Vocal drum sounds in human beatboxing: An acoustic and articulatory exploration using electromagnetic articulography

Annalisa Paroni, Nathalie Henrich Bernardoni, Christophe Savariaux, Hélène Loevenbruck, Pascale Calabrese, Thomas Pellegrini, Sandrine Mouysset, Silvain Gerber

JASA ARTICLE

Vocal drum sounds in human beatboxing: An acoustic and articulatory exploration using electromagnetic articulography

Annalisa Paroni,[1] Nathalie Henrich Bernardoni,[1,a)] Christophe Savariaux,[1] Hélène Lœvenbruck,[2] Pascale Calabrese,[3] Thomas Pellegrini,[4] Sandrine Mouysset,[4] and Silvain Gerber[1]

[1]*Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France*

[2]*Univ. Grenoble Alpes, Univ. Savoie Mont-Blanc, CNRS, LPNC, F-38000 Grenoble, France*

[3]*Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France*

[4]*IRIT, Toulouse, France*

ABSTRACT:

Acoustic characteristics, lingual and labial articulatory dynamics, and ventilatory behaviors were studied on a beatboxer producing twelve drum sounds belonging to five main categories of his repertoire (kick, snare, hi-hat, rimshot, cymbal). Various types of experimental data were collected synchronously (respiratory inductance plethysmography, electroglottography, electromagnetic articulography, and acoustic recording). Automatic unsupervised classification was successfully applied on acoustic data with t-SNE spectral clustering technique. A cluster purity value of 94% was achieved, showing that each sound has a specific acoustic signature. Acoustical intensity of sounds produced with the humming technique was found to be significantly lower than their non-humming counterparts. For these sounds, a dissociation between articulation and breathing was observed. Overall, a wide range of articulatory gestures was observed, some of which were non-linguistic. The tongue was systematically involved in the articulation of the explored beatboxing sounds, either as the main articulator or as accompanying the lip dynamics. Two pulmonic and three non-pulmonic airstream mechanisms were identified. Ejectives were found in the production of all the sounds with bilabial occlusion or alveolar occlusion with egressive airstream. A phonetic annotation using the IPA alphabet was performed, highlighting the complexity of such sound production and the limits of speech-based annotation.

 https://doi.org/10.1121/10.0002921

(Received 10 April 2020; revised 19 November 2020; accepted 24 November 2020; published online 7 January 2021)

[Editor: Susanne Fuchs] Pages: 191–206

I. INTRODUCTION

Human beatboxing (HBB) is a vocal art belonging to the Hip-Hop culture. Born in the USA in the early 1980s, it aims to reproduce the sounds of electronic drum machines. The original purpose of HBB was instrumental mimicry: in the absence of the actual instruments, the human substituted for the beatbox machine. Beatboxers reproduced those sounds with their voice to create the rhythmic accompaniment to singing and rapping. HBB has now spread all over the world. While maintaining its instrument mimicry core, it has rapidly evolved in complexity and diversity, both for sound qualities and vocal techniques.

The basic categories of HBB drum sounds, called "effects," comprise mostly plosive and fricative sounds. They are named after the drum kit sounds they imitate: *kick* (kick drum), *snare* (snare drum or side drum), *rimshot* (a percussion technique performed on the snare drum), *cymbal*, *open* and *closed hi-hat* (matching pair of two cymbals mounted on a stand, held apart: *open hi–hat* or together: *closed hi–hat*, by means of a foot pedal). Beatboxers can perform the rhythmic line alone or together with other sounds or a melody that propagate through the nose (humming technique). This terminology is widely shared within the international HBB community. In addition, a prerogative of each beatboxer is to experiment with their own voice and discover new and innovative sounds, so that their repertoire of HBB sounds is in constant evolution. In this respect, HBB is a very prolific environment for the experimentation and creation of human sound production. HBB learning often begins with training on speech plosives, syllables, or sentences. For instance, basic kick sounds are often learned from [p], hi-hat sounds from [t] for closed hi-hat or [ts] for open hi-hat, and rimshot sounds from [k]. The articulatory adaptation that enables transformation of a linguistic sound into a HBB sound remains mostly unexplored.

From a scientific perspective, this human-voice sound production is captivating, because beatboxers explore all the possibilities of their vocal instrument unrestrained by style or language. However, the existing literature on HBB comprises only a few published studies. The earliest works dealt with automatic recognition and classification of basic HBB

a)Electronic mail: nathalie.henrich@gipsa-lab.fr, ORCID: 0000-0002-3944-611X.

sounds based on acoustic data. Kapur *et al.* (2004) exploited acoustic features of some HBB sounds and rhythmic information for a new approach on music information retrieval (MIR). Sinyor *et al.* (2005) tested the use of the autonomous classification engine (ACE) for classifying some basic HBB percussion sounds. More recently, Picart *et al.* (2015) and Evain *et al.* (2020) investigated the automatic recognition of pre-recorded HBB drum and instrument sounds. Other studies have investigated articulatory aspects of HBB production. De Torcy *et al.* (2014) and Sapthavee *et al.* (2014) conducted endoscopic investigations on the laryngeal structures involved and the overall behavior of the larynx during beatboxing. They showed very active laryngopharyngeal dynamics and a dissociated mobilization of the laryngopharyngeal structures. These authors pointed out the use of extreme articulatory configurations in the laryngopharynx region, a piston-like action of the closed glottis that accompanies the production of some plosive sounds (De Torcy *et al.*, 2014) as well as articulatory behaviors that can protect against glottal injury (Sapthavee *et al.*, 2014). Some studies have explored the articulatory mechanisms of HBB in the vocal tract mid-sagittal plane. Proctor *et al.* (2013) analyzed the articulatory mechanisms of 17 HBB drum sounds belonging to the repertoire of a professional beatboxer. They found that they were similar to those exploited in speech, such that the authors were able to annotate each sound using the International Phonetic Alphabet (IPA) which was devised to represent the sounds of spoken language. However, further data (Blaylock *et al.*, 2017) showed that beatboxers employ an extremely wide variety of articulatory mechanisms, in terms of both place and manner of articulation, as well as airstream mechanisms, often non-attested in speech but some of which were recently mentioned in vocal imitations of non-speech sounds (Friberg *et al.*, 2018; Helgason, 2014). In addition, the higher the level of expertise, the better the control of articulatory and airstream mechanisms (Patil *et al.*, 2017). All three studies (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013) described the use of ejective productions of several plosive sounds, wherein the closed glottis acts like an upward moving piston to compress the air trapped between the glottal closure and a supraglottal closure (Ladefoged and Maddieson, 1996) to produce a more intense sound upon supraglottal closure release than a pulmonic plosive would produce.

The few available studies so far that investigate HBB production mechanisms have employed techniques such as endoscopy (De Torcy *et al.*, 2014; Dehais Underdown *et al.*, 2019; Sapthavee *et al.*, 2014) and rtMRI (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013). While providing valuable information on the general behavior of the articulators, neither technique allows the study of the dynamics of a given flesh point on an articulator. Both techniques are also limited by a relatively low sampling frequency. More precise and quantitative evaluation of the articulatory dynamics could be performed using electromagnetic articulography (EMA), a widely used technique in speech research to measure the position and movement over time of selected points on articulators (Barbier *et al.*, 2020; Brunner *et al.*, 2010; Savariaux *et al.*, 2017; Tiede *et al.*, 2019a).

This study is part of an ongoing effort to understand the production of HBB drum sounds by exploring lingual and labial articulatory dynamics in relation to acoustic characteristics and ventilatory behavior on a beatboxer producing five categories of drum sounds belonging to his repertoire (kick, snare, hi-hat, rimshot, cymbal). We rely on the technique of electromagnetic articulography to explore the kinematics of tongue and lip-flesh points. A database comprising synchronized recordings of respiratory, phonatory and articulatory signals is presented in Sec. II. Acoustic and articulatory characterizations provided in Secs. III A and III B lead to a phonetic description of the ways these vocal drum sounds are produced. Section IV discusses the specifics of HBB sound production.

II. MATERIAL AND METHODS

A. Subject

The subject is a 28 year-old left-handed male native speaker of French. He has been practicing HBB for 9 years at an amateur level. He occasionally performs in concerts, however, he has never participated in HBB competitions. Because of having often experienced vocal fatigue and discomfort after practicing, the subject learned the diaphragmatic breathing technique (Leanderson and Sundberg, 1988; Leanderson *et al.*, 1984). He reports benefiting from this breathing technique in his HBB practice and that he no longer suffers from vocal fatigue.

B. Corpus and protocol

The protocol started with an interview of the beatboxer prior to the recording session. His experience in HBB and his vocal habits were collected. The experimental details were presented to him. The HBB effects of interest for the study—five categories of drum sounds (kick, hi-hat, snare, rimshot, and cymbals) and their variants—were discussed with him. He stated that he could produce more than one sound for each effect: a humming variant and a non-humming one (that he called *power*), and an inhaled and exhaled variant. All the humming sounds were produced without superimposed melody or voicing [audio files are available online (Henrich Bernardoni and Paroni, 2020)].

The following HBB vocal drum sounds were then recorded:

(1) **Kick**: humming and power variants
(2) **Hi-hat**: humming and power variants, open and closed for the power variant
(3) **Snare**: humming and power variant, exhaled and inhaled for the power variant
(4) **Rimshot**: humming and power variants
(5) **Cymbal**: exhaled and inhaled

https://doi.org/10.1121/10.0002921

Each sound was repeated at least 15 times, while following the tempo provided by a metronome set at 80 beats per minute (bpm), and varying loudness when possible. A phonetic description of these sounds, based on the acoustic and articulatory findings of this study, is provided in Table II.

C. Experimental setting and apparatus

The data collection took place in a semi-anechoic room during a one-hour session.

After being interviewed and signing an informed consent form, the subject was placed in the recording room [Fig. 1(b)], wearing a waistcoat for respiratory inductance plethysmography (VISURESP system, RBI, France), and sitting on an adapted chair that assured the stabilization of the head inside of the magnetic field of an electromagnetic articulograph (EMA) (WAVE, NDI, Canada). To collect the articulatory data, 12 coils were positioned as follows [Fig. 1(a)]:

- three coils were placed midsagitaly on the tongue: 1 coil about 1 cm from apex (TIP), 1 coil on the blade about 3 cm from apex (MID), and 1 coil on the dorsum about 5 cm from apex (DORS);
- one coil on the medial lower incisors (JAW);
- two coils on the upper lip (mid, UML and left, ULL), two coils on the lower lip (mid, LML and left, LLL);
- one reference coil on the upper incisors;
- two reference coils on the mastoid processes behind both right and left ears;
- one reference coil on the nasion.

The EMA signal was sampled at 400 Hz. After recording, EMA data were post-processed in two steps (for more details on the method, see Tiede *et al.*, 2019b). As a first step, movement of the head was corrected with MATLAB software using the four reference coils glued on the nose, upper incisor, and behind both ears. As a second step, a rotation and a translation were applied to reference the data in the coordinate system of the beatboxer [Fig. 1(d)].

Two pairs of electrodes [Glottal Enterprise EG2 dual-channel electroglottograph (Rothenberg, 1992)] were positioned on the neck of the subject in the larynx region for measuring vocal-fold contact and detecting laryngeal movements [Figs. 1(a) and 1(b)]. An AKG microphone and a 1/2 in. prepolarized free-field microphone (B&K 4189) connected to a microphone preamplifier (B&K 2669 C) and NEXUS conditioning amplifier (B&K 2690) were placed at a distance of approximately 20 cm from the subject's mouth in order to capture the audio signal and derive intensity level after calibration. Both electroglottographic (EGG) and audio signals were sent to a BIOPAC unit (MP150) and sampled at 40 kHz. The respiratory inductance plethysmographic (RIP) signals were recorded on two devices: a computer

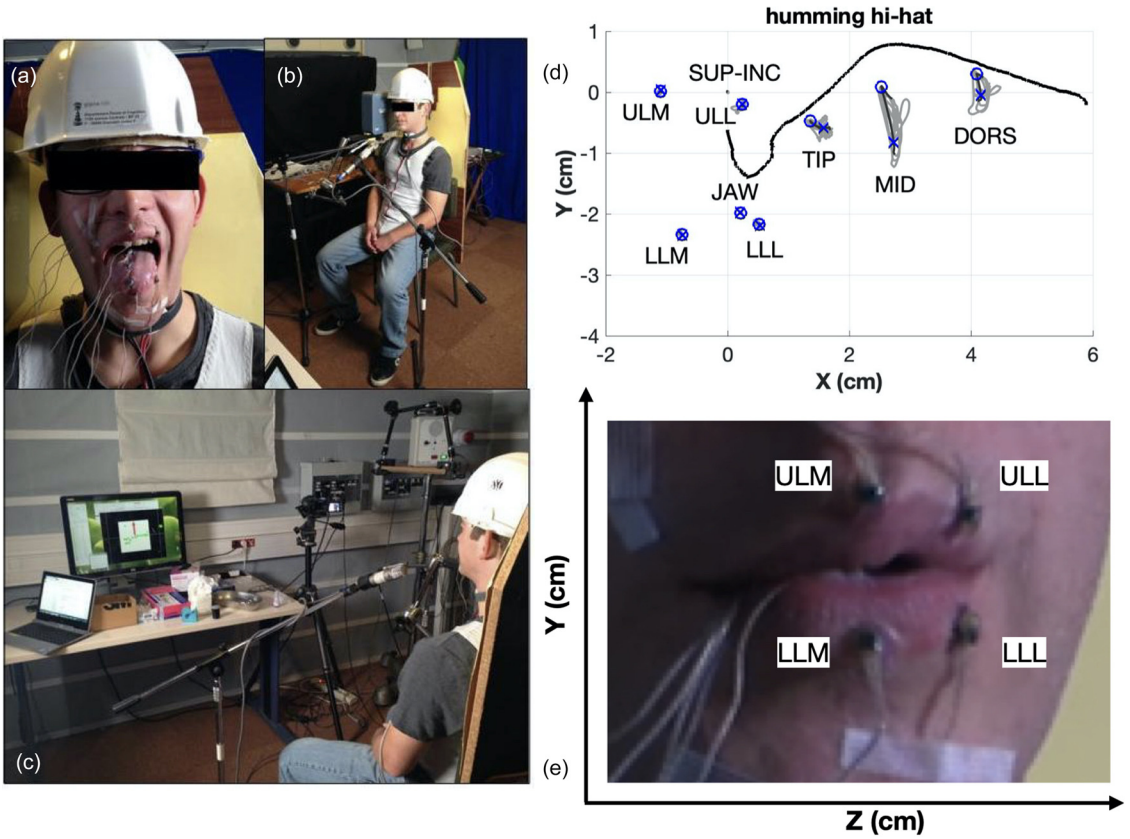

FIG. 1. (Color online) (a) Coils layout; (b) and (c) experimental setting; (d) sagittal (XY) view of hard palate contour, coil trajectories, and corresponding labeling; (e) frontal (ZY) view of lip coils and corresponding labeling.

dedicated to VISURESP system (at 40 Hz sampling frequency) and the BIOPAC unit so as to be synchronized with audio and EGG signals (at 40 kHz sampling frequency).

A camera was facing the subject for the video recordings at 25 fps. During recording, an acoustic trigger signal (20 ms square wave) was manually launched by an external electronic device and captured by each system prior to and after each task, so as to allow data synchronization in post-processing.

At the end of the recording session, a coil manually traced the mid-sagittal plane from the back of the palate to the front of the upper incisors to obtain the palatal contour.

D. Methods

All data were synchronized. Audio files were manually segmented and phonetically annotated using the software PRAAT (version 6.0.49) (Boersma, 2006). The phonetic annotations were carried out inspecting audio, video, and EGG data. Audio files were segmented using the following criteria: the left boundary was placed in correspondence with the burst and the right boundary in correspondence with the last visible oscillation on the waveform. Boundaries subsequently provided timestamps for the meaningful quantities investigated. The phonetic annotations were performed by the first author who is a speech therapist and has also received a training as a linguist. The alphabet used was the Worldbet Alphabet (Hieronymus, 1993), which is the translation of IPA into symbols compatible with automatic data processing.

A clustering technique was used to test whether HBB sounds are distinguishable on the basis of the acoustic signal. Spectral clustering on 12 t-SNE-whitened Mel frequency cepstral coefficients (MFCCs) was applied. t-SNE (Maaten and Hinton, 2008), which stands for t-distributed stochastic neighbor embedding, is a recent and efficient non-linear projection technique (SC) (Von Luxburg, 2007). The first coefficient (C0) was removed, as it measures signal loudness that is not relevant to characterize the frequency content of interest. The MFCCs were extracted every 6.25 on 25 ms frames, with 50 and 8000 Hz as minimum and maximum extreme frequency values to compute the Mel bands.

EMA data were processed using the commercial software package MATLAB (MATLAB, 2018). The spatial trajectories of the eight coils positioned on the tongue, jaw, and lips were computed. A visual inspection of the trajectories was carried out to characterize the articulation of each HBB sound. Corrections to the phonetic annotation were introduced when needed.

The EGG signal is composed of both a high–frequency component, which reflects vibration of the vocal folds (voicing) and a low–frequency component corresponding to slow vertical motions of the larynx (e.g., during swallowing). For a recent review on EGG use in research, see Herbst (2020). The EGG signal was visually inspected to detect vocal fold vibration phases.

Respiratory inductance plethysmography (RIP) measures thoracic and abdominal cross sectional area changes. RIP data were calibrated following the method used and described by Eberhard *et al.* (2001) and Calabrese *et al.* (2007). The thoracic and abdominal signals measured with RIP were simultaneously recorded, together with the airflow signal measured by a flowmeter (Fleishhead No. 1, Emka Technologies, Paris, France), a differential transducer (163PC01D36, Micro Switch, Honeywell, United States), and a face mask worn by the subject while breathing spontaneously for approximately one minute. Thoracic and abdominal signals recorded with RIP were subsequently linearly combined to obtain the ventilatory volume signal. The linear coefficients were estimated from the least square method to fit the airflow signal recorded with the flowmeter.

Several parameters were extracted from the annotated data: sound duration and vocal intensity (from acoustic signal), maximum of tangential speed and acceleration (from EMA signals). Three statistical analyses were performed using the R software (R Core Team, 2013). First, a test was run to inspect if a difference in intensity (response variable, in logarithmic scale) exists between variants (humming vs power) of the same effect (kick, snare, hi-hat, rimshot). Second, an analysis was carried out to test what kind of relationship exists between duration (response variable, in logarithmic scale) and intensity in each HBB sound (12 modalities: humming kick, humming snare, humming hi-hat, humming rimshot, power kick, power snare, power inward snare, power closed hi-hat, power open hi-hat, inhaled cymbal, exhaled cymbal). Last, an analysis was carried out on the HBB sounds to inspect whether a significant difference exists among the means of the maximum speed of pairings of lingual articulators (TIP, MID, DORS) and of lip articulators (JAW-LLL, LLL-ULL). The considered factors are the coils (eight modalities: TIP, MID, DORSUM, JAW, ULL, UML, LLL, LML) and the HBB sounds (12 modalities: as above) and their interaction. Each analysis was run using the lme function of the nlme package (in R). This function takes into account potential differences in residual variances across HBB vocal drum sounds, or possible correlations among coils in the third analysis. Repetition is considered as a random effect. All the p-values reported in Sec. III are provided by the glht function of the multcomp package (Hothorn *et al.*, 2008) calculated from the corresponding model. For the first and the third model, the estimated differences of the comparisons and their estimated standard errors are provided. For the second analysis, the estimated values of the slopes and their estimated standard errors are provided.

III. RESULTS

In this section, prototypical examples are presented. The corresponding audio examples and video files can be found online (Henrich Bernardoni and Paroni, 2020).

https://doi.org/10.1121/10.0002921

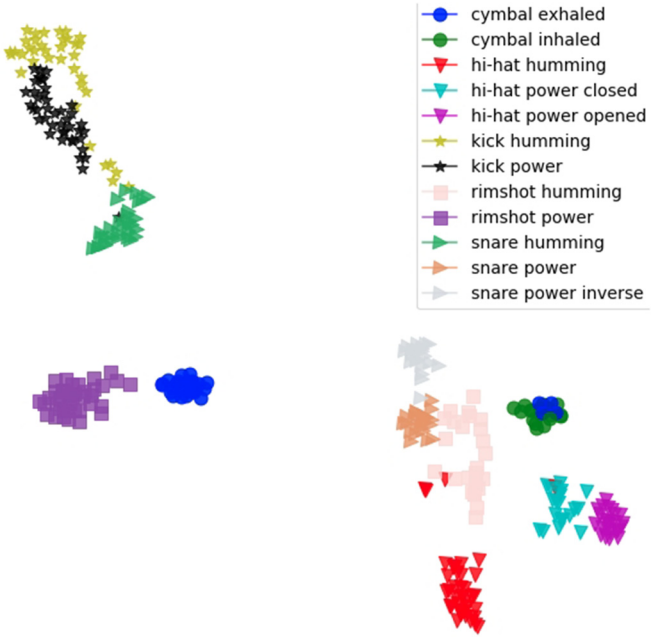

FIG. 2. (Color online) Visualization obtained with the t-SNE projection technique. Although the x axis and y axis are arbitrary scales, one can see that the different sounds are clearly grouped into distinct clusters [color version available online (Henrich Bernardoni and Paroni, 2020)].

A. Acoustic characterization

341 sound realizations of twelve HBB sounds were analysed. Acoustic characterization performed through spectral clustering achieved a 94% clustering purity value. Nineteen samples out of the 341 realizations were misclassified. Out of these 19 misclassifications, 12 were annotation errors. For instance, four exhaled cymbal realizations were wrongly annotated as inhaled cymbal. The remaining misclassifications were confusions, among which the most frequent was between humming kick and humming snare. Figure 2 shows the data points after a two-dimension reduction with t-SNE. In this plot, the x axis and y axis are the output of the t-SNE projection technique and thus, they are arbitrary scales. Each data point is plotted using shape and color according to its sound label [colored version of the figure available online (Henrich Bernardoni and Paroni, 2020)]. Pure and meaningful compact clusters can clearly be identified. In general, variants of a same HBB effect are also close together, e.g., the points for power kick and humming kick lie in the same region. Cymbal (in particular the inhaled variant) and hi-hat points are close together, which makes sense, as the two sounds have a similar acoustic signature [see Fig. 3 and audio files online (Henrich Bernardoni and Paroni, 2020)].

This classification accuracy, i.e., the fact that each sound can be correctly assigned to its corresponding cluster via unsupervised methods, demonstrates that each HBB vocal drum sound has its own characteristic acoustic signature. Figure 3 illustrates these signatures with the waveform and spectrogram of a representative token for each HBB sound explored in the present study.

Most sounds have a duration shorter than 200 ms (Figs. 3 and 4). Only three sounds were associated with longer duration, ranging from 300 to 700 ms. Six sounds are impulsive sounds, most often produced with a strong burst: humming and power kick, humming and power rimshot, humming and power closed hi-hat. The others are characterized by an impulse attack followed by a more or less protracted friction noise: power snare and inward snare, exhaled cymbal and inhaled cymbal, power open hi–hat. Some sounds show a vibration component, either for the whole sound (humming kick and snare) or for the attack (power snare, inward snare, exhaled cymbal). The EGG signal does not show any signs of vocal-fold vibration (Figs. 7, 9, and 11), hence indicating that the vibratory source is located elsewhere than the glottis. The vibratory-source nature will be discussed in Sec. IV F.

As shown in Fig. 4 and Table I, HBB sound duration ranges from 37 $\pm$11 ms for humming kick to 595 $\pm$139 ms for power open hi-hat. Sound intensity ranges from 41 $\pm$1 dB for the softest (power open hi-hat) to 60 $\pm$1 dB for the loudest one (power snare), as shown in Fig. 5 and Table I. Large variability among the sound realizations is clearly visible, especially for the power inward snare. The power version of all the effects is always produced at a higher intensity than the humming ones. The difference in intensity between power and humming variants of the same sound category is significant for the kick (0.1973 ± 0.0110, $p < 0.001$), snare (0.1412 ± 0.0144, $p < 0.001$), hi-hat (0.1179 ± 0.0145, $p < 0.001$) and rimshot (0.2034 ± 0.0133, $p < 0.001$) effects.

The ANCOVA analysis shows that sound duration and vocal intensity do not correlate with each other in most cases, except for three sounds (humming rimshot, power inward snare, and power closed hi-hat). Sound duration negatively correlates with intensity for humming rimshot (-0.0845 ± 0.01467, $p < 0.001$) and power closed hi-hat (-0.0559 ± 0.0114, $p < 0.001$), whereas a positive yet weaker correlation is found for power inward snare ($0.0151 \pm 0.0.0049$, $p < 0.05$).

B. Articulatory characterization

Based on acoustic, EGG, video, respiratory, and EMA data [see also multimedia material available online (Henrich Bernardoni and Paroni, 2020)], the HBB drum sounds could be qualitatively interpreted as corresponding to a variety of articulatory and phonatory gestures, ranging from bilabial ejectives to lateral clicks, in addition to more common ones for French such as oral occlusives and fricatives. Some non–linguistic mechanisms were also observed. Figures 6, 8, and 10 show the displacements of the lips and tongue sensors during five repetitions of the same sound. The trajectory of a representative gesture is highlighted in black. In general, coil trajectories are rather consistent over the repetitions, meaning that the articulatory pattern of each sound is stable.

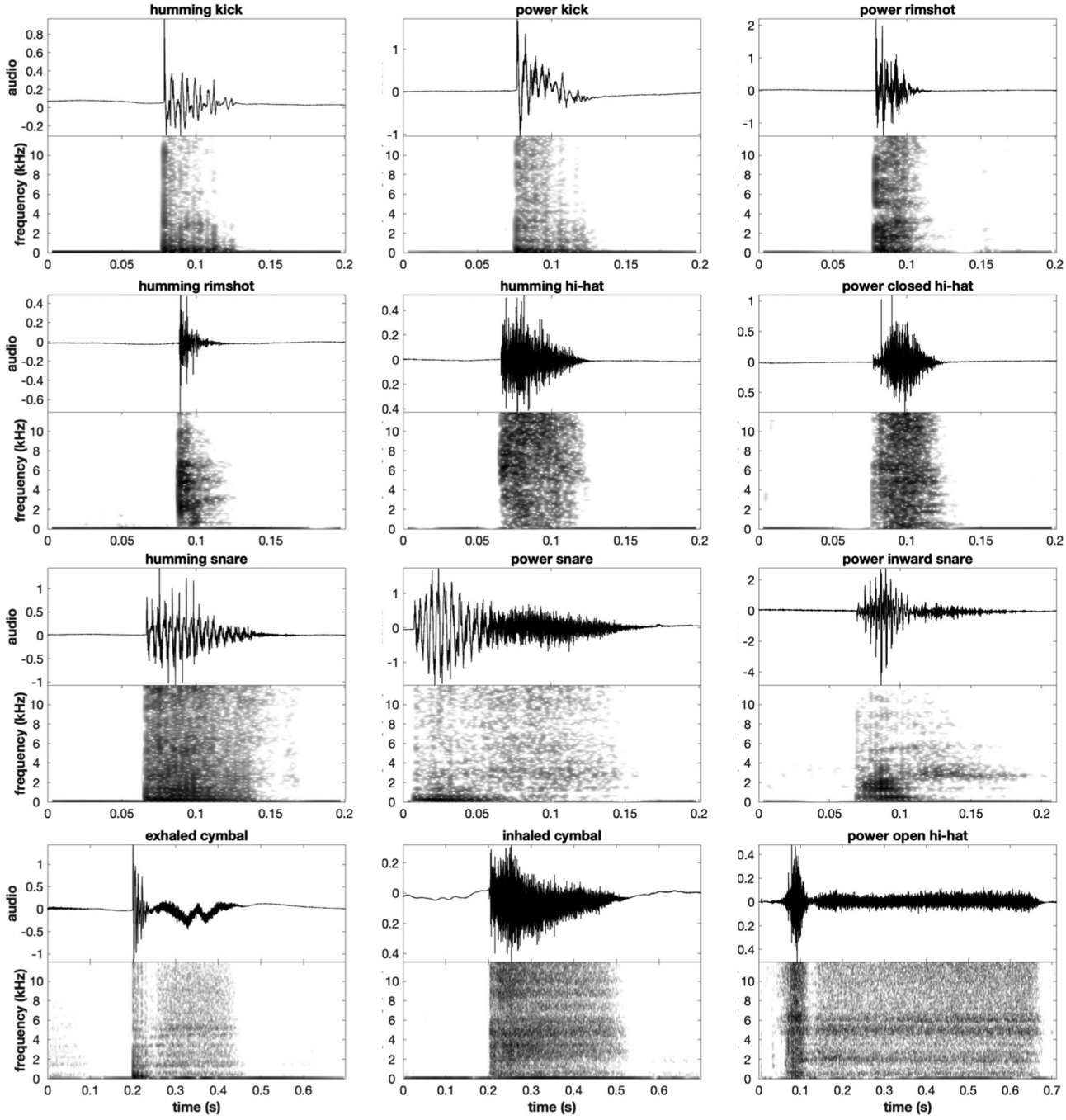

FIG. 3. Audio waveforms and spectrograms of a representative token for each of the twelve HBB sounds. Spectrogram parameters: view range: 0–12 kHz; window length: 5 ms; dynamic range: 50 dB.

1. Lip articulations

Five HBB sounds were produced with complete lip occlusion: humming and power kick, humming and power snare, and exhaled cymbal. The release is lateralized to the left portion of the lips, as evidenced by EMA and video data.

The lips undergo relatively large and fast protrusion displacements during the realization of the humming and power kicks, whereas their movements are smaller for the humming and power snares (Figs. 6 and 7) and the exhaled cymbal (Fig. 10). The tongue is very active in the articulation of both humming and power kicks and snares (Fig. 6): the tongue sensors display considerable movements along regular trajectories that are similar for humming kick and humming snare and for power kick and power snare, but differ between humming and power. For the humming sounds, the tongue is raised in the dorsal region against the palate, suggesting a back closure isolating the oral cavity from the rest of the vocal tract. The coil trajectories suggest a pushing action of the tongue from back to front and from right to left toward the point at the lips where the occlusion is released.

https://doi.org/10.1121/10.0002921

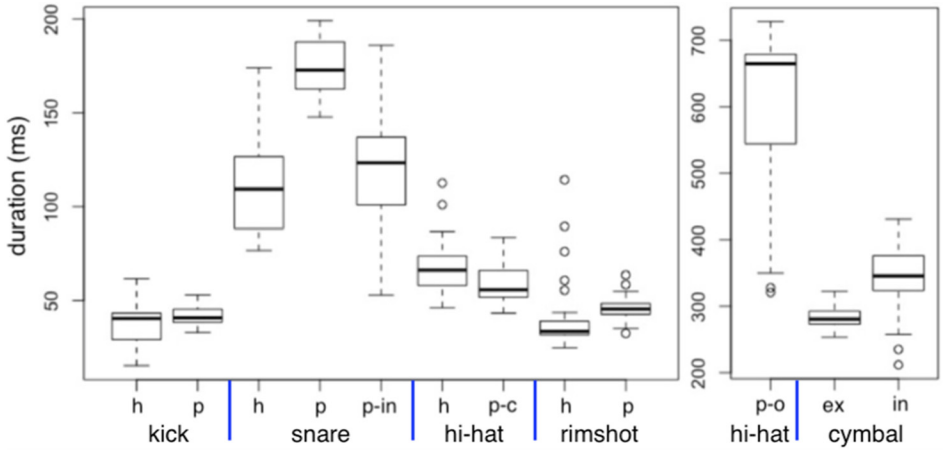

FIG. 4. (Color online) Distribution of duration for the twelve HBB sounds. Legend: h = humming; p = power; c = closed; o = open; in = inward/inhaled; ex = exhaled.

RIP data (Fig. 7) show that humming sound production takes place during both inhalation (increasing VR values) and exhalation (decreasing VR values), suggesting that sound production and breathing are dissociated. This supports the hypothesis that the airflow used in producing the sound is non-pulmonic, originating in the oral cavity. The articulatory pattern of the tongue suggests that it is lingual egressive. The realization of the power sounds is achieved with a flatter tongue that moves from an overall lower to a higher (almost by 2 cm) position in the oral cavity. A laryngeal elevation is evidenced on the video. This movement is probably due to the use of an ejective mechanism. The shorter sound duration of the power kick compared with the power snare does not reduce the overall tongue vertical displacement by much. Decreasing ventilatory volume (VR) values during sound production indicate that the airstream mechanism is egressive for both sounds (Fig. 7). For the power snare, the fricative portion of the sound (Figs. 3 and 7) is likely produced with a pulmonic egressive airstream. Video and acoustic data show that the stricture of close approximation related to this friction is created between the left portion of the lower lip and the upper teeth. In the exhaled cymbal (Fig. 10), the tongue, although moving slightly from a lower to a higher position during sound production, especially its posterior portion, assumes an almost horizontal position, revealing a laminar articulation of the fricative portion of the sound (Fig. 3). As for the power snare, the airstream is egressive (decreasing VR values) (Fig. 11). Video data show slight larynx elevation, suggesting the use of an ejective articulation for the bilabial.

TABLE I. Mean and standard deviation (in brackets) of the sound duration and vocal intensity.

Sound	Duration (ms)	Intensity (dB)
Humming kick	37 (11)	47 (3)
Power kick	42 (5)	58 (2)
Humming snare	112 (28)	52 (4)
Power snare	174 (15)	60 (1)
Power inward snare	120 (28)	53 (9)
Humming hi-hat	68 (14)	43 (3)
Power closed hi-hat	59 (11)	48 (2)
Power open hi-hat	595 (139)	41 (1)
Humming rimshot	42 (21)	49 (3)
Power rimshot	46 (6)	59 (2)
Exhaled cymbal	283 (17)	52 (2)
Inhaled cymbal	339 (52)	43 (2)

2. Anterior tongue articulations

Four sounds were produced with complete occlusion of the vocal tract in the alveolar or post-alveolar region: humming hi-hat, power closed and open hi-hat, inhaled cymbal. Different tongue positions and the use of different airstream mechanisms differentiate the realization of these sounds.

The articulatory data for humming hi-hat (Fig. 8) show that the tongue forms a cavity in the mid-region, suggesting that a pocket of air is trapped between the alveolar/post-alveolar and dorsal regions. The mid-region of the tongue is then rapidly pushed upward (Fig. 9) during sound production, suggesting that the oral airflow is indeed generated by a pushing action of the tongue. RIP data (Fig. 9) show that sound production takes place during both exhalation (decreasing VR values) and inhalation (increasing VR values). This is evidence for the use of a non-pulmonic airflow

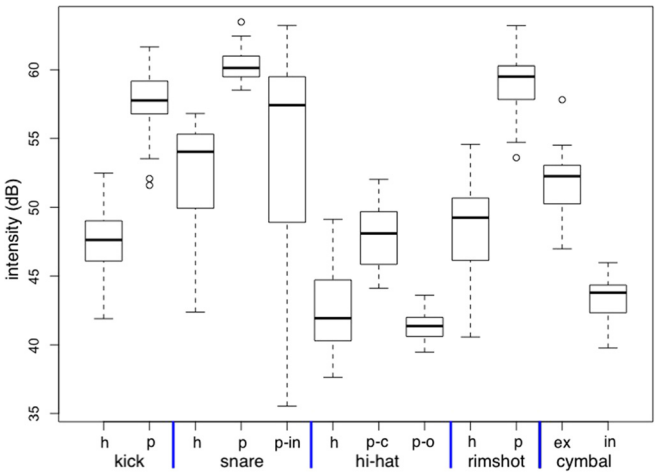

FIG. 5. (Color online) Distribution of vocal intensity for the twelve HBB sounds. Legend: h = humming; p = power; c = closed; o = open; in = inward/inhaled; ex = exhaled.

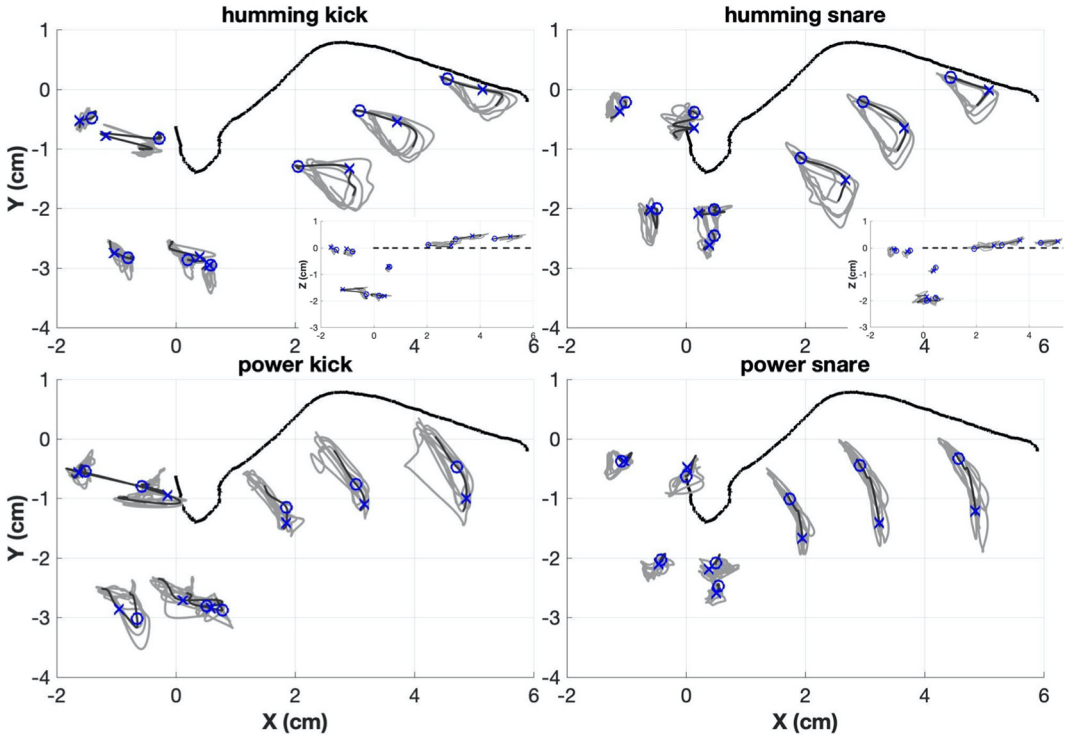

FIG. 6. (Color online) Sagittal (XY) and transversal (XZ) views of trajectories for five repetitions of kick sounds (humming/power) and snare sounds (humming/power). Displayed coils: four lip coils, three tongue coils, jaw coil (see Fig. 1). Solid and dotted black lines: trace of the palate on the mid-sagittal plane. Black segment: trajectory of a representative token (same as Fig. 3). Grey lines: trajectories of the two tokens preceding and the two tokens following the representative token. Cross: start of sound. Circle: end of sound. Animation is available online as multimedia material (Henrich Bernardoni and Paroni, 2020).

that allows some dissociation between sound production and ventilation. The combination of the articulatory pattern and the breathing behavior suggests that this gesture is produced via a lingual egressive airstream mechanism. The lip coils hardly move, meaning that the lips are not active in the articulation of this sound. The posterior seal may take place in the velar region, further back than the DORS coil. The anterior seal may take place in the alveolar or post-alveolar region and may be apical rather than laminal. This would explain the almost horizontal trajectory of TIP coil during sound production.

The articulatory movements of the power closed hi-hat are quite subtle and mainly restrained to the tip region (Fig. 8), especially during sound production, while the tongue assumes a generally flat position in the middle of the oral cavity. The vertical movements of the tongue, especially in its mid and dorsal regions, may be due to an upward movement of the larynx evidenced on the video and likely related to an ejective mechanism.

The power open hi-hat (Fig. 8) is produced similarly to the closed version, but the alveolar occlusive is followed by a laminal constriction. Again, the vertical displacement of the tongue during the first part of the sound production may be related to the upward movements of the larynx [Fig. 8 and Henrich Bernardoni and Paroni (2020)]. The airstream is clearly egressive (decreasing VR values), likely glottal at first, then pulmonic.

The inhaled cymbal is realized with the tongue in an arched and higher position than the other sounds (Fig. 8). The airstream used is pulmonic ingressive (increasing VR values during sound production, Fig. 9).

3. Posterior tongue articulations

Three sounds were articulated with complete occlusion of the vocal tract in the posterior region of the oral cavity: power inward snare, humming rimshot, and power rimshot.

The front portion of the tongue is held against the hard palate in the production of the power inward snare and humming rimshot while the occlusion is released in the dorsal region (Fig. 10).

Sound production during both exhalation (decreasing VR values) and inhalation (increasing VR values) (Fig. 11) indicates that the airstream of the humming rimshot is non-pulmonic. The aggregation of articulatory (Fig. 10), ventilatory (Fig. 11), and acoustic [Fig. 3 and Henrich Bernardoni and Paroni (2020)] data implies that the airstream is lingual ingressive (or velaric).

The power inward snare shows a downward motion of the jaw and lower lip. Increasing VR values during sound production (Fig. 11) suggest that the airstream is pulmonic ingressive. In the power rimshot, only the posterior part of the tongue is in contact with the palate. Before the burst, the motion of the sensors suggests that the tongue is pushed upward and forward, while the occlusion is being held. When the burst occurs, the tongue dorsal region (DORS coil) reaches its highest and most advanced position while the jaw is

https://doi.org/10.1121/10.0002921

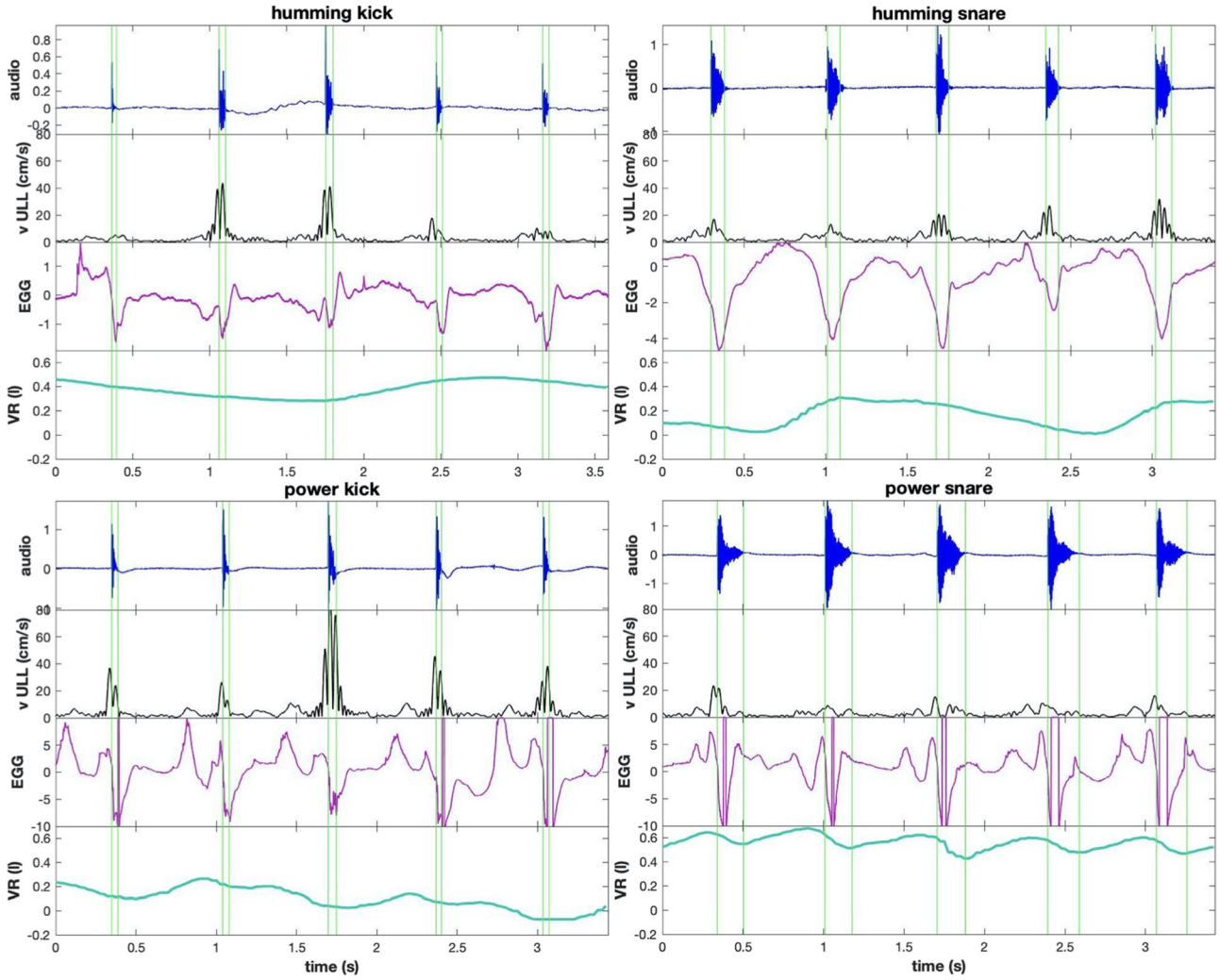

FIG. 7. (Color online) Synchronized audio, lip-coil speed (vULL), EGG, and RIP data (ventilatory volume VR) of five repetitions of kicks (humming/power) and snares (humming/power) (same as Figs. 3 and 6).

lowered together with the lower lip (JAW, LLL, and LLM coils). Systematic decrease of ventilatory volume during sound production (Fig. 11) indicates that the airstream of the power rimshot is egressive. Upward movements of the larynx suggest the use of an ejective mechanism.

C. Articulatory dynamics

The analysis of maximum speed distribution is presented in Fig. 12. Lips are the articulators that reach the highest values of speed, especially on their left side (ULL and LLL coils). Power variants show faster moves than humming ones. In the power kick, the left upper lip has an average maximum speed of 45 cm/s, but it can reach maximum velocities as high as 90 cm/s. Humming and power snares both involve a bilabial occlusive, however, the order of magnitude of lip speed is smaller (15–17 cm/s for the upper left lip for both variants) than the kicks, possibly because the lips are still engaged in a stricture of close approximation after the release of the occlusion.

The data show that the tongue is almost always involved in the articulation of the explored HBB sounds, either as the main articulator or accompanying the lip dynamics. However, it never reaches the highest speed values of the lips. Our analyses point out that the tongue, either as a whole or in part, is the main articulator for the production of both the humming and the power variant of the hi-hat and rimshot effects, the power inward snare as well as the inhaled cymbal. The regions of the tongue that reach the highest velocities typically match the main place of articulation, i.e., where the occlusion is released. However, in the humming hi-hat, the mid-portion of the tongue appears to be the fastest moving articulator, moving at an average maximum speed of about 20 cm/s. As discussed in the previous section, this is likely the place where the airflow is generated and not the place where the anterior occlusion is released.

The sounds for which the tongue is not the main articulator are both the humming and the power variants of kick and snare, as well as the exhaled cymbal. In these cases, a general tendency seems to emerge that the tongue moves as a whole, with all three regions showing comparable average maximum velocities.

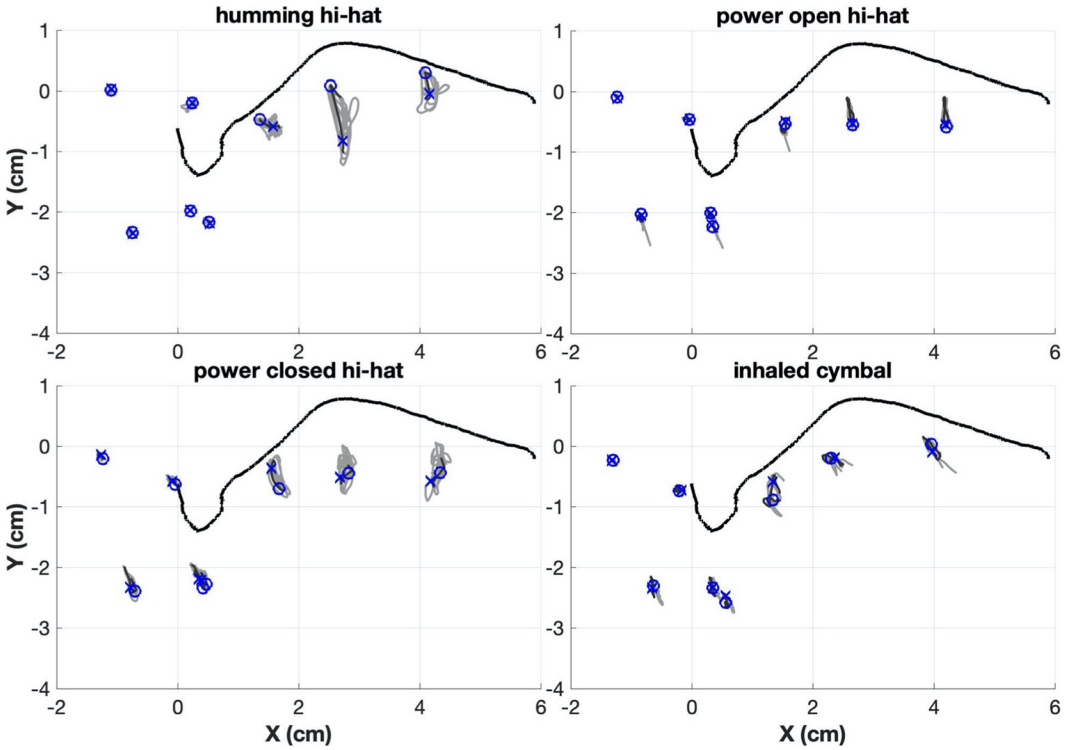

FIG. 8. (Color online) Sagittal (XY) views of trajectories of 5 repetitions of power closed hi-hat, power open hi-hat, humming hi-hat, inhaled cymbal. Legend: see Fig. 6.

The analysis of speed distribution demonstrates limited dynamics for the jaw. This articulator almost never reaches high speeds, moving at an average maximum speed of approximately 5 cm/s across all the examined sounds. The jaw dynamics seem to be quite independent of the dynamics of the left lower lip (LLL coil) in all bilabial effects. The statistical analysis shows that the JAW coil reaches significantly lower maximum speed values than the LLL coil in all these sounds (humming kick: -1.5258 ± 0.0672, $p < 0.001$; power kick: -1.4389 ± 0.0513, $p < 0.001$; humming snare: -1.1822 ± 0.0601, $p < 0.001$; power snare: -0.9178 ± 0.0519, $p < 0.001$; exhaled cymbal: -1.4924 ± 0.1075, $p < 0.001$).

Only in the articulation of two HBB sounds, i.e., power rimshot and power inward snare, does the jaw move more quickly, reaching approximately 10 cm/s for the former and slightly less than 15 cm/s for the latter. In both cases, the jaw dynamics possibly accompanies the lower lip dynamics, as the two articulators (JAW, LML and LLL coils) on average show the same maximum velocities.

D. Phonetic description

A phonetic annotation was performed using the IPA alphabet. The results are presented in Table II.

In general, either the symbols utilized do not belong to French or, if they are present in French, some diacritics were needed, because of the occurrence of perceptible phonetic effects or modifications. A symbol was assigned to each sound. Especially the non-speechlike articulatory and airstream mechanisms required the use of diacritics. Neither the IPA nor the extIPA (Ball *et al.*, 2018) provide a notation for lingual egressive articulations. Hence, the symbol for the corresponding click (always ingressive in speech) was used in combination with the symbol for an egressive airflow. The vibratory aspects revealed by the acoustic investigation (Sec. III A) are annotated as a voiceless bilabial trill (B̥). Due to the lack of a symbol for a lingual egressive mechanism, the difference in airstream mechanism presented in Sec. III B cannot be reported in this table. Further, the frequency of lip vibration of these sounds seems higher than that of speech bilabial trills.

Some sounds presented similarities with French phonemes: bilabial, alveolar and velar stops, labiodental and alveolar fricatives. However substantial features differentiate the HBB sounds from the French phonemes. The power kick is similar to the French [p] in that it is a bilabial stop. It is, however, ejective and lateralized. Similarly, the power closed hi–hat and the power rimshot are similar to the French [t] and [k], except for the airstream mechanism.

IV. DISCUSSION

A. Feasibility and suitability of multimodal synchronized physiological measurements in HBB

Although limited to one subject [as is often the case for HBB, e.g., Blaylock *et al.* (2017) and Proctor *et al.* (2013)], the present study suggests that the recording of multimodal (EMA, EGG, RIP, audio, video) and synchronized data is compatible with HBB production and paramount in the exploration and understanding of the production mechanisms of this peculiar vocal art.

https://doi.org/10.1121/10.0002921

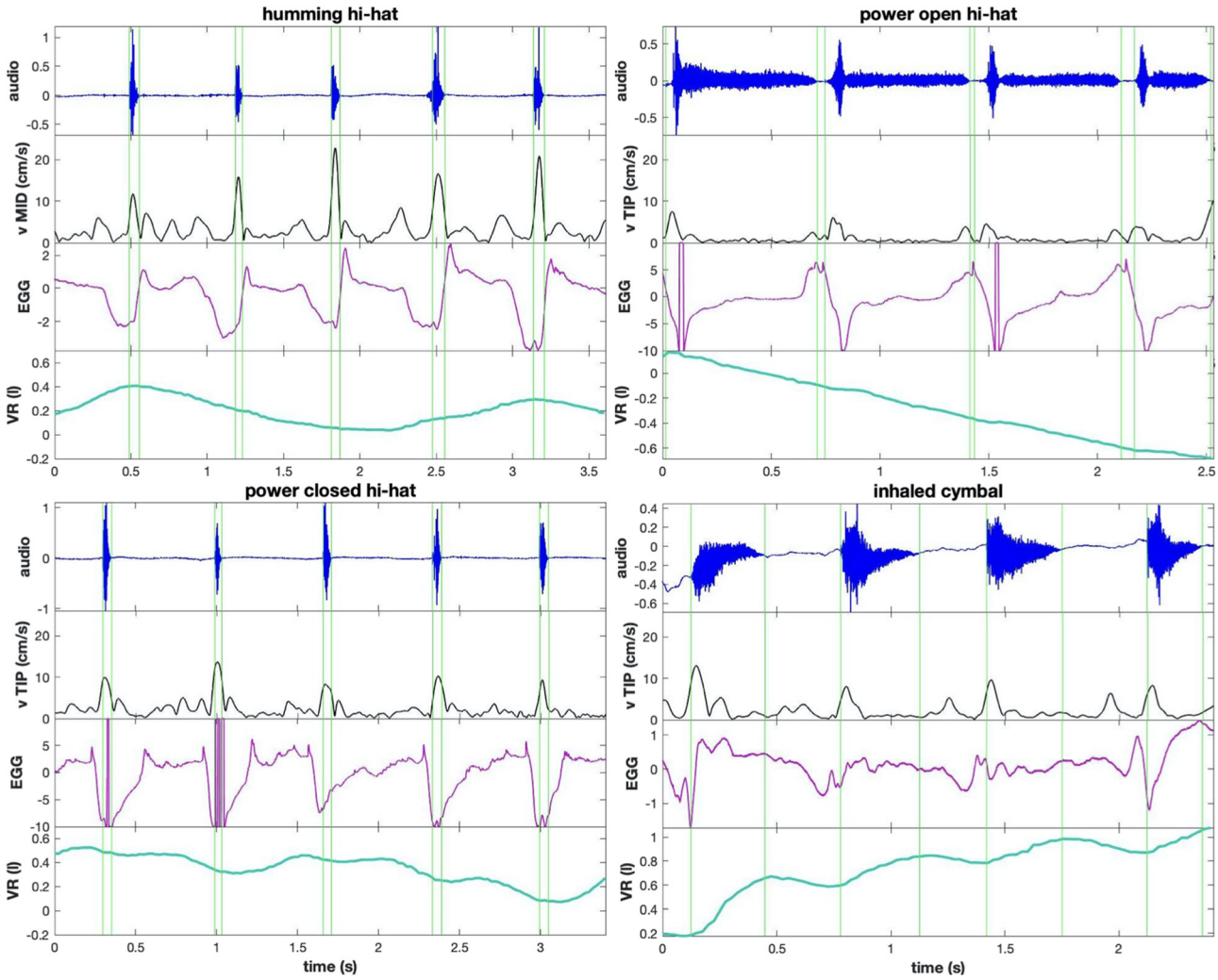

FIG. 9. (Color online) Synchronized audio, speed, EGG, and RIP data of five repetitions of power closed hi-hat, power open hi-hat, humming hi-hat, inhaled cymbal (same as Figs. 3 and 8).

The beatboxer was able to produce more than one hour of sounds with the coils firmly attached to the lips and tongue. The coil wires were uncomfortable for him at first, but he got used to them and managed to produce all the HBB sounds in the protocol. The measurements consisted of three-dimensional articulatory movements. Being able to compute tangential speed with all three spatial components (x, y, z) was particularly relevant for lip dynamics in HBB, which presented several lateral articulations such as lateralization of occlusion release.

B. Boxemes, distinct sound units

The acoustic data outlined different spectral signatures for every sound. These differences were such that an unsupervised classifier was able to automatically detect each sound and correctly assign it to a category in agreement with those provided by the beatboxer. The articulatory and ventilatory data also showed different behavior that distinguishes each sound from the others, in terms of place and/or manner of articulation, and airstream mechanism. Our results indicate that each one of the twelve HBB drum sounds investigated in this study was substantially different from the others, supporting the idea that they make sense as distinct sound units. We propose that these sound units be called *boxemes*, by analogy with speech phonemes. They constitute the building blocks of a HBB musical phrase. Considering HBB as a musical language structured similarly to human speech calls for future research that goes far beyond the present study.

The few studies that have proposed an IPA transcription of speech-like HBB sounds show some degree of agreement with the transcription proposed in the present investigation. Kicks that correspond to power kick in this study often involve bilabial ejectives [p'] (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013), snares corresponding to power snare are a double articulation of a bilabial ejective and labiodental fricative [p'f] or [pf'] (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013), hi-hats corresponding to power closed and open hi-hat often involve an alveolar stop [t] or [ts] (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013), rimshots corresponding to power

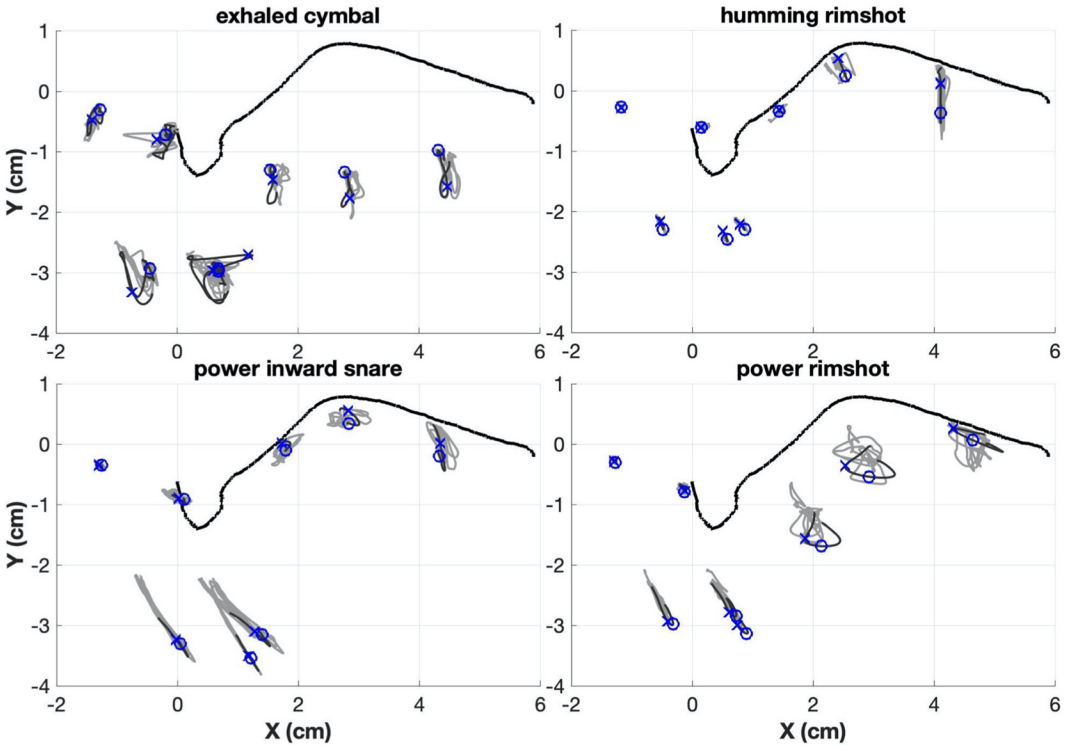

FIG. 10. (Color online) Sagittal (XY) views of trajectories of five repetitions of humming rimshot, power rimshot, exhaled cymbal, power inward snare. Legend: see Fig. 6.

rimshot are often velar stops [k] or [k'] (Proctor *et al.*, 2013). Even though Blaylock *et al.* (2017) do not provide an IPA transcription, similarities in non-linguistic articulations can be found between power inward snare and inward K, humming kick and lip pop, humming hi-hat, and forced hi-hat. Such an agreement suggests that similar acoustic and/or articulatory strategies are used for the same sound among different beatboxers, regardless of the beatboxer's native language.

C. Complex articulatory behaviors

Our data demonstrated a variety of articulatory gestures, many of which elicited labial dynamics. Lingual dynamics was also clearly manifest, both when the tongue was the main articulator and when accompanying lip dynamics. This suggests complex tongue-lip synergies. On the contrary, jaw dynamics was often limited in our corpus of HBB drum sounds, possibly due to the absence of vocalic sounds.

The investigated sounds seem to be produced on two different time scales: 9 sounds were short, generally not exceeding 200 ms, 3 were longer, up to 750 ms. However, even the shorter sounds could be produced as a double articulation of a plosive attack, generally due to the release of a complete occlusion, followed by a friction noise.

Our data showed the use of quite a wide variety of manners of articulation. Despite the small number of HBB drum sounds explored, ejectives, clicks, stops and fricatives were observed. Most of the produced sounds did not belong to the phonology of French, the language spoken by our subject. Some are found in other world's languages (Ladefoged and Maddieson, 1996), others have never been attested in any language.

D. Mastering pulmonic and non–pulmonic airstreams

In agreement with the existing literature (Blaylock *et al.*, 2017; Proctor *et al.*, 2013), both egressive and ingressive airstreams were observed. In some cases, the opposite airstream was used as compared to what is generally observed for the speech counterparts of the same sounds. This occurred mainly in the articulation of stop and fricative sounds, where a pulmonic ingressive airstream could be used. Stowell and Plumbley (2010) also described the use in HBB of a given sound produced with both pulmonic ingressive and egressive airstream in the case of oral stops.

All the humming sounds were produced via a lingual ingressive (velaric) or egressive airstream. The latter has already been described by Blaylock *et al.* (2017), but it has never been observed in speech so far. A lingual airflow initiation grants some dissociation between sound production and articulation. This allows the beatboxer to perform multiple actions at the same time, such as breathing or producing a melodic line through the nose without being silent.

However, this has a cost in terms of intensity, as humming variants were always significantly quieter than their power counterparts.

E. Evidence for ejective productions

Our data argue in favor of an ejective production of all the non–humming sounds that imply a bilabial occlusion

https://doi.org/10.1121/10.0002921

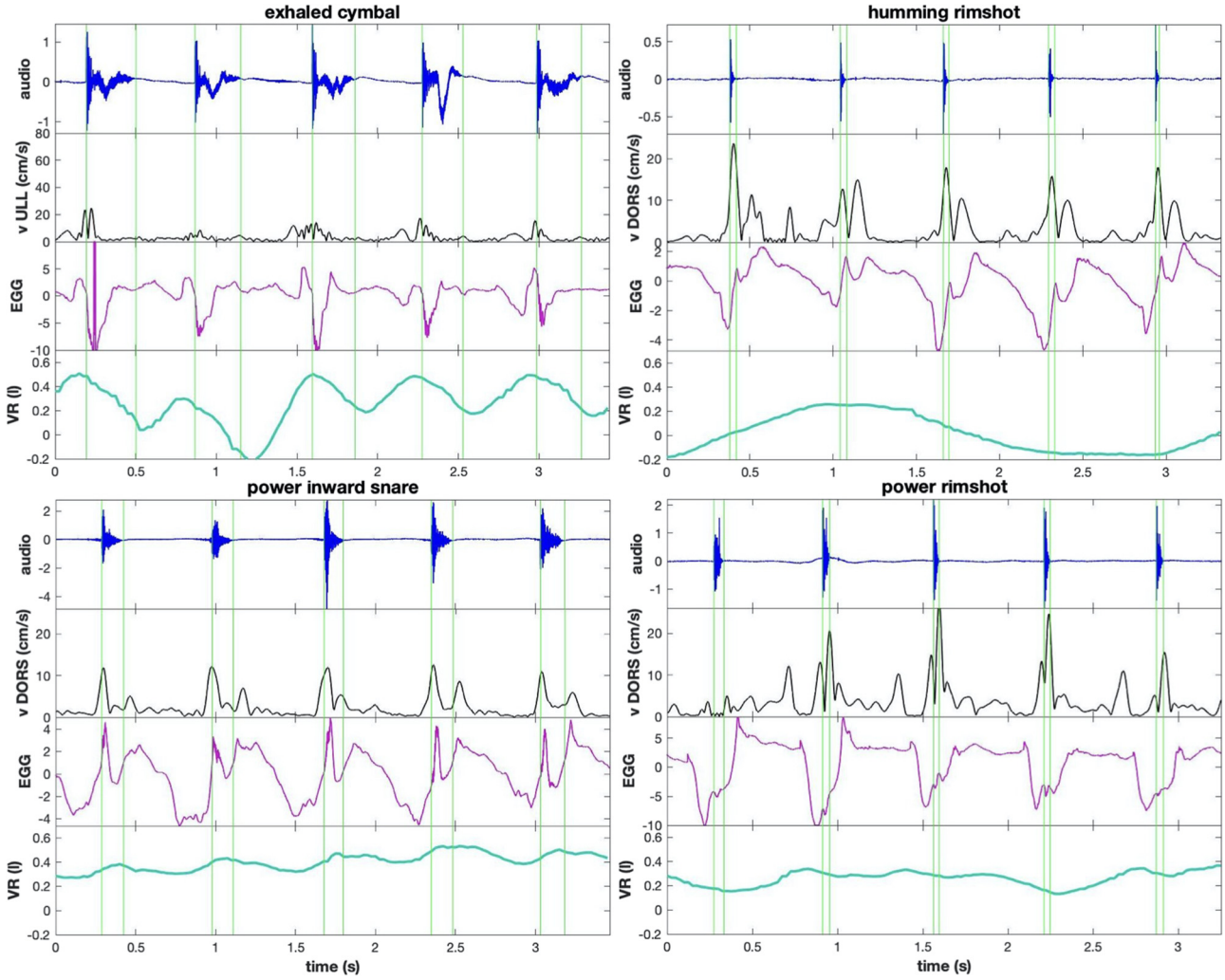

FIG. 11. (Color online) Synchronized audio, speed, EGG, and RIP data of five repetitions of humming rimshot, power rimshot, exhaled cymbal, power inward snare (same as Figs. 3 and 10).

(e.g., kick and snare effects), or alveolar occlusion with egressive airstream (e.g., power closed hi-hat). An ejective production of the power rimshot could not be precluded as a possibility. As mentioned in Sec. I, the use of ejectives in HBB was already attested [even though not systematically employed by all beatboxers (Patil *et al.*, 2017)] by a few articulatory studies that exploited different imaging techniques [video-fiberscopy: De Torcy *et al.* (2014), Dehais Underdown *et al.* (2019), Sapthavee *et al.* (2014); MRI: Blaylock *et al.* (2017), Patil *et al.* (2017), Proctor *et al.* (2013)]. In particular, these studies characterize several kick and snare sounds as ejectives, when produced as bilabial occlusives, closed hi-hats as alveolar ejectives as well as a rimshot sound as a velar ejective. Proctor *et al.* (2013) characterized three kick sounds as stiff ejectives, with different amounts of lingual retraction during laryngeal lowering and a different final lingual posture. They also suggested that tongue and larynx may be used in concert to produce a more effective pushing action. Our data support this hypothesis of a lingual action in the articulation of ejective sounds.

F. Vibration and lateralization

In some cases, acoustic data revealed a clear vibratory pattern that did not originate from glottal vibration as attested by the EGG signal. The combination of acoustic, articulatory, and video data suggests that the vibration was produced at the place of occlusion, namely, the lip area for the humming kick and snare, power snare, and exhaled cymbal, and possibly the lateral rim of the tongue for the power inward snare.

All the bilabial sounds were laterally released on the left side of the lips. This consistent lateralization may be explained by the fact that beatboxers need to control lip tension in order to produce self-oscillation at adequate vibratory frequency (Stowell and Plumbley, 2010). Shortening the lip portion that can vibrate may provide a better control and the possibility to produce vibrations at higher frequencies. The resulting effect is reminiscent of the way vocal folds are controlled to modify f0.

Furthermore, the lateralization of the bilabial occlusion release seemed to affect the articulation of the following fricative at least in the case of the power snare. Here, the labiodental fricative was also articulated on the left.

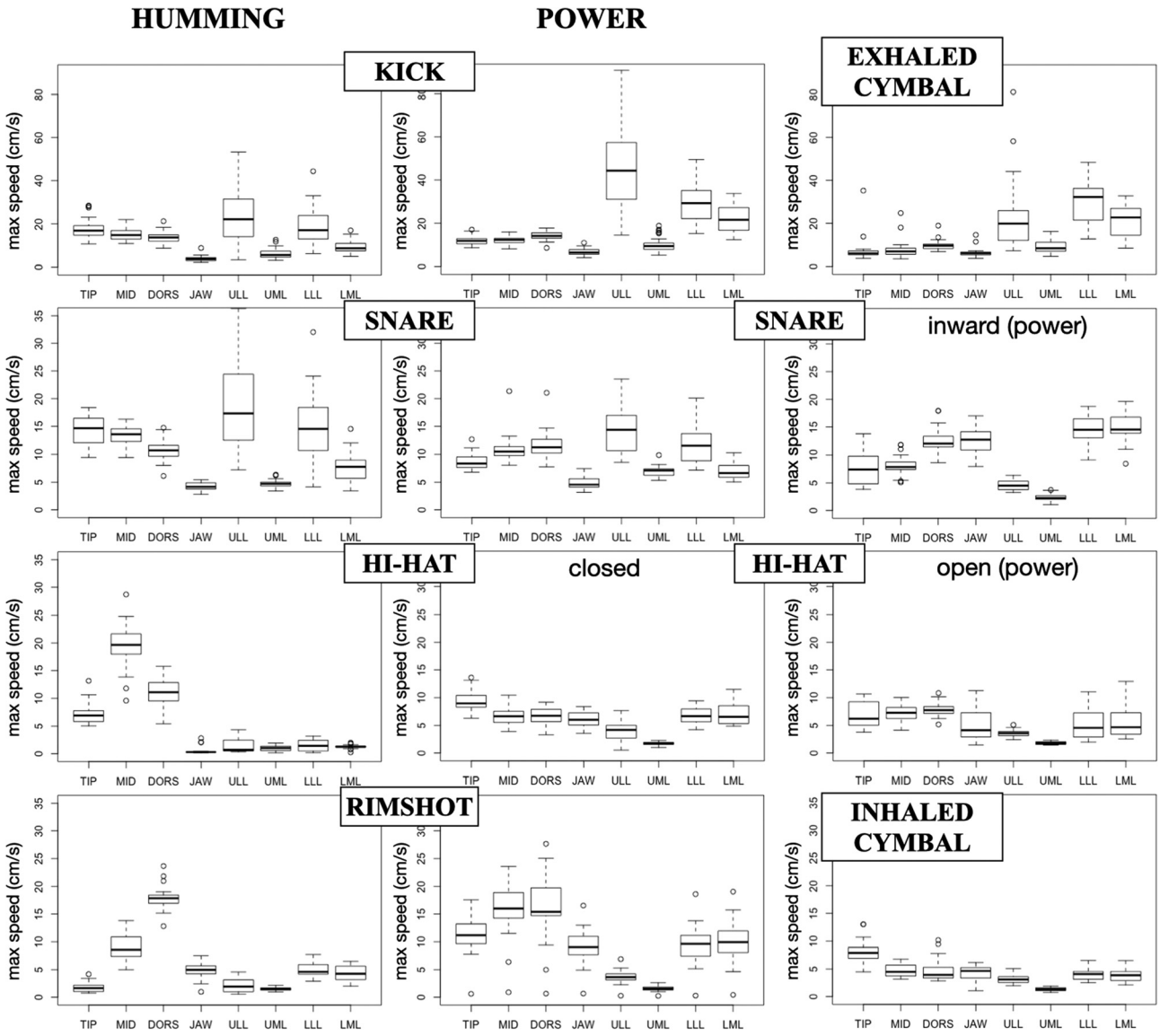

FIG. 12. Maximum speed distribution (in cm/s) of the coils for the twelve HBB sounds. Left column: humming variants; center and right column: power variants and cymbals. Note that the first row of panels has a wider *y* axis scale, because of faster lip movements for kick and exhaled cymbal sounds.

G. HBB sound annotation

Using IPA to annotate HBB sounds was not straightforward, in agreement with Blaylock *et al*. (2017). Some basic HBB sounds may stem from speech sounds (e.g., classic kick—or power kick according to our beatboxer's terminology). They may share the same mechanisms, as suggested by Proctor *et al*. (2013), but they are substantially modified to induce a non-linguistic or para-linguistic connotation. Further, our data displayed the use of sources of vibration other than the glottal one, suggesting that the simple distinction between voiced and voiceless sounds is not sufficient in HBB to fully characterize the acoustic production. Moreover, even if the international HBB community shares a considerable amount of coded sounds, a prerogative of each beatboxer is to experiment with their own vocal instrument to create new sounds, never produced before and more and more difficult to articulate. As a consequence, a much more subtle and adapted notation system is needed in order to capture the acoustic and articulatory richness of HBB production. An articulatory-based writing system seems promising and has recently been used for beatbox-sound automatic recognition purpose (Evain *et al*., 2020).

V. CONCLUSION AND PERSPECTIVES

Acoustic, articulatory, and ventilatory properties of twelve different HBB drum sounds were investigated on a French beatboxer. Electromagnetic articulography, an experimental technique widely used in speech research, was successfully used to capture the articulatory dynamics. It was combined with acoustic measurements, electroglottography, and respiratory inductance plethysmography to get a deeper understanding of articulatory and airstream mechanisms underlying these complex vocal sound productions.

https://doi.org/10.1121/10.0002921

TABLE II. Phonetic characterization and brief articulatory description of the HBB sounds.

		Description				
Sound	IPA	Voicing	Airstream	Place	Manner	Articulation
Humming kick	[ʘ͡ʙˡ↑]	voiceless	lingual egressive	lateral bilabial	stop	double
		voiceless	lingual egressive	lateral bilabial	trill	
Humming snare	[ʘ͡ʙˡːˑ↑]	voiceless	lingual egressive	lateral bilabial	stop	double
		voiceless	lingual egressive	lateral bilabial	trill	
Power kick	[pʼˡ]	voiceless	glottalic egressive	lateral bilabial	stop	simple
Power snare	[p͡fˡːˑ]	voiceless	glottalic egressive	lateral bilabial	stop	double
		voiceless	pulmonic egressive	lateral labio–dental	fricative	
Exhaled cymbal	[ʙˡsːˑ]	voiceless	glottalic egressive	lateral bilabial	trill	double
		voiceless	pulmonic egressive	laminal	fricative	
Humming hi-hat	[|↑]	voiceless	lingual egressive	alveolar	stop	simple
Power closed hi-hat	[tʼ]	voiceless	glottalic egressive	alveolar	stop	simple
Power open hi-hat	[tʼs͡ːˑ]	voiceless	glottalic egressive	alveolar	stop	double
		voiceless	pulmonic egressive	alveolar	fricative	
Inhaled cymbal	[sːˑ↓]	voiceless	pulmonic ingressive	alveolar	stop	double
		voiceless	pulmonic ingressive	alveolar	fricative	
Humming rimshot	[||]	voiceless	lingual ingressive	lateral	stop	simple
Power rimshot	[kʼ]	voiceless	glottalic egressive	velar	stop	simple
Power inward snare	[k͡ɬˑ↓]	voiceless	pulmonic ingressive	velar	stop	double
		voiceless	pulmonic ingressive	lateral	fricative	

In agreement with the existing literature, a wide variety of articulatory gestures were observed, most of which do not belong to the phonology of the beatboxer's language, nor to any known phonology. Our data revealed the use of multiple airstream mechanisms, the possibility of dissociating breathing and sound production, a pronounced labial dynamics, or a lingual dynamics that accompanies the labial dynamics when the principal articulator is not the tongue.

The notion of *boxeme* has been suggested, as building blocks of human beatboxing considered as a musical language. This calls for further research.

This investigation was conducted with a single beatboxer. The next step is to collect and analyze HBB articulatory behavior from multiple beatboxers with several training levels in order to generalize our findings and relate them to the HBB level of practice. It would also be very interesting to study the impact of the native language on vocal drum sound production.

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency and NeuroCoG in the framework of the "Investissements d'avenir" program (Grant No. ANR-15-IDEX-02). Our gratitude goes to the beatboxer who kindly agreed to participate in the experiment. We thank Maëva Garnier who contributed to the protocol elaboration and to the recordings. We thank the Associate Editor Susanne Fuchs and three anonymous reviewers for their many helpful suggestions and corrections. We are very grateful to Joe Wolfe and Annie Devlin for proofreading.

Ball, M. J., Howard, S. J., and Miller, K. (**2018**). "Revisions to the extipa chart," J. Int. Phon. Assoc. **48**(2), 155–164.

Barbier, G., Baum, S. R., Ménard, L., and Shiller, D. M. (**2020**). "Sensorimotor adaptation across the speech production workspace in response to a palatal perturbation," J. Acoust. Soc. Am. **147**(2), 1163–1178.

Blaylock, R., Patil, N., Greer, T., and Narayanan, S. S. (**2017**). "Sounds of the human vocal tract," in *Proceedings of Interspeech*, pp. 2287–2291.

Boersma, P. (**2006**). "Praat: Doing phonetics by computer," http://www.praat.org/ (Last viewed April 8, 2020).

Brunner, J., Hoole, P., Guenther, F., and Perkell, J. S. (**2010**). "Dependency of compensatory strategies on the shape of the vocal tract during speech perturbed with an artificial palate," Proc. Mtgs. Acoust. **9**, 060003.

Calabrese, P., Besleaga, T., Eberhard, A., Vovc, V., and Baconnier, P. (**2007**). "Respiratory inductance plethysmography is suitable for voluntary hyperventilation test," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE), pp. 1055–1057.

De Torcy, T., Clouet, A., Pillot-Loiseau, C., Vaissiere, J., Brasnu, D., and Crevier-Buchman, L. (**2014**). "A video-fiberscopic study of laryngopharyngeal behaviour in the human beatbox," Logoped. Phoniatr. Vocol. **39**(1), 38–48.

Dehais Underdown, A., Buchman, L., and Demolin, D. (**2019**). "Acoustico-physiological coordination in the Human Beatbox: A pilot study on the beatboxed Classic Kick Drum," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, https://hal.archives-ouvertes.fr/hal-02284132 (Last viewed August 25, 2020).

Eberhard, A., Calabrese, P., Baconnier, P., and Benchetrit, G. (**2001**). "Comparison between the respiratory inductance plethysmography signal derivative and the airflow signal," in *Frontiers in Modeling and Control of Breathing* (Springer, Berlin), pp. 489–494.

Evain, S., Contesse, A., Pinchaud, A., Schwab, D., Lecouteux, B., and Henrich Bernardoni, N. (**2020**). "Reconnaissance de parole beatboxée à l'aide d'un système HMM-GMM inspiré de la reconnaissance automatique de la parole" ("Beatboxed speech recognition using a HMM-GMM system based on automatic speech recognition"), in *Journées d'Études sur la Parole (Speech Study Days)*, Vol. 1 of *6th JEP-TAL-RECITAL Conference*, edited by C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla, and S. Schneider, ATALA, Nancy, France, https://hal.archives-ouvertes.fr/hal-02798538 (Last viewed August 25, 2020), pp. 208–216.

Friberg, A., Lindeberg, T., Hellwagner, M., Helgason, P., Salomão, G. L., Elowsson, A., Lemaitre, G., and Ternström, S. (**2018**). "Prediction of three articulatory categories in vocal sound imitations using models for auditory receptive fields," J. Acoust. Soc. Am. **144**(3), 1467–1483.
Helgason, P. (**2014**). "Sound initiation and source types in human imitations of sounds," in *Proceedings of FONETIK*, pp. 83–88.
Henrich Bernardoni, N., and Paroni, A. (**2020**). "Vocal drum sounds in Human Beatboxing: An acoustic and articulatory exploration using electromagnetic articulography," Zenodo, Dataset https://doi.org/10.5281/zenodo.4264747 (Last viewed September 11, 2020).
Herbst, C. T. (**2020**). "Electroglottography—An update," J. Voice **34**(4), 503–526.
Hieronymus, J. L. (**1993**). "Ascii phonetic symbols for the world's languages: Worldbet," J. Int. Phon. Assoc. **23**, 72.
Hothorn, T., Bretz, F., and Westfall, P. (**2008**). "Simultaneous inference in general parametric models," Biometr. J.: J. Math. Meth. Biosci. **50**(3), 346–363.
Kapur, A., Benning, M., and Tzanetakis, G. (**2004**). "Query-by-beat-boxing: Music retrieval for the DJ," in *Proceedings of the International Conference on Music Information Retrieval*, pp. 170–177.
Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford).
Leanderson, R., and Sundberg, J. (**1988**). "Breathing for singing," J. Voice **2**(1), 2–12.
Leanderson, R., Sundberg, J., and von Euler, C. (**1984**). "Effects of diaphragm activity on phonation during singing," in *Transactions of the 13th Annual Symposium on Care of the Professional Voice* (The Voice Foundation, New York), pp. 165–169.
Maaten, L. v. d., and Hinton, G. (**2008**). "Visualizing data using t-sne," J. Mach. Learn. Res. **9**, 2579–2605.
MATLAB (**2018**). MathWorks: Bioinformatics Toolbox: User's Guide (R2018b), Cybernet Systems Co., Ltd.
Patil, N., Greer, T., Blaylock, R., and Narayanan, S. S. (**2017**). "Comparison of basic beatboxing articulations between expert and novice artists using real-time magnetic resonance imaging," in *Proceedings of Interspeech*, pp. 2277–2281.
Picart, B., Brognaux, S., and Dupont, S. (**2015**). "Analysis and automatic recognition of human beatbox sounds: A comparative study," in *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 4255–4259.
Proctor, M., Bresch, E., Byrd, D., Nayak, K., and Narayanan, S. (**2013**). "Paralinguistic mechanisms of production in human 'beatboxing': A real-time magnetic resonance imaging study," J. Acoust. Soc. Am. **133**(2), 1043–1054.
R Core Team (**2013**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/ (Last viewed April 8, 2020).
Rothenberg, M. (**1992**). "A multichannel electroglottograph," J. Voice **6**(1), 36–43.
Sapthavee, A., Yi, P., and Sims, H. S. (**2014**). "Functional endoscopic analysis of beatbox performers," J. Voice **28**(3), 328–331.
Savariaux, C., Badin, P., Samson, A., and Gerber, S. (**2017**). "A comparative study of the precision of Carstens and Northern Digital Instruments electromagnetic articulographs," J. Speech Lang. Hear. Res. **60**(2), 322–340.
Sinyor, E., Rebecca, C. M., Mcennis, D., and Fujinaga, I. (**2005**). "Beatbox classification using ace," in *Proceedings of the International Conference on Music Information Retrieval*, Citeseer.
Stowell, D., and Plumbley, M. D. (**2010**). "Delayed decision-making in real-time beatbox percussion classification," J. New Music Res. **39**(3), 203–213.
Tiede, M., Chen, W.-R., and Whalen, D. (**2019a**). "Fundamental frequency correlates with head movement evaluated at two contrasting speech production rates," J. Acoust. Soc. Am. **146**(4), 3085–3085.
Tiede, M., Mooshammer, C., and Goldstein, L. (**2019b**). "Noggin nodding: Head movement correlates with increased effort in accelerating speech production tasks," Front. Psychol. **10**, 2459.
Von Luxburg, U. (**2007**). "A tutorial on spectral clustering," Stat. Comput. **17**(4), 395–416.