Etude comparative acoustique et articulatoire de la plosion entre parole et beatbox

Annalisa Paroni, Nathalie Henrich Bernardoni, Hélène Loevenbruck, Silvain Gerber, Pierre Baraduc, Christophe Savariaux

Etude comparative acoustique et articulatoire de la plosion entre parole et beatbox

Annalisa Paroni[1] Nathalie Henrich Bernardoni[1] Hélène Lœvenbruck[2]
Silvain Gerber[1] Pierre Baraduc[1] Christophe Savariaux[1]
(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
nathalie.henrich@gipsa-lab.fr

RÉSUMÉ

Au cœur de la pratique du Human Beatbox, les trois sons de percussion vocale *kick*, *hi-hat* et *rimshot* sont souvent introduits par le dévoisement des syllabes /pu/, /ti/ et /ka/ dans les débuts de l'apprentissage. Nous nous intéressons ici à comparer les caractéristiques acoustiques et articulatoires de la plosion entre ces consonnes parlées et leurs homologues beatboxés. Des séquences de répétition de ces sons produits par quatre beatboxeurs expérimentés ont été analysées. Nous montrons que la durée de la consonne est plus courte pour les consonnes parlées que pour celles beatboxées. La plosion est plus intense en beatbox qu'en parole. Les signatures temporelles et spectrales diffèrent nettement. Si le lieu d'articulation est similaire, la dynamique articulatoire les distingue. Ces résultats pointent du doigt le fait que, même si les beatboxeurs parlent et partent de sons de la parole, ils aboutissent à des mécanismes de production sonore bien différents en beatbox.

ABSTRACT

Comparative acoustical and articulatory study of the plosion between speech and beatbox

At the heart of Human Beatbox practice, the three vocal drum sounds *kick*, *hi-hat* and *rimshot* are often introduced by unvoicing the syllables /pu/, /ti/ and /ka/ in the early stages of learning. This study aims to compare the acoustical and articulatory characteristics of the plosion between speech consonants and their beatboxed counterparts. Repetition sequences of these sounds produced by four experienced beatboxers were analysed. We show that the consonant duration is shorter for spoken consonants than for beatboxed ones. The plosion is more intense in beatbox than in speech. The temporal and spectral signatures differ markedly. While the place of articulation is similar, the articulatory dynamics distinguishes them. These results point to the fact that, although beatboxers speak and start from speech sounds, they end up with quite different sound production mechanisms in beatboxing.

MOTS-CLÉS : beatbox, consonnes, articulation, acoustique.

KEYWORDS: beatboxing, consonants, articulation, acoustics.

1 Introduction

La percussion vocale est pratiquée dans de nombreuses cultures vocales à travers le monde (Patel & Iversen, 2003). Nous nous intéressons ici à la pratique de la percussion vocale dans le Human Beatbox, un art vocal qui s'est fortement répandu dans les pays occidentaux et dont la pratique est structurée par

des championnats nationaux et internationaux (des battles). Le Human Beatbox s'apprend en atelier ou de façon auto-didacte à partir de tutoriaux proposés sur le web. Pour introduire les premiers sons de base du beatbox que sont l'imitation de la grosse caisse (le *kick*), du charleston (le *hi-hat*), et du son typique de la technique du rimshot, des images vocales sont proposées, qui s'appuient bien souvent sur des syllabes parlées que l'on dévoise. Ainsi le kick peut s'introduire à partir de la syllabe [pu], le hi-hat de la syllabe [ti] et le rimshot de la syllabe [ka]. La pratique du beatbox s'inscrit ainsi pleinement entre parole et musique. Nous nous sommes donc intéressés à comparer ces sons consonantiques que l'on pourrait qualifier de « beatboxés » à leurs homologues parlés. Plusieurs études (Blaylock *et al.*, 2017; Patil *et al.*, 2017; Proctor *et al.*, 2013; Dehais Underdown *et al.*, 2019; Sapthavee *et al.*, 2014; De Torcy *et al.*, 2014) ont montré que souvent ces sons sont glottaux egressifs. Une étude (Dehais Underdown *et al.*, 2019) a fourni des détails sur la physiologie du *kick* d'un beatboxeur, avec des données acoustiques et aérodynamiques. Une exploration conjointe des caractéristiques acoustiques, de la dynamique articulatoire linguale et labiale et des comportements ventilatoires d'un beatboxeur amateur produisant les principaux sons de batterie de son répertoire - et en particulier le kick, le hi-hat et le rimshot - a permis d'introduire la notion de boxème et mis en évidence des signatures acoustiques spécifiques à chaque boxème, en lien avec des gestes articulatoires complexes et non observés en parole (Paroni *et al.*, 2021). Les études comparatives entre beatbox et parole sont rares dans la littérature. Une comparaison entre des phrases parlées ou beatboxées pour ce même beatboxeur a montré que la dynamique articulatoire des consonnes beatboxées se distingue nettement de celle des consonnes parlées (Paroni *et al.*, 2020a). Nous nous proposons d'étendre ici cette étude comparative entre parole et beatbox au cas de plusieurs beatboxeurs expérimentés et dans un contexte plus contrôlé de répétition syllabique (ou *boxique*). Le corpus constitué pour cette étude et les méthodes d'analyse seront présentées en partie 2. Nous comparerons les signatures acoustiques et spectrales des consonnes parlées à celles de leurs homologues beatboxées en partie 3.1. Nous nous intéresserons à la durée des consonnes parlées et beatboxée en partie 3.2 et détaillerons les différences de comportement articulatoire en partie 3.3.

2 Matériel et méthodes

2.1 Sujets

Quatre beatboxeurs masculins expérimentés ont participé aux expériences. Leurs caractéristiques sont précisées dans le Tableau 1. Ils ont une pratique régulière du beatbox, journalière pour S01, S03 et S04. Tous ont appris le beatbox en auto-didacte, avec parfois le soutien de vidéos YouTube (S01, S04). Deux beatboxeurs (S02 et S03) sont des professionnels du beatbox. Deux beatboxeurs (S02 et S04) ont remporté des titres de champion lors des championnats nationaux et internationaux.

2.2 Tâches

Pour chaque consonne parlée ou beatboxée, les sujets ont du répéter des séquences de sons précédées de la phrase [sasɛlə]. Les trois syllabes parlées /pu/, /ti/ et /ka/ ont été répétées par deux séquences de six répétitions chacune. Les trois boxèmes kick (**P**), hi-hat (**t**) et rimshot (**K**) ont été répétés 12 fois à la suite.

Sujet	Age	Latéralité	Années de pratique	Niveau pro	Niveau compet
S01	21	D	5	amateur	1
S02	38	D		pro	3
S03	31	G	13	pro	2
S04	20	D	7	amateur	3

TABLE 1 – Informations sur les sujets. Le sujet est considéré comme pro s'il gagne sa vie de sa pratique. Le niveau de compétition est quoté sur la participation à des compétitions officielles : (1) pas ou peu (1 ou 2) ; (2) oui (>2), mais sans titre ; (3) oui, avec titre(s).

2.3 Constitution du corpus

Les données ont été acquises en laboratoire dans une salle semi–anéchoique (plateforme BEDEI, GIPSA–lab, Grenoble). Le protocole expérimental a été validé par le Comité d'Ethique pour les Recherches Grenoble Alpes.

Les signaux acoustiques ont été enregistrés à partir d'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). Ce microphone de mesure a été positionné à 30 cm de la bouche du beatboxeur.

Les gestes articulatoires ont été enregistrés à partir d'un articulographe électromagnétique 3D (système WAVE, société NDI). Neuf bobines ont été collées sur des points de chair d'intérêt de la langue et des lèvres (voir Figure 1) : — trois bobines sur la langue dans le plan médio–sagittal, une dans la région apicale (TA), une dans la région dorso–palatale (TM) et une dans la région dorso–vélaire (TB) ; — deux bobines sur la langue dans la région dorso–palatale à droite et gauche du plan médio-sagittal (TR et TL) — quatre bobines sur les lèvres supérieure et inférieure, de façon médiane (LMH, LML) et latérale (LLH, LLL).

Une bobine a été rattachée à l'incisive inférieure pour suivre les mouvements de la mâchoire (JAW). Enfin, une bobine de référence a été positionnée au niveau du nasion. Les signaux EMA ont été enregistrés à 200 Hz pour les sujets S01, S02 et S04 et à 100 Hz pour le sujet S03.

L'expérience a été contrôlée par un caméscope (SONY, HDR-XR500E) placé face au sujet.

2.4 Analyse des données

Les données audio et électroglottographiques ont été segmentées manuellement et annotées sous Praat (Boersma, 2006). En particulier, chaque consonne a été annotée en plaçant la borne de début de la consonne à l'instant du burst et la borne de fin soit (i) à la reprise du voisement (consonne sourde parlée dans un contexte syllabique CV), soit (ii) à l'extinction du son (consonne beatboxée). Les annotations des TextGrid ont été importées sous MATLAB, afin de déterminer les fenêtres d'analyse du signal acoustique.
Les signaux EMA ont été importés et traités sous MATLAB. En particulier, les trajectoires moyennes et leurs covariances ont été calculées sur les différentes répétitions consonantiques.

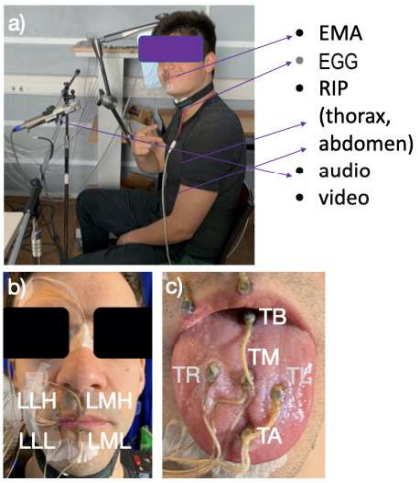

FIGURE 1 – Présentation du cadre expérimental et des dispositifs de mesure : audio, articulographe életromagnétique (EMA), électroglottographe (EGG), pléthysmographie à variation d'inductance (RIP).

2.5 Modélisation statistique

Nous avons regardé l'impact d'une variable explicative *consonne* à 6 modalités (p, t, k, kick, hi-hat, rimshot) sur la variation d'intensité ou de durée. Un modèle linéaire mixte a été élaboré en utilisant la fonction *lme* du package *nlme* du logiciel de statistique R. Un tel modèle mixte permet à la fois de tenir compte de la répétition des mesures, de la variabilité inter-beatboxeurs et de la variance résiduelle qui peuvent changer d'une consonne à l'autre. Pour les variables de durée (durée absolue et durée relative de la consonne) dont la distribution est asymétrique, nous avons utilisé leur logarithme. Une analyse de contrastes a été menée sur les modèles établis avec la fonction *glht* du package *multcomp*, selon la méthode présentée par Hothorn *et al.* (2008). L'objectif des comparaisons multiples est d'explorer s'il existe une différence entre une consonne parlée et son équivalente beatboxée.

3 Résultats

3.1 Signatures acoustiques et spectrales

La Figure 2 présente les signatures acoustiques et spectrales des consonnes parlées et de leurs homologues beatboxées, extraites de la troisième répétition de chaque son.

L'onde acoustique est systématiquement d'amplitude plus élevée pour les sons consonantiques en beatbox qu'en parole. Nous avons quantifié cette différence en calculant l'intensité de la consonne. La Figure 3 présente l'intensité des sons consonantiques parlés et beatboxés. La différence d'intensité est fortement significative pour chaque paire de sons : de manière globale et en moyenne, [p] est moins intense que **P** de 20.9$\pm$1.6 dB ($p<0.001$), [t] est moins intense que **t** de 12.7$\pm$1.4 dB ($p<0.001$), [k]

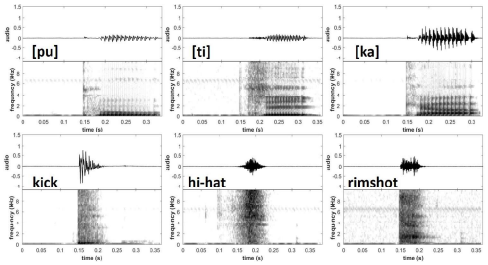

FIGURE 2 – Formes temporelles et spectrales du signal audio pour la troisième répétition de chaque consonne parlée ou beatboxée par le sujet S03. Paramètres du spectrogramme : gamme fréquentielle 0–10 kHz ; fenêtre d'analyse FFT : 9 ms ; gamme dynamique : 30 dB.

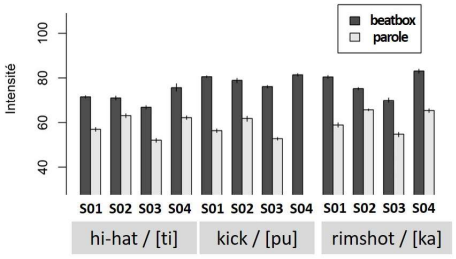

FIGURE 3 – Intensité en dB des consonnes parlées et beatboxées pour les quatre sujets.

est moins intense que **R** de 16.0±1.4 dB (p<0.001).

3.2 Durée absolue et relative des consonnes

La Figure 4 présente les mesures de durée moyennes et leurs écart-types pour les quatre sujets et les trois comparaisons.

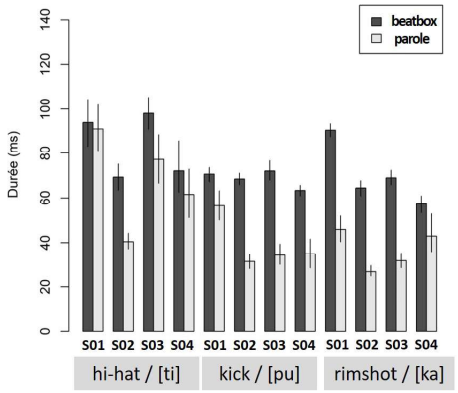

FIGURE 4 – Durée moyenne des consonnes parlées et beatboxées en ms pour les quatre sujets.

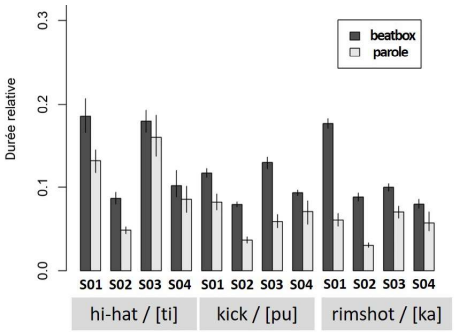

FIGURE 5 – Durée relative au tempo des consonnes parlées et beatboxées pour les quatre sujets.

La différence est significative pour toutes les paires de sons (valeurs estimées données sur l'échelle logarithmique) : bilabiaux (-0.6134±0.096, p<0.001), apico-alvéolaires (-0.2522±0.0978, p<0.05) et vélaires (-0.6782±0.0936, p<0.001).

Comme le rythme de la répétition consonantique était laissé à l'appréciation du beatboxeur pendant la séquence, nous avons également rapporté la durée du son consonantique à la durée d'un cycle de ré-

pétition, prenant ainsi en compte le tempo naturel de la séquence de répétition sonore. Les différences sont encore plus accentuées pour ces rapports de durée (voir Figure 5). De même que précédemment, la différence est significative pour toutes les paires de sons : bilabiaux (-0.5758±0.1259, p<0.05), apico-alvéolaires (-0.3182±0.1274, p<0.001) et vélaires (-0.7283±0.1243, p<0.001).

3.3 Comportement articulatoire

L'analyse des données articulatoires a confirmé que, pour tous les beatboxeurs, le lieu d'articulation est similaire entre sons beatboxés et sons parlés : l'occlusion orale a lieu au niveau des lèvres dans le cas de **P** et [p] ; dans la région de TA pour **t** et [t] ; dans la région de TB pour **K** et [k]. La Figure 7 exemplifie le cas de S02. La technique d'EMA ne permet néanmoins pas de donner le point précis de l'occlusion. Il ne sera donc pas nécessairement le même entre beatbox et parole. Par exemple de façon générale dans le cas de **K** la bobine TB entre en contact avec le palais plus en avant que pour [k] (voir Fig. 7). Souvent les sons beatboxés sont produits grâce à des mouvements plus amples et rapides en beatbox, nonobstant l'absence de sons vocaliques.

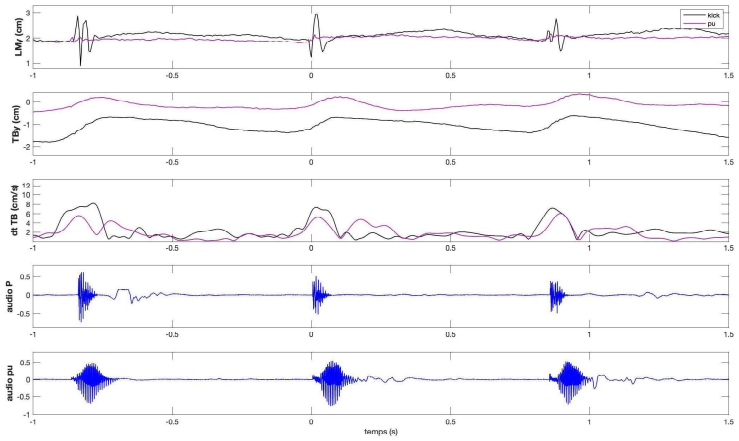

FIGURE 6 – Observation comparative des signaux temporels articulatoires pour les deux sons consonantiques kick et [pu], respectivement beatboxé et parlé par le sujet S02.

Le cas de la consonne beatboxée **P** est particulièrement intéressant. Durant la phase de production acoustique, la distance interlabiale (visualisée en Figure 6 par le tracé LM, distance entre les bobines centrales des lèvres superieure et inférieure) varie rapidement et grandement après le relâchement de l'occlusion, tandis que dans le cas de [p] les deux lèvres ne s'écartent jamais considérablement, même au moment du relâchement. Ce mouvement bien différent des lèvres est mis en évidence sur la Figure 7 pour la bobine médiane de la lèvre inférieure.

Comme pour les lèvres, la langue présente des mouvements rapides, notamment au niveau de TB, chez tous les beatboxeurs, pouvant atteindre les 20 cm/s, alors que les vitesses mesurées pour [pu] sont de l'ordre de 5–6 cm/s. Ces mouvements de remontée commencent avant le burst et se poursuivent jusqu'à la fin, voir après la fin du son. Cependant, la langue adopte une position généralement plus basse à l'intérieur de la cavité orale par rapport à [p]. En superposant les trajectoires de toutes les

occurrences de la séquence, les bobines de la langue dessinent des boucles liées à ces mouvements de remontée. De telles boucles sont plus marquées chez certains beatboxeurs (S01, S03) que chez d'autres (S02, S04). Durant l'articulation de [pu], ces boucles peuvent apparaître, mais sont souvent moins prononcées et moins stéréotypées. Le **K** a montré les stratégies articulatoires plus variées et s'éloignant le plus de celle de la parole. Chez S02 et S04, l'occlusion se fait dans la région de TB, mais bien plus en arrière pour S04 par rapport à S02, et la signature acoustique du son est très différente. Chez S02 et S04, TA est en contact avec le palais et il le reste au moment du relâchement de l'occlusion dans la région de TB. Dans ce cas aussi, la signature acoustique des deux sons est bien différente. Indépendamment des différentes stratégies articulatoires, les vitesses associées à **K** sont systématiquement moindres par rapport à celles de [ka]. Des boucles articulatoires sont bien visibles chez tous les beatboxeurs pour l'articulation de [ka]. Elles peuvent apparaître dans le cas de **K** (S01-S03), mais sont généralement moins amples (S03) et se orientent selon un axe plus vertical qu'en parole (Fig. 7). Un comportement articulatoire plus similaire entre beatbox et parole se trouve dans le cas de **t**. Toutefois, en beatbox, la langue présente une position globalement plus basse et la bobine TB peut remonter légèrement après le burst. Quant aux vitesses articulatoires, deux cas de figure se présentent. Dans la plupart des cas (S01, S03, S04), le relâchement au niveau de TA est plus rapide en beatbox qu'en parole. Néanmoins chez S02, le relâchement est plus rapide en parole.

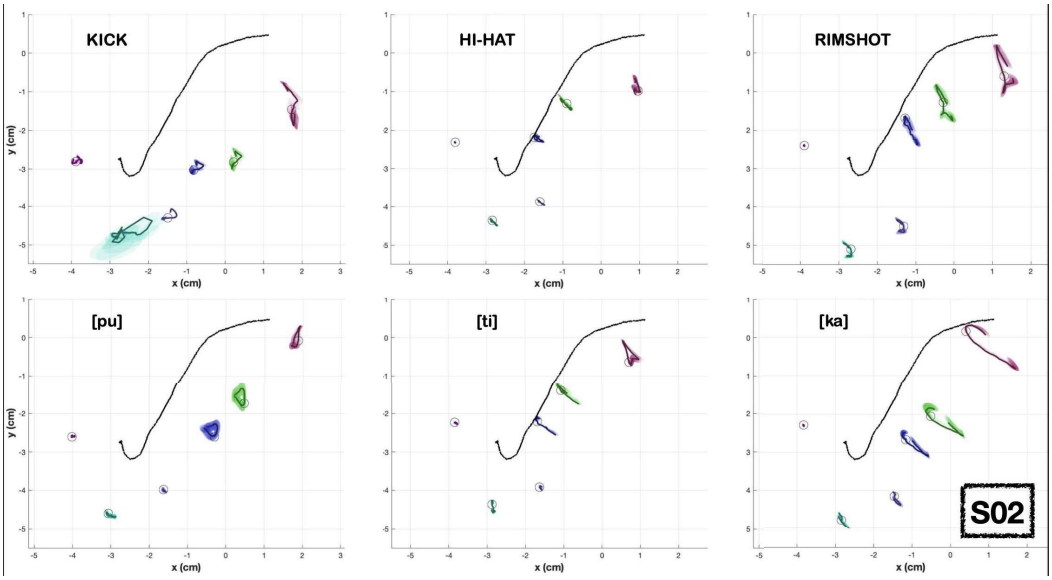

FIGURE 7 – Illustration des trajectoires articulatoires moyennes mises en jeu dans la production des sons consonantiques beatboxés (en haut) et parlés (en bas) pour le sujet S02. Les positions à l'instant du burst sont représentées par les cercles. La fenêtre temporelle visualisée est de 300 ms avant et après le burst.

4 Discussion et conclusion

Les résultats obtenus confirment les observations faites précédemment dans le cas d'un unique beatboxeur (Paroni *et al.*, 2020a). Les sons consonantiques beatboxés se distinguent nettement de leurs homologues parlés, tout à la fois du point de vue acoustique (signature temporelle et spectrale, durée et intensité du son) et articulatoire (trajectoire et vitesse articulatoire). Si le lieu d'articulation est le même pour les trois sons beatboxés comparativement à ceux parlés, des différences subsistent quant au point précis d'articulation, qui peut être légèrement décalé.

Les stratégies articulatoires différencient les consonnes beatboxées des consonnes parlées. Nous retrouvons des boucles articulatoires différentes, probablement en lien avec une stratégie d'activation différente des muscles de la langue. Nous retrouvons aussi les mouvements de remontée de la langue, notamment dans le cas du **P**, qui ne peuvent pas être liés à la coarticulation avec un son vocalique, comme dans le cas de la parole où la consonne [p] est coarticulée avec la voyelle [u]. Quant à la nature de ce mouvement de remontée, nous avons suggéré qu'il puisse être lié au mouvement de remontée du larynx typique d'une production éjective (Paroni *et al.*, 2020b).

A des comportements articulatoires différents correspondent des signatures acoustiques qui distinguent clairement les sons beatboxés de leurs homologues parlés. De manière générale, les sons beatboxés sont significativement plus intenses que leurs homologues parlés, ce qui est en accord avec des productions éjectives, notoirement plus intenses que leurs équivalents pulmoniques (Ladefoged & Maddieson, 1996). Les sons beatboxés sont aussi plus longs que les consonnes parlées. La coarticulation entre consonne et voyelle dans les séquences parlées contribue très certainement à cette différence. Dans les séquences beatboxées, les sons consonantiques sont coarticulés entre eux, sans passer par une configuration vocalique. Ils peuvent donc se développer plus librement. Il serait intéressant de vérifier si cette différence de durée subsiste dans une tâche moins naturelle, mais phonétiquement plus équilibrée de répétition de la seule consonne parlée, sans coarticulation avec la voyelle. Cette tâche permettrait aussi de mieux comparer les vitesses articulatoires : en correspondance du lieu d'articulation, elles sont moindres en parole qu'en beatbox dans un contexte phonétique où la langue ne doit pas beaucoup se déplacer pour atteindre la position vocalique ([pu], [ti]), mais deviennent plus importantes quand la langue doit se déplacer davantage ([ka]). Cette différence de vitesse articulatoire serait-elle donc intrinsèquement liée aux mécanismes de production du beatbox ou bien au contexte phonétique ?

Remerciements

Ce travail a été soutenu par l'Agence Nationale de la Recherche dans le cadre du programme d'Investissements d'avenir (ANR-15-IDEX-02). Il n'aurait pas pu exister sans l'investissement des quatre beatboxeurs. Un grand merci à eux pour leur forte implication dans ce projet et leur patience lors des mesures.

Références

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

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

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.

HOTHORN T., BRETZ F. & WESTFALL P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal : Journal of Mathematical Methods in Biosciences*, **50**(3), 346–363.

LADEFOGED P. & MADDIESON I. (1996). *The sounds of the world's languages*. Blackwell Oxford.

PARONI A., BERNARDONI N. H., SAVARIAUX C., BARADUC P. & LŒVENBRUCK H. (2020a). Beatboxer, est-ce parler ? ce que nous en dit l'étude de la dynamique articulatoire d'un beatboxer. In *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). Volume 1 : Journées d'Études sur la Parole*, p. 472–479 : ATALA ; AFCP.

PARONI A., HENRICH BERNARDONI N., SAVARIAUX C., BARADUC P., CALABRESE P. & LOEVENBRUCK H. (2020b). Beatboxing, is it talking ? In *Proceedings of the 12th International Seminar on Speech Production*, Providence (virtual), United States : Haskins Laboratories.

PARONI A., HENRICH BERNARDONI N., SAVARIAUX C., LŒVENBRUCK H., CALABRESE P., PELLEGRINI T., MOUYSSET S. & GERBER S. (2021). Vocal drum sounds in human beatboxing : An acoustic and articulatory exploration using electromagnetic articulography. *The Journal of the Acoustical Society of America*, **149**(1), 191–206.

PATEL A. & IVERSEN J. (2003). Acoustical and perceptual comparison of speech and drum sounds in the North India tabla tradition : An empirical study of sound symbolism. *Proceedings of the 15th International Congress of Phonetic Sciences*.

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 *Proceedings of 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.