ACOUSTICO-PHYSIOLOGICAL COORDINATION IN THE HUMAN BEATBOX: A PILOT STUDY ON THE BEATBOXED CLASSIC KICK DRUM.

Alexis Dehais Underdown[1], Lise Crevier Buchman[1,2], Didier Demolin[1]

[1]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS – Sorbonne Nouvelle), [2]Hopital Foch, Univ. VSQ.
alexis.dehais-underdown@sorbonne-nouvelle.fr

ABSTRACT

This paper presents a pilot study of physiological, acoustic and aerodynamic characteristics of the beatboxed classic kick drum by a single artist excerpted from our database of his beatboxing repertoire. We recorded 25 repetitions with synchronized aerodynamic (intraoral pressure, oral airflow) acoustic, electroglottographic and laryngoscopic data. Results show that the classic kick drum is produced either as a bilabial voiceless glottalic egressive plosive or as a bilabial voiceless glottalic egressive affricate. Laryngoscopic data show a glottal adduction, a slight supraglottal constriction and a laryngeal raising. We also observed an occlusion, a high amplitude burst and low frequencies in the spectrum. We discuss the coordination of the gestures involved in the classic kick production based on the multi-instrumental recording. Future analysis of the beatboxed database will help us understand beatboxing physiology and the extent of the vocal tract capacities.

Keywords: Human Beatbox, phonetics, aerodynamics, laryngoscopy, acoustics.

1. INTRODUCTION

The Human Beatbox (HBB) make use of the vocal tract in order to imitate musical sonorities. Few studies mention physiological aspects of this peculiar technique. MRI studies of beatboxing [1, 8, 10] give a description of basic and complex articulations of HBB production. Laryngoscopic studies of laryngeal behaviour during beatboxing [3, 11] show that beatboxers manage to recruit laryngopharyngeal articulators (i.e. pharynx constrictor, glottis, ventricular folds, aryepiglottic folds) separately and optimally. Understanding this technique would provide valuable knowledge about the extent of the vocal tract capacities' and its limits aside from its use for linguistic purpose.

This paper aims to describe physiological aspects of the beatboxed classic kick drum by a French beatboxer using multi-instrumental simultaneous recordings with aerodynamic, acoustic and laryngoscopic data. To our knowledge there is no existing study of the beatboxing vocal technique using multidimensional synchronous data. This multiparametric study was found useful to understand the physiology of the beatboxed classic kick drum, in particular the coordination of supralaryngeal and laryngeal gestures. The beatboxed classic kick drum is described [1, 3, 8, 10, 11] as a voiceless glottalic egressive bilabial produced as a plosive [p'] or an affricate [p'ɸ].

2. METHODS

The subject was a 35 y.o. male who volunteered for the experiments after signing an informed consent in accordance with the ethical committee of our hospital (N° 1922081). The subject was asked to perform his repertoire of beatboxed sounds. First he gave a description of the sounds, then produced them in isolation and finally, he combined the sounds in a beatboxed sequence of various sounds from his repertoire. For analytical purposes and to understand how the synchronisation of simultaneous parameters (articulation, aerodynamics, acoustics) is involved, we chose to focus only on the classic kick drum production.

2.1. Aerodynamic and electroglottographic data

Aerodynamic data was collected with the EVA2 workstation (SQLab-LPL, Aix en Provence, France) which allows synchronized aerodynamic, acoustic and electroglottographic (EGG) data collection. Intraoral pressure (Po), expressed in hectopascal (1 hPa = 1,02 cm H_2O), was obtained inserting a small tube into the mouth; we took care to place the tube at the corner of the lip so it would not interfere with Oral airflow. Oral Airflow (Oaf), expressed in cubic decimeter per second (dm^3/s), was collected using a flexible silicone mask pressed on the subject's mouth.

The EGG signal (Glottal Enterprise) was obtained by placing two electrodes on both sides of the thyroid cartilage. The EVA2 station is equipped with a microphone allowing to retrieve the acoustic signal. We analyzed manually the data with PHONEDIT Signaix [9] to get (1) the duration based on the audio waveform, (2) the peak of Po for each token, (3) the peak of Oaf and the air volume in cm^3 for each token. We analyzed the global shape of the EGG signal to confirm that the beatboxed classic kick drum uses a glottalic egressive airstream.

2.2. Acoustic data

We used Praat [2] to run an acoustic analysis based on the audio signal together with the EGG signal that were both recorded during the aerodynamic experiment. First we segmented the audio on a textGrid and for each token we labelled the acoustic phase of the classic kick drum: the occlusion, the burst (i.e. impulsional noise produced when the lips come apart), the release noise (i.e. noise generated by the air passing through the lips after the lip aperture) and the frication noise (i.e. the noise generated during a fricative consonant, only for [p'ɸ]). Finally, for each phase we were able to extract the duration in milliseconds (ms), the intensity in decibels (dB) and the FFT spectrum (25ms window) in order to get the centre of gravity (CoG in Hz), the skewness (i.e. spectral asymmetry distribution) and kurtosis (i.e. peakedness of the distribution). Since a silicone mask was used during the experiment, we were able to keep the same distance between the participant mouth and the microphone, thus allowing us to analyze the intensity values.

2.3. Laryngoscopic data

The laryngeal examination was performed by the second author using a nasal flexible fiberscope (Kay-Pentax®FNL10RP3) with a DigitalStrobe®, RLS91000 (Kay Elemetrics, Lincoln Park, NJ, USA) station. The video rate was 25 frames per second (fps). Laryngoscopy is an invasive procedure that allows visualizing laryngeal structures without interfering with phonation or articulation. The subject did not ask for any anesthetic. We used VirtualDub software to visualize the video (13,72s duration) and to extract each image: 13,72s x 25 fps (frames per second) = 343 images. We focused on describing glottal adduction and supraglottal contraction.

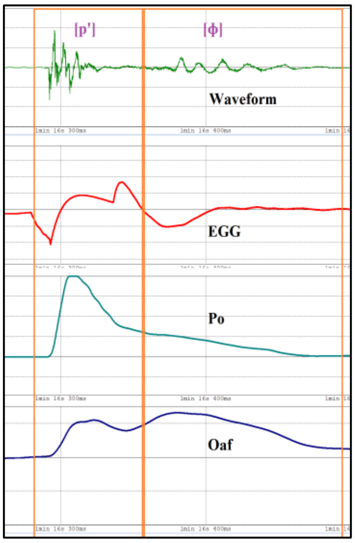
3. RESULTS

We found two types of classic kick drum, one is a bilabial voiceless glottalic egressive [p’]; the second was produced as a bilabial voiceless affricate [p'ɸ]. The second sound appears only in a beatboxed sequence of kick-drums, hi-hats (i.e. [t͡s']) and snare drum (i.e. [k’]); for now, we do not know if it is a stylistic variant or an effect of respiratory control.

3.1 Aerodynamic and EGG data

We analyzed 24 productions of beatboxed classic kick drum, 18 were unaffricated; 6 were affricated.

Figure 1: Acoustic waveform, EGG signal, Po (hPa) and Oaf (dm^3/s) of an affricated classic kick [p’ɸ]. The red dotted line indicates the acoustic release.



On Fig. 1, Po increases up to 20 hPa indicating that an occlusion occurs. The EGG signal shows that the glottis is closed until the end of [p’]. Shortly after the occlusion occurs, we note that Oaf is positive, there is an egressive airstream. Taken together, the data prove that the beatboxer is using a glottalic egressive airstream to produce the classic kick drum effect. As shown in Fig.1, the affricated classic kick drum adds a pulmonic fricative [ɸ]. After the release of this type of kick drum, Po starts dropping but as soon as the glottis opens on the EGG signal the Po drops more slowly and Oaf exhibits another peak. This second peak of Oaf indicates a second constriction is created in the vocal tract. It results in an increased velocity of the volume passing through the constriction point, made by the lips. The coordination of the occlusion and release of the beatboxed classic kick drum is very peculiar. Indeed, while Po still builds up, Oaf shows that air is coming out of the mouth. There seem to be a leakage during the lip occlusion. We hypothesize that while maintaining the occlusion, the beatboxer’s lips become slightly opened by a little slit, thus allowing the pressure to keep raising and the air to come out. This type of coordination (i.e. closure and leakage) is normally not found in normal and casual speech. This appears to be a specificity of beatboxing or at least a specificity of the beatboxer’s production since all tokens showed this type of coordination.

Table 1 shows the mean and standard deviation of the duration (ms), based on the acoustic waveform, Po (hPa), Oaf (dm^3/s) and the air volume for [p'] and [p'ɸ]. Although the Po saturated at 20 hPa, it is consistent with an ejective production [4]. Concerning Oral Airflow, the values are stable and similar for both [p'] and [ɸ]. The volume of air for the fricative is higher and more variable than for the ejective. The use of different airstream mechanisms, one glottalic egressive and the other pulmonic, explains the discrepancy between volume values. Indeed, the pulmonic fricative uses a larger air volume in the lungs whereas the ejective uses only the volume contained between the glottis and the lips. We hypothesize that this alternation between pulmonic and glottalic airstreams may help to keep a fluid rhythm while beatboxing.

Table 1: Mean (standard deviation) duration (ms), Po (hPa), Oaf (dm^3/s) and air volume (cm^3) for [p'] and [p' ɸ].

	unaffricated	affricated	
	[p']	[p']	[ɸ]
Duration (ms)	65(9)	50(7)	133(40)
Po (hPa)	19,7(0,7)	19,7(0,4)	4,1(1,4)
Oaf (dm^3/s)	1,1(0,1)	1,2(0,1)	1(0,1)
Volume (cm^3)	48(11)	43(6)	102(38)

3.2 Acoustic data

We extracted the audio data collected during the aerodynamic experiments and analyzed (1) the duration of the occlusion, burst and frication, (2) the intensity, (3) the resonance frequency.

Table 2: Mean (standard deviation) of duration (ms), intensity (dB), centre of gravity (Hz), Skewness and Kurtosis.

	Closure (n=18)	Burst (n=25)	Noise (n=25)	Frication (n=8)
Duration	6(1)	14(15)	41(12)	137(49)
Intensity		68(4)	60(5)	49(6)
CoG		243(44)	176(91)	531(268)
Skewness		9(1,8)	11(4,6)	4,5(2)
Kurtosis		133(41)	239(121)	39(32)

Concerning the duration of the occlusion (Table 2), due to the particular coordination of occlusion/release, not all of the classic kick drum showed an occlusion preceding the release. Indeed, only 18 productions (16 unaffricated and 2 affricated classic kick drums) showed evidence of closure. The mean duration and standard deviation of the closure is 7(1) ms which indicates a quick occlusion and little variation. The mean duration and standard deviation of the burst is 14(15)ms which indicates there is variation across the duration of the release. Finally, the duration of the frication of [p'ɸ] is longer, the mean and standard deviation are 137(49)ms. This difference of duration may be explained by the fact that to sustain a continuous frication, articulatory adjustments are necessary. Indeed, as pointed out by Signorello et al. [12], the size of the constriction is to be controlled and both intraoral differential pressure (i.e. $\Delta P_{intraoral} = P_{intraoral} - P_{atmospheric}$) and subglottic differential pressure ($\Delta P_{subglottic} = P_{subglottic} - P_{intraoral}$) must reach a specific threshold to produce and sustain audible frication.

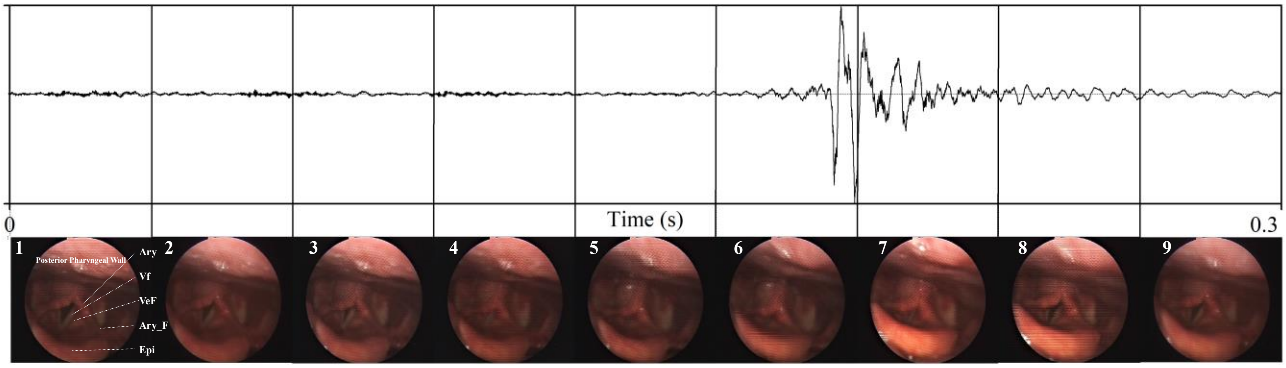
Regarding intensity values, the burst shows the highest values with little variation across tokens. It can be explained by the fact that the classic kick drum uses a glottalic egressive airstream in order to increase Po resulting in a strong intensity burst. The noise release intensity is lower and indicates that intensity is dropping after the lips come apart. Finally, the frication noise is the less intense but the more variable. Signorello et al. [12] reported values between 52 dB and 64 dB for [f]; they used the same experimental methods. We hypothesize that the frication noise is in fact a bilabial fricative rather than a labiodental one. Indeed, the turbulence may be created by soft tissues (i.e. lips) and not by hard tissues (i.e. teeth).

Mean (standard deviation) in the spectral analysis revealed low frequencies and a peaked spectrum. The CoG is 209(35) Hz and 142(72) Hz at the offset. The burst and noise of release show similar values concerning the CoG and the skewness. The classic kick seems to be characterized by low frequency components. Concerning the kurtosis values we values are different, the burst is a bit flatter than the noise. Conversely, the frication noise shows a different pattern. The fricative component has a higher CoG, a lower skewness and a very low kurtosis when compared to the burst and the noise. It indicates a flat low frequency spectrum.

3.3 Laryngoscopic data

The laryngoscopic experiment took place during a second session on the same day. We analyzed 25 repetitions (22 unaffricated and 3 affricated) of the beatboxed classic kick drum in order to describe the laryngeal behavior. Fig. 2 presents an unaffricated classic kick drum together with the acoustic signal extracted from the video. During the initial position (Fig. 2 frame 1) the vocal folds are abducted and the arytenoid cartilages are apart. The pharyngeal

Figure 2: Acoustic waveform and its synchronized laryngoscopic images of an unaffricated classic kick-drum. Frame 1 to 3 show glottal adduction involving vocal folds and ventricular folds, frame 4 to 6 show supraglottal compression involving arytenoid cartilages and the aryepiglottic folds, frame 7 to 9 indicate glottal abduction involving vocal folds and ventricular folds (Frame 1: Vf = Vocal folds, VeF = Ventricular Folds, Ary = Arytenoids, Ary_F = aeryepiglottic folds, Epi = Epiglottis).



cavity is 'open', indeed the hypopharynx shows no sign of contraction. Then, a full vocal fold adduction and an incomplete adduction of the ventricular folds is achieved (Fig. 2 frame 1 to 3). We noted an incomplete anteroposterior closure (i.e. arytenoids are getting closer to the epiglottis) of the arytenoid cartilages and the aryepiglottic folds (Fig. 2 frame 4 to 6). This diminishes both the vocal folds' and the epilaryngeal tube's length. Light contrast (i.e. increased luminosity) on frame 7, 8, 9 may suggest that the larynx has raised. The acoustic burst occurs at the end of frame 6 and the acoustic release corresponds to frame 7 where the glottis opens. During the unaffricated classic kick drum (figure 2) the vocal folds remain adducted while the larynx is raising, whereas during the affricated classic kick drum (not shown) the larynx raising starts with an adducted glottis that opens before the larynx reaches its highest position. This difference may be a sign of coordination between a glottalic egressive airstream and a pulmonic egressive airstream. The raising ends with a wide opened glottis compared to the neutral position. This wide aperture results in a larger volume of air coming from the lung to produce the turbulent airflow of the bilabial fricative component.

4. DISCUSSION

With a multiparametric methodology we were able to describe the physiology the beatboxed classic kick drum produced by a single beatboxer. Our findings concur with other studies about beatboxing. Concerning the laryngeal behavior, [3, 11] also found a glottal adduction and the absence of major supraglottal compression. We found a retraction of the epiglottis, possibly due to a tongue posteriorization. This might be a specific gesture of this present study's subject. This laryngeal behavior is not specific to beatboxing since the same gesture is commonly used to achieve glottal stops [5, 6]. Beatboxing studies [1, 3, 8, 10, 11] concluded that the classic kick drum is produced with a glottalic egressive airstream, our data suggest the same conclusion.

Aerodynamic data brought light on the coordination of gestures used to achieve the classic kick drum. Normally the production of a plosive is defined by a closure in the vocal tract increasing the pressure behind the constriction point and a decrease in Po after the constrictions' release. In our case the bilabial closure is simultaneous with a leakage suggesting an atypical coordination of common gestures (i.e. closure and aperture the Vocal Tract).

The discussion about the coordination of the gestures implicated in the classic kick production brings us to the heart of the problem: what are the primitives of the system? Is Beatboxing relying on gestures or features? On which mental representations does the Beatboxer rely on? All those issues will be resolved by future studies on this technique with multi-instrumental experiments. Indeed, our multiparametric study was useful because we could not only come to the same general conclusion as in [1, 3, 8, 10, 11] about the classic kick drum (i.e. glottalic egressive airstream) but also provide finer phonetic details such as the gestures involved and their coordination. In future studies we will use the same methods in order to gather more data from more subjects and assess the reproducibility of the present results. We hope to be able to answer the previous questions about the primitives of beatboxing.

5. CONCLUSION

This pilot study based on aerodynamics, EGG, acoustics and laryngoscopy showed the valuable information extracted from the data to discuss fine phonetic details in beatboxed sounds and to understand the gestures and their coordination.

6. AKNOWLEDGEMENT

We thank the French beatboxer JOOS for participating to this experiment.

7. REFERENCES

[1] Blaylock, R., Patil, N., Greer, T., Narayanan, S. 2017. Sounds of the Human Vocal Tract. *INTERSPEECH 2017*, (pp. 2287-2291). Stockholm, Sweden.

[2] Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer. Version 6.0.21, recupéré le 25 Septembre 2016 sur http://www.praat.org/.

[3] de Torcy, T., Clouet, A., Pillot Loiseau, C., Vaissière, J., Brasnu, D. & Crevier-Bushman, L. 2013. A videofibroscopy study of laryngo-pharyngeal behaviour in the human beatbox. *Logopedics Phoniatrics Vocology, 39(1)*, 38-48.

[4] Demolin, D. 2004. Acoustic and aerodynamic characteristics of ejectives in Amharic. *The Journal of the Acoustical Society of America, 115(5)*, 2610-2610.

[5] Edmondson, J. A., & Esling, J. H. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, 23(2), 157-191.

[6] Esling, J. H., Fraser, K. E., & Harris, J. G. 2005. Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics, 33(4)*, 383-410

[7] Fant, G. (1970). Acoustic Theory of Speech With Calculations based on X-Ray Studies. The Hague, Netherlands: Mouton.

[8] Patil, N., Greer, T., Blaylock, R., & Narayanan, S. 2017. Comparison of Basic Beatboxing Articulations between Expert and Novice Artists using Real-Time Magnetic Resonance Imaging. *INTERSPEECH 2017*, (pp. 2277-2281). Stockholm, Sweden.

[9] PHONEDIT Signaix (Version 4.2.0.8). Computer program, Aix en-Provence, France: Parole et Langage Laboratory, Retrieved on February 28, 2017 from http://www.lpl-aix.fr/lpldev/phonedit/.

[10] Proctor, M., Bresch, E., Byrd, D., Nayak, K. & Narayanan, S. 2013. Paralinguistic mechanisms of production in human “beatboxing”: A real-time magnetic resonance imaging study. *The Journal of the Acoustical Society of America; 133 (2)*, 1043-1054.

[11] Sapthavee, A., Yi, P., & Sims, H. S. 2014. Functional endoscopic analysis of beatbox performers. Journal of Voice, 28(3), 328-331.

[12] Signorello, R., Hassid, S. & Demolin, D. 2018. Toward an aerodynamic model of fricative consonants. Journal of Acoustical Society of America, 143(3), 386-392.