# Beatboxer, est-ce parler ? Ce que nous en dit l'étude de la dynamique articulatoire d'un beatboxer

Annalisa Paroni, Nathalie Henrich Bernardoni, Christophe Savariaux, Pierre Baraduc, Hélène Lœvenbruck

▶ **To cite this version:**

Annalisa Paroni, Nathalie Henrich Bernardoni, Christophe Savariaux, Pierre Baraduc, Hélène Lœvenbruck. Beatboxer, est-ce parler ? Ce que nous en dit l'étude de la dynamique articulatoire d'un beatboxer. JEP-TALN-RECITAL 2020 - 6e conférence conjointe 33e Journées d'Études sur la Parole, 27e Traitement Automatique des Langues Naturelles, 22e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, Jun 2020, Nancy, France. pp.472-479. hal-02798574v3

**HAL Id: hal-02798574**
**https://hal.science/hal-02798574v3**

Submitted on 23 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),*
*Traitement Automatique des Langues Naturelles (TALN, 27e édition),*
*Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)*
*Nancy, France, 08-19 juin 2020*

# Beatboxer, est-ce parler ? Ce que nous en dit l'étude de la dynamique articulatoire d'un beatboxeur

Annalisa Paroni[1] Nathalie Henrich Bernardoni[1] Christophe Savariaux[1]
Pierre Baraduc[1] Hélène Lœvenbruck[2]
(1) Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA–lab, F–38000 Grenoble
(2) Univ. Grenoble Alpes, Univ. Savoie Mont–Blanc, CNRS, LPNC, F–38000 Grenoble
annalisa.paroni@gipsa-lab.fr nathalie.henrich@gipsa-lab.fr

RÉSUMÉ

Les consonnes plosives sont parmi les phonèmes les plus représentés dans l'inventaire phonologique des langues du monde. Outre leur rôle linguistique, elles remplissent également un rôle paralinguistique dans la pratique instrumentale et vocale, notamment au sein de la pratique vocale du Human Beatbox. Cet article apporte un éclairage sur les similitudes et différences dans la dynamique articulatoire de trois consonnes plosives du français et des sons percussifs correspondants du Human Beatbox. Si ces deux modes de production vocale ont une racine commune, une dynamique articulatoire différente est mise en évidence pour le Human Beatbox. Nous retrouvons des indices d'un mécanisme éjectif, qui a un impact sur la dynamique linguale.

ABSTRACT

**Beatboxing, is it talking ? What the study of the articulatory dynamics of a beatboxer tells us**

Plosive consonants are among the most commonly-found phonemes in the phonological inventory of the world's languages. In addition to their linguistic role, they also fulfil a paralinguistic role in instrumental and vocal practice, especially within the practice of Human Beatbox. This article sheds light on the similarities and differences in the articulatory dynamics of three plosive consonants of French and the corresponding percussive sounds of Human Beatbox. These two modes of vocal production share a common root. Yet, different articulatory dynamics are highlighted for Human Beatbox. Clues of an ejective mechanism are found, that impact the lingual dynamics.

MOTS-CLÉS : plosive, son percussif, beatboxer, human beatbox, dynamique articulatoire, articulographie électromagnétique.

KEYWORDS: plosive, percussive sound, beatboxing, human beatbox, articulatory dynamics, electromagnetic articulography.

## 1 Introduction

Les consonnes plosives /p, t, k/ font partie des phonèmes les plus représentés dans l'inventaire phonologique des langues du monde (Maddieson & Disner, 1984). Du point de vue articulatoire, les plosives sont produites grâce au relâchement plus ou moins soudain d'une occlusion du conduit vocal (CV). Cette occlusion se situe dans le conduit oral, au niveau des lèvres, du palais dur, du velum ou dans l'espace pharyngé. Ce point d'articulation a une influence sur les caractéristiques acoustiques du bruit de plosion (Forrest *et al.*, 1988) qui joue un rôle important dans la discrimination des consonnes

Volume 1 : Journées d'Études sur la Parole, pages 472–479.
HAL : hal-02798574.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

plosives (CNET-ENST, 1989). En français les points d'articulation discriminatifs pour les plosives sont les lèvres (plosives bilabiales), les alvéoles (plosives alvéolaires) ou le vélum (plosives vélaires). A la marge de leur rôle linguistique, les plosives ont aussi un rôle paralinguistique. Les syllabes avec consonnes plosives sont utilisées dans la pratique instrumentale, pour le jeu des instruments à vent et des percussions. Elles soutiennent également la pratique vocale, comme par exemple dans le cas du scat ou du konnakol indien. L'origine linguistique de ces sons a un impact sur l'articulation de ces mêmes sons quand ils sont utilisés de façon paralinguistique. Certaines études se sont intéressées aux différences articulatoires des plosives utilisées dans la pratique instrumentale en fonction de la langue maternelle de l'instrumentiste (Lamkin, 2005; Heyne & Derrick, 2014, 2015; Heyne *et al.*, 2019).

Dans cette étude, nous nous intéressons à la pratique du Human Beatbox (HBB), un art vocal émergent qui s'appuie sur l'instrument vocal humain pour produire toutes sortes de son dans le but de faire de la musique. Les sons percussifs y sont très exploités. Bien souvent, l'apprentissage du HBB commence par un travail sur des plosives, des syllabes ou des phrases de la parole. Tel est le cas du kick, le son utilisé pour imiter la grosse caisse de la batterie, appris à partir d'un [p], du hi–hat, qui reproduit le son du charleston, travaillé à partir d'un [t] ou [ts], ou encore de la technique du rimshot, sur la base d'un [k]. Quelles sont les similitudes et les différences dans la dynamique articulatoire entre parole et HBB ? La base de données exploitée pour apporter une première réponse à cette question est présentée en Section 2. Nous présenterons les résultats en terme de trajectoires articulatoires en Section 3. Nous comparerons la production des plosives et des sons percussifs ayant même point d'articulation.

# 2 Matériel et méthodes

Nous présentons les résultats d'une étude pilote menée sur un sujet beatboxeur à partir de la technique expérimentale de l'articulographie électromagnétique (EMA) combinée à des mesures acoustiques, électroglottographiques et ventilatoires. Au sein de cet article, nous nous concentrons uniquement sur les mesures articulatoires. Le sujet est un beatboxeur de 28 ans ayant une pratique amateur de 9 ans en HBB et dont la langue maternelle est le français.

Nous nous sommes intéressés à la production en parole des plosives /p, b, t, d, k, g/ et de leur équivalent en HBB en termes de point d'articulation : le kick, le hi–hat et le rimshot. Trois phrases – *des p'tits cookies des gros cookies* (Cookies) ; *pâtes au pesto* (Pâtes) ; *boots and cats* (BootsAndCats) – ont été répétées plusieurs fois à partir d'un mode parlé jusqu'à un mode beatboxé, et inversement (Cookies, 8 répétitions parlées et 8 répétitions beatboxées ; Pâtes : 8 répétitions parlées et 17 répétitions beatboxées ; BootsAndCats : 16 répétitions parlées et 23 répétitions beatboxées). Ces phrases ont été choisies car elles sont assez couramment utilisées par les beatboxeurs dans l'apprentissage de leur art vocal.

Les enregistrements ont eu lieu dans la salle semi–anéchoique de la plateforme BEDEI du laboratoire GIPSA–lab de Grenoble, lieu de recherches biomédicales autorisé par l'ARS Auvergne–Rhône–Alpes depuis Mars 2016. Le protocole expérimental a reçu un avis favorable du Comité d'Ethique pour les Recherches Grenoble Alpes. Après signature du consentement éclairé, le sujet a été équipé d'un gilet pour la mesure pléthysmographique respiratoire à variation d'inductance (système VISURESP, RBI). Il a été installé dans la salle semi–anéchoique sur une chaise adaptée pour la stabilisation de la tête à l'intérieur du champ magnétique de l'articulographe électromagnétique 3D (EMA WAVE, NDI).

Pour le recueil des données articulatoires, 12 bobines ont été collées sur des points de chair d'intérêt :

— 3 bobines sur la langue dans le plan médio–sagittal : 1 dans la région apicale (TIP), 1 dans la région dorso–palatale (MID) et 1 dans la région dorso–vélaire (DORS) ;
— 1 bobine sur la mâchoire (JAW) rattachée à l'incisive inférieure ;
— 4 bobines sur les lèvres de façon médiane ou latérale : 2 sur la lèvre supérieure (ULL et ULM) et 2 sur la lèvre inférieure (LLL et LLM) ;
— 4 bobines de référence : 1 sur l'incisive supérieure, 1 derrière chaque oreille et 1 sur le front.

Le signal EMA a été enregistré à $400Hz$. Des électrodes de mesure du contact glottique par électroglottographie (EGG) ont été placées sur son cou au niveau du larynx. Un microphone AKG et un microphone omnidirectionnel pré-polarisé 1/2" (B&K 4189) connecté à un pré–amplificateur (B&K 2669C) et à un amplificateur de conditionnement NEXUS (B&K 2690) ont été positionnés à environ 20 cm de la bouche du beatboxeur pour mesurer le signal audio et son intensité. Une caméra a été placée face au sujet pour des enregistrements vidéo à 25 images par seconde.

Les données audio ont été segmentées manuellement et annotées sous Praat (Boersma, 2006). Les annotations phonétiques ont permis de repérer les instants où avaient lieu les bursts des plosives. Ces marqueurs temporels ont été sauvegardés dans un fichier TextGrid. Les données articulatoires ont ensuite été analysées sous MATLAB (MATLAB, 2018). Afin de visualiser la variabilité des mouvements des articulateurs, la trajectoire médiane et l'écart interquartile des bobines d'intérêt ont été calculés en prenant en compte toutes les occurrences du même son pour un même mode de production, en choisissant comme instant de référence le moment du burst détecté sur les données audio. Les données EMA sont mesurées en 3D, mais nous avons choisi de présenter ici l'analyse selon l'axe vertical (y) indiquant les mouvements bas–haut. La médiane de la vitesse tangentielle (3D) et l'écart interquartile ont été également calculés.

Notre étude de la dynamique articulatoire porte sur l'analyse de l'évolution dans le temps de la trajectoire d'une bobine donnée qui se déplace dans l'espace, ainsi que de sa vitesse. Nous menons cette analyse sur une fenêtre temporelle de $400ms$ pour la phrase Cookies et $800ms$ pour BootsAndCats, centrée sur le burst.

# 3 Résultats

L'analyse des données articulatoires (EMA et vidéo) montre que le lieu d'articulation de chacune des plosives reste sensiblement le même en parole et en HBB.

Les Figures 1 et 2 illustrent la dynamique articulatoire de la réalisation de la plosive /p/ et de son équivalent percussif, le kick, prononcée dans la phrase Cookies dans sa version parlée "*des **p**'tits cookies ...*" ou réalisée en HBB par la série "***p** t k t t **p** k t*". Ces sons sont obtenus par occlusion complète des lèvres. Dans les deux modes de production (parlé et beatboxé), cette occlusion précède le moment du burst d'environ $100ms$. En parole, un comportement articulatoire reproductible est observé pour les différentes répétitions de la phrase. Il est mis en évidence sur la Figure 1 par un écart interquartile faible. Un comportement plus variable est observé en HBB à proximité du burst pour ce même point d'articulation (milieu des lèvres). Il est intéressant de noter qu'un comportement homogène (écart interquartile faible) similaire à celui observé en parole se retrouve au niveau de la moité gauche des lèvres (bobines LL). Le relâchement de ce son percussif bilabial est donc latéralisé à gauche, ce que confirment les données vidéo.

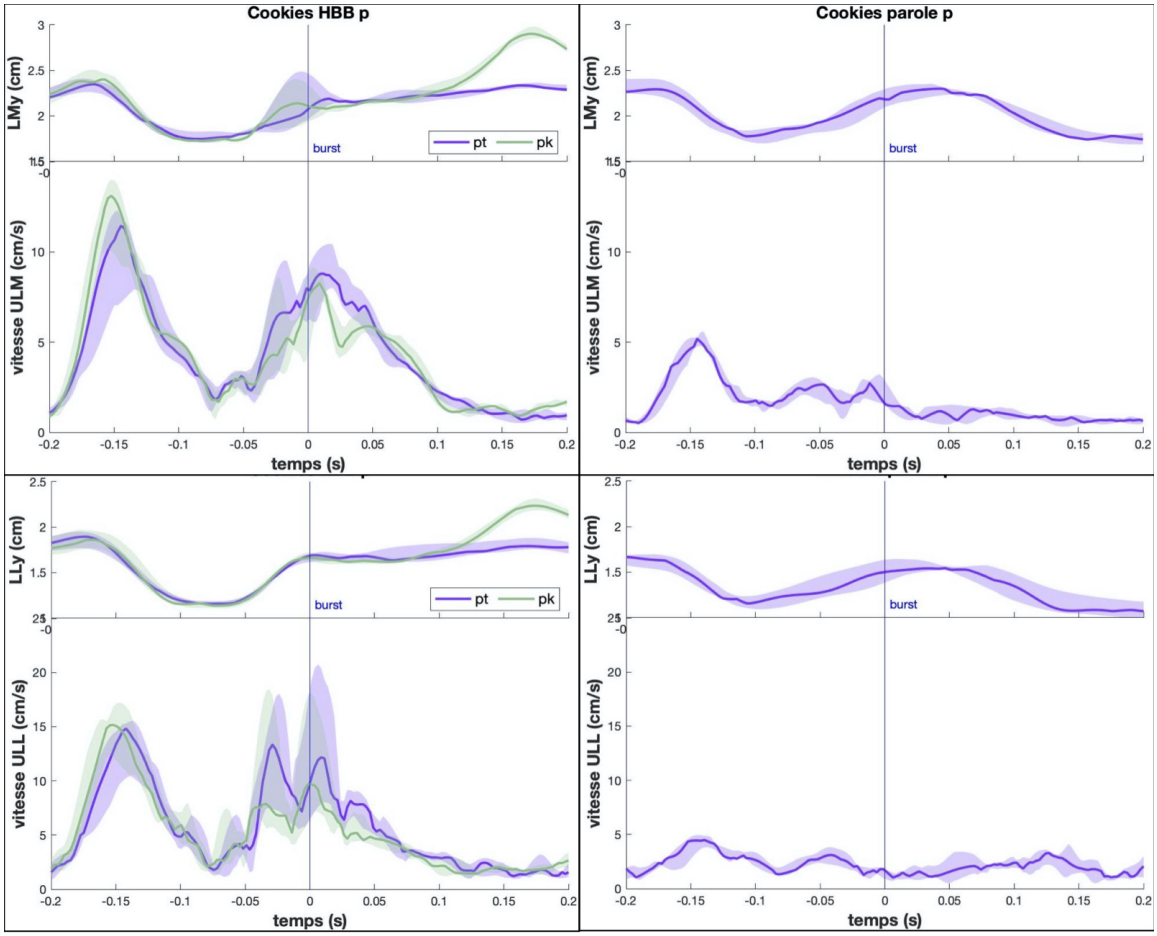


FIGURE 1 – Haut : médiane et écart interquartile de la distance interbobine du milieu des lèvre (LM), du côté gauche des lèvres (LL) au long de l'axe y, calculés à partir de toutes les occurences de la plosive bilabiale pour la phrase Cookies et sa contrepartie beatboxée. Bas : médiane et écart interquartile de la vitesse tangentielle (3D) de la bobine d'intérêt.

La vitesse de relâchement de l'occlusion bilabiale est nettement plus grande en HBB qu'en parole, avec une augmentation de la vitesse de la bobine rattachée au milieu de la lèvre du haut (ULM sur Fig. 1) qui peut atteindre 10 cm/s aux alentours du burst, contre 2,5 cm/s en parole. Quant à la bobine de la lèvre supérieure gauche (ULL sur Fig. 1), plus proche du lieu du relâchement de l'occlusion en HBB, elle peut atteindre 20 cm/s au moment du burst, alors qu'en parole sa vitesse est quasiment nulle. Une différence entre parole et HBB est également observée au niveau de la dynamique linguale, ainsi que l'illustre la Figure 2 pour la bobine posée sur le dos de la langue. Si le mouvement de cette bobine suit le burst en parole, il le précède en HBB. Nous pouvons noter un mouvement de déplacement vertical du dos de la langue marqué en HBB, s'accompagnant d'une variabilité très faible de la trajectoire au moment du burst, mais différent en fonction du son qui suit.

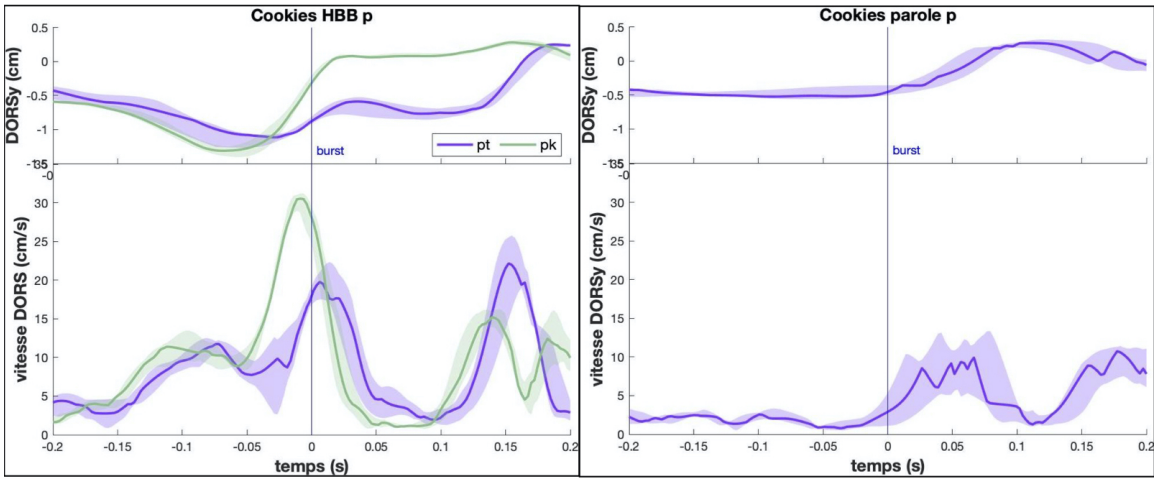


FIGURE 2 – Haut : médiane et écart interquartile de la bobine DORS au long de l'axe y, calculés à partir de toutes les occurences de la plosive bilabiale pour la phrase Cookies et sa contrepartie beatboxée. Bas : médiane et écart interquartile de la vitesse tangentielle (3D) de la bobine DORS.

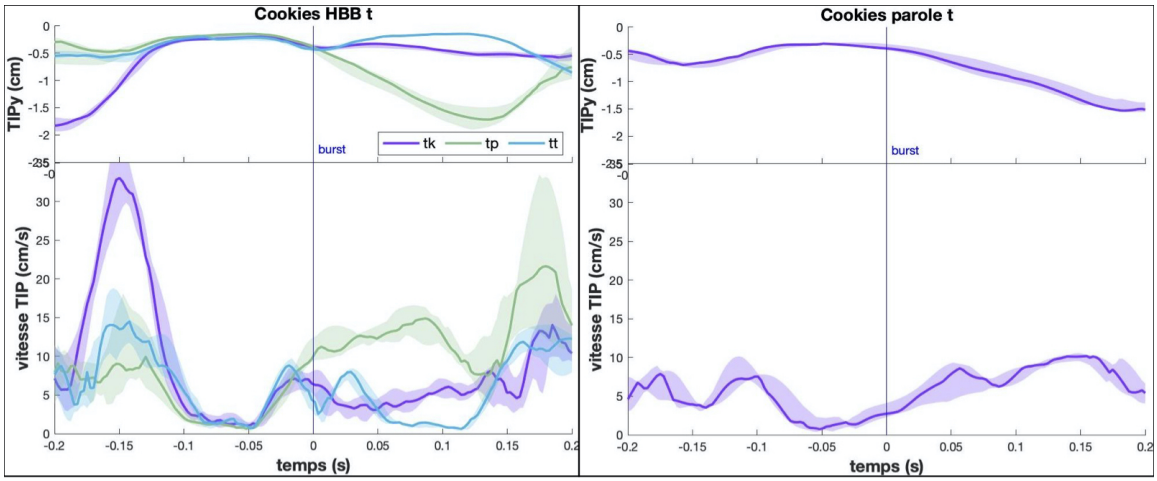


FIGURE 3 – Haut : médiane et écart interquartile de la bobine TIP au long de l'axe y, calculés à partir de toutes les occurences de la plosive alvéolaire pour la phrase Cookies et sa contrepartie beatboxée. Bas : médiane et écart interquartile de la vitesse tangentielle (3D) de la bobine TIP.

La Figure 3 illustre la dynamique articulatoire de la pointe de la langue pour la production de la plosive /t/ et de son équivalent en HBB, le hi–hat, dans le cas de la même phrase "*des p'**t**its cookies* ..." et de la série "*p **t** k **t** **t** p k **t***". Un plateau bien marqué est observé en HBB en correspondance duquel la vitesse de la bobine est presque nulle. Ceci suggère que l'occlusion du CV a bien lieu dans la région alvéolaire ou post–alvéolaire et qu'elle est maintenue un court instant. Ce plateau ne se retrouve pas en parole. Dans les deux modes de production, le relâchement de l'occlusion tel que détecté par la bobine TIP précède le burst détecté sur le signal audio. Dans les deux cas, la variabilité de la trajectoire est réduite. En particulier, l'écart interquartile est faible avant le burst. Bien que moins marquée que dans le cas du kick, la vitesse de déplacement de la bobine d'intérêt au moment du relâchement est plus grande en HBB qu'en parole.

Concernant la plosive /k/ et son équivalent rimshot en HBB, nous présentons le cas de la phrase BootsAndCats dans sa version parlée "*boots and **c**ats*", ou beatboxée comme suit : $p^{(n)}$ *ts* ***k*** *ts*.

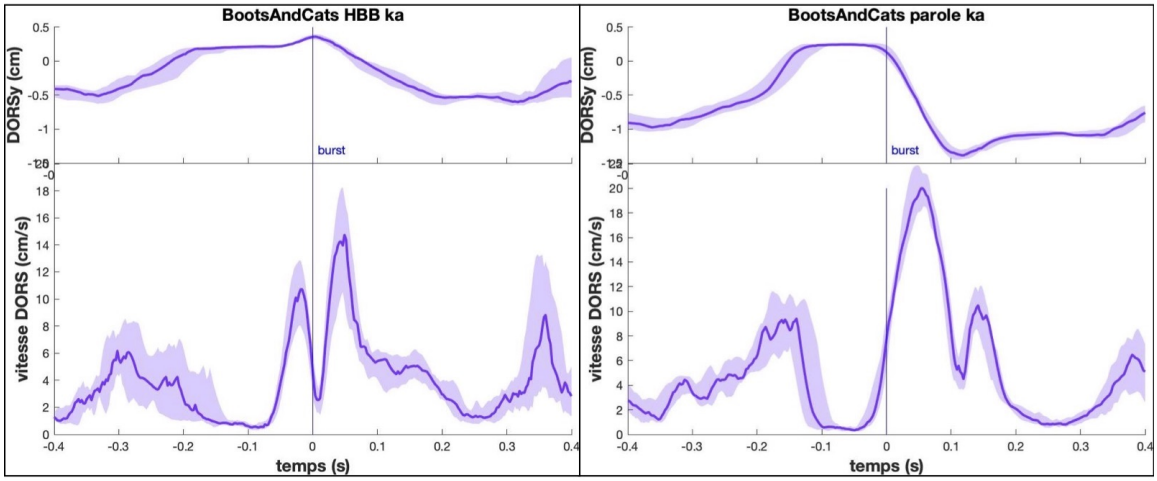


FIGURE 4 – Haut : médiane et écart interquartile de la bobine DORS au long de l'axe y, calculés à partir de toutes les occurences de la plosive vélaire pour la phrase BootsAndCats et sa contrepartie beatboxée. Bas : médiane et écart interquartile de la vitesse tangentielle (3D) de la bobine DORS.

En parole comme en HBB (Fig. 4), les trajectoires du dos de la langue (bobine DORS) montrent un plateau avant le burst, indiquant l'occlusion du CV au niveau de la région palato–vélaire. En parole, ce plateau se maintient jusqu'au burst, où la bobine montre un mouvement descendant vers la position de la voyelle suivante ([a]). En HBB, un mouvement ultérieur de la bobine DORS est observé avant le burst : vers la fin de l'occlusion, la bobine subit un déplacement vers le haut qui se termine en correspondance du burst. Ce mouvement n'est pas aléatoire, comme l'indique l'écart interquartile très faible. Il induit un pic de vitesse avant le burst (environ 10 cm/s), qui n'est pas observé en parole. En effet en parole, le seul pic de vitesse aux alentours du burst est celui lié au déplacement de la langue de l'occlusion vélaire vers la position beaucoup plus basse de la voyelle [a]. En HBB, deux pics de vitesse sont observés, l'un juste avant et l'autre juste après le burst, ce qui fait qu'au moment du burst la bobine DORS est en train de décélérer.

# 4 Discussion et conclusion

Nos résultats montrent que la production des sons percussifs du HBB se distingue clairement de celle des sons plosifs de la parole, même s'il est possible d'entrevoir une racine comune notamment au niveau du point d'articulation et de la trajectoire générale.

Les détails de l'articulation des sons percussifs du HBB revèlent une dynamique articulatoire propre au HBB et différente de celle de la parole. Dans le cas du kick comparé à une plosive sourde bilabiale, l'occlusion est latéralisée. Cette latéralisation du relâchement de l'occlusion bilabiale s'observe chez certains beatboxeurs, mais pas de façon systématique. Le type de latéralité (gauche ou droite) dépend du beatboxeur. Ce geste labial a probablement pour but de mieux contrôler la tension des lèvres au moment du relâchement. La langue est également très active même si l'occlusion s'effectue au niveau des lèvres. Les vitesses de déplacement des articulateurs sont souvent plus grandes en HBB qu'en parole, surtout avant le burst. La dynamique linguale des sons percussifs du HBB semble être cohérente avec l'utilisation d'un mécanisme éjectif. La dynamique linguale très active et reproductible durant l'articulation des sons percussifs bilabiaux du HBB ne semble pas pouvoir être complètement expliquée par la coarticulation, puisque à égalité d'environnement phonétique en parole (notamment /pt/), nous n'observons pas de mouvements comparables. En outre, toutes les trois bobines de la langue sont impliquées quasiment de la même manière, indiquant un déplacement global de la langue. Cela semble plutôt suggérer un mouvement de la langue en relation avec la remontée du larynx propre au mécanisme éjectif. En effet, plusieurs études ont attesté l'utilisation de ce mécanisme dans l'articulation des sons percussifs du HBB (Proctor *et al.*, 2013; De Torcy *et al.*, 2014; Sapthavee *et al.*, 2014; Blaylock *et al.*, 2017; Patil *et al.*, 2017; Dehais Underdown *et al.*, 2019). Il en va de même pour les mouvements de remontée de la langue à la fin de la phase d'occlusion des sons percussifs vélaires du HBB. Quant à la nature de ces mouvements linguaux, Proctor et collègues (Proctor *et al.*, 2013) font l'hypothèse que la langue soit utilisée avec le larynx pour produire une action de poussée plus efficace. En effet, un mécanisme éjectif optimalisé pourrait permettre d'accroître l'efficacité sonore du son produit, ce qui est très important en HBB. En revanche, nos données n'ont pas mis en évidence ce type de mouvements lors de l'articulation des plosives alvéolaires d'aucune des trois phrases beatboxées. Cela pourrait s'expliquer par le fait qu'en ces cas la remontée du larynx n'a pas d'influence sur la langue de part sa position articulatoire (apex plus elevé que le dos). Il se peut aussi que notre beatboxeur n'utilise pas de mécanisme éjectif pour la réalisation des alvéolaires. De manière générale et à notre connaissance, peu d'études ont exploré l'impact du mécanisme éjectif sur la dynamique linguale, en parole comme en HBB. Grâce à une dynamique articulatoire très active et à la large utilisation de ce mécanisme, l'étude de la production des sons percussifs du HBB permet de mettre en évidence des phénomènes articulatoires au niveau lingual.

# Références


BLAYLOCK R., PATIL N., GREER T. & NARAYANAN S. S. (2017). Sounds of the human vocal tract. In *INTERSPEECH*, p. 2287–2291.

BOERSMA P. (2006). Praat : doing phonetics by computer. *http://www.praat.org/*.

CNET-ENST C. (1989). La parole et son traitement automatique.

DE TORCY T., CLOUET A., PILLOT-LOISEAU C., VAISSIERE J., BRASNU D. & CREVIER-BUCHMAN L. (2014). A video–fiberscopic study of laryngopharyngeal behaviour in the human beatbox. *Logopedics Phoniatrics Vocology*, **39**(1), 38–48.

DEHAIS UNDERDOWN A., BUCHMAN L. & 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 (ICPhS)*, Melbourne, Australia. HAL : hal-02284132.

FORREST K., WEISMER G., MILENKOVIC P. & DOUGALL R. N. (1988). Statistical analysis of word-initial voiceless obstruents : preliminary data. *The Journal of the Acoustical Society of America*, **84**(1), 115–123.

HEYNE M. & DERRICK D. (2014). Some initial findings regarding first language influence on playing brass instruments. In *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, p. 180–183.

HEYNE M. & DERRICK D. (2015). The influence of tongue position on trombone sound : A likely area of language influence. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)* : University of Canterbury. New Zealand Institute of Language, Brain & Behaviour.

HEYNE M., DERRICK D. & AL-TAMIMI J. (2019). Native language influence on brass instrument performance : An application of generalized additive mixed models (gamms) to midsagittal ultrasound images of the tongue. *Frontiers in Psychology*, **10**, 2597.

LAMKIN L. L. (2005). An examination of correlations between flutists' linguistic practices and their sound production on the flute. In *Proceedings of the Conference on Interdisciplinary Musicology*.

MADDIESON I. & DISNER S. F. (1984). *Patterns of sounds*. Cambridge university press.

MATLAB, 2018 (2018). Mathworks : Bioinformatics toolbox : User's guide (r2018b).

PATIL N., GREER T., BLAYLOCK R. & NARAYANAN S. S. (2017). Comparison of basic beatboxing articulations between expert and novice artists using real-time magnetic resonance imaging. In *Interspeech*, p. 2277–2281.

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.

SAPTHAVEE A., YI P. & SIMS H. S. (2014). Functional endoscopic analysis of beatbox performers. *Journal of Voice*, **28**(3), 328–331.