Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies," in *International Conference on Voice Physiology and Biomechanics*, 2004, pp. 55–58.

[9] M. Uecker, S. Zhang, D. Voit, A. Karaus, K.-D. Merboldt, and J. Frahm, "Real-time mri at a resolution of 20 ms," *NMR in Biomedicine*, vol. 23, no. 8, pp. 986–994, 2010.

[10] J. Kingston, "The phonetics and phonology of the timing of oral and glottal events," Ph.D. dissertation, University of California, Berkeley, 1985.

[11] Z. Hermes, M. Fu, S. Rose, R. Shosted, and B. Sutton, "Representations of place and airstream mechanism: A real-time mri study of tigrinya ejectives," in *Proceedings of the 15th Conference on Laboratory Phonology, Ithaca, USA*, vol. 133, 2016.

[12] D. Demolin, H. Ngonga-Ke-Mbembe, and A. Soquet, "Phonetic characteristics of an unexploded palatal implosive in hendo," *Journal of the International Phonetic Association*, vol. 32, no. 1, pp. 1–15, 2002.