

Rythme et contrôle articulatoire : étude préliminaire du Human Beatbox.

Alexis Dehais Underdown¹ Paul Vignes¹, Lise Crevier-Buchman^{1,2} & Didier Demolin¹
(1) Laboratoire de Phonétique et Phonologie, (CNRS / Sorbonne Nouvelle), 19 Rue des Bernardins, 75005 Paris, France
(2) Unité d'exploration fonctionnelle Voix-Parole – Service d'ORL & Chirurgie Cervico-Faciale, 40 rue Worth, BP 36, 92151 Suresnes, France
alexis.dehais-underdown@sorbonne-nouvelle.fr, vignes.paul@gmail.com,
lise.buchman1@gmail.com, didier.demolin@sorbonne-nouvelle.fr

RÉSUMÉ

Dans cette étude nous nous intéressons à l'analyse spectrale d'imitation de grosses caisses, de charleston et de caisse claire dans un paradigme d'augmentation de la vitesse de production. La vitesse de production a été contrôlée en utilisant un métronome à vibration paramétré à 90, 120 puis 150 battements par minute. Le Centre de Gravité spectral et le coefficient d'asymétrie ont été mesurés pour inférer la stabilité et la variabilité articulatoire des sons produits dans les patterns beatboxés. Les grosses caisses sont les sons les plus contrôlés suivi par les caisses claires puis les charlestons.

ABSTRACT

Rhythm and articulatory Control : Preliminary study of Human Beatboxing.

In this study we were interested in the spectral analysis of kick drums, hi-hat and snare drum imitations in a speed increasing production task. The production speed was controlled using a vibration metronome set at 90, 120 and 150 beat per minute. Center of Gravity and Skewness were measured in order to infer articulatory stability and variability of the sounds in the beatboxed patterns. Kick drum were the most controlled sounds followed by snare drums and hi-hats.

MOTS-CLÉS : HUMAN BEATBOX, VITESSE DE PRODUCTION, ACOUSTIQUE, RYTHME.

KEYWORDS: HUMAN BEATBOXING, SPEED RATE, ACOUSTICS, RHYTHM.

1 Introduction

La parole est un accommodement entre des principes de production et de perception pour délivrer un message linguistique. L'utilisation du Conduit Vocal (CV) humain est soumise à des contraintes qui permettent de minimiser le coût de la production afin de maximiser la perception des sons produits. La parole est dépendante d'une organisation complexe de différents niveaux linguistiques (lexical, morphosyntaxique, phonologique ...). Le Human Beatbox (HBB) n'obéit pas à ses restrictions, on ne beatboxe pas pour véhiculer un message linguistique. Cela implique que la production n'est pas soumise à des contraintes liées aux langues. Cette différence entre parole et HBB a un impact fondamental sur la façon dont on utilise le conduit vocal lorsque l'on beatboxe. L'HBB n'est pas

restreint à l'ensemble des gestes articulatoires répondant au compromis linguistique entre production et perception. Cependant, il doit aussi exister une relation entre production et perception dans le HBB dont on ignore, à ce jour, la nature.

Dans cette étude nous nous intéressons à l'analyse spectrale d'imitation d'instruments percussifs dans un paradigme d'augmentation de la vitesse de production. Il existe à ce jour peu d'études sur le HBB. Concernant le signal acoustique des sons beatboxés la littérature est encore plus restreinte. On retiendra celles de Picart, Brognaux Dupont, 2015 ; Stowell & Plumbley, 2010 s'intéressant à la classification automatique des sons, celle de Stowell & Plumbley (2008) décrivant de façon impressionniste les sons et celles de Dehais Underdown, Crevier Buchman & Demolin (2019) et Dehais Underdown et al. (2019) qui ont analysé les caractéristiques spectrales de la grosse caisse [p'] et la caisse claire [pf']. Enfin, à partir du même corpus Dehais Underdown et al. (2020) ont montré que la durée des séquences beatboxées tend à diminuer lorsque la vitesse de production augmente ($r = 0.97$) ; de plus ils ont trouvé un faible taux d'erreurs de production de 7,6% ($n=1083$).

Les sons de notre corpus sont non-voisés, il faut donc s'interroger sur l'analyse acoustique de la production d'une suite de sons sourds. En l'absence de sons vocaliques il est difficile d'utiliser certains indices acoustiques comme le VOT (Cho & Ladefoged, 1999), les équations de locus (Sussman, McCaffrey & Matthews, 1991) ainsi que d'autres corrélats destinés à l'étude des consonnes glottiques (Kingston, 1985 ; Wright et al., 2002). Nous avons donc choisi de porter notre analyse sur les bruits des sons que nous avons analysé en s'intéressant aux moments spectraux 1 et 3, c'est-à-dire le Centre de Gravité spectral (CdG) et le coefficient d'asymétrie (Forrest et al. 1988). Ces paramètres permettent de distinguer les lieux d'articulation à la fois des occlusives et des fricatives sibilantes (Forrest et al. 1988 ; Jongman et al. 2000 ; Nitrouer, 1995 ; Lousada et al. 2012). Il existe quatre moments spectraux décrivant la répartition d'énergie dans le spectre. Le premier correspond au CdG, c'est une moyenne de la composition spectrale indiquant la fréquence où l'énergie est concentrée. Le deuxième moment décrit la variance de la distribution. Le troisième moment (i.e. coefficient d'asymétrie ou « skewness » en anglais) renseigne sur la symétrie ou l'asymétrie de la distribution. Enfin le quatrième moment (i.e. coefficient d'aplatissement ou kurtosis) décrit l'aplatissement de la distribution. Les fortes corrélations entre le CdG et les paramètres articulatoires impliqués dans la production des fricatives dans l'étude de Tabain (2001) nous permettent d'inférer que le CdG est un bon indicateur pour définir le lieu d'articulation. Nous prendrons le coefficient d'asymétrie comme paramètre indiquant la variabilité de composition spectrale et indirectement les variations de configuration du CV.

L'objectif de cette étude est d'analyser l'influence de la vitesse de production sur des sons beatboxés. Nous faisons l'hypothèse qu'en augmentant la vitesse de production cela entraînera des perturbations dans la production : (1) nous pensons que les sons coronaux montreront plus de variation de CdG (i.e. indicateur du lieu d'articulation) que les vélares et ensuite que les labiales ; (2) que les labiales montreront plus de variation du coefficient d'asymétrie (i.e. indicateur de la configuration du conduit vocal) ; enfin (3) la position du son dans une structure beatboxée est un facteur influençant la variation de ces deux paramètres spectraux.

2 Méthodologie

2.1 Corpus & protocole

Un beatboxeur professionnel de 31 ans a pris part à cette étude préliminaire. Les enregistrements se sont déroulés dans la chambre sourde du Laboratoire de Phonétique et Phonologie avec un microphone cardioïde AKG C520 avec une fréquence d'échantillonnage de 44kHz. Les séquences rythmiques (i.e. pattern de beatbox) à l'étude ont été extraites d'un corpus plus large. Le pattern de beatbox désigne une structure composée de sons beatboxés qui peut avoir un rythme binaire ou asymétrique et une organisation plus ou moins complexe en fonction du nombre de constituants de cette structure. La structure de pattern de beatbox créé pour cette étude est composée de 9 frappes (i.e. production du son) sur 4 battements. Chaque frappe a été numérotée de 1 à 9 sous forme de position dans le pattern de beatbox. Nous avons contrôlé la vitesse de production (i.e. Tempo) en utilisant un métronome à vibration paramétré à 90, 120 puis 150 battements par minute (BPM). Chaque pattern de beatbox (PB) se compose d'une même structure : Grosse Caisse, Charleston, Caisse Claire, Charleston, Grosse Caisse, Grosse Caisse, Charleston, Caisse Claire, Charleston. Nous avons choisi de transcrire les sons beatboxés à l'aide de l'API (Association Phonétique Internationale, 1999) et de ses extensions destinées à la parole pathologique (Ball, Howard & Miller, 2018).

1. [p' ts' ↓kʌ ts' p'p' ts' ↓kʌ ts'],
2. {η [↑⊙ ↑| * ↑| ⊙↑⊙ ↑| * ↑| η]} ({η ... η} indique que l'artiste peut fredonner en même temps qu'il beatboxe, [*] est un click où le relâchement est postérieur décrit par (Tuhuse & Traill, 1999)),
3. [t' ts' tʃi' ts' t' t' ts' tʃi' ts'].

Chaque pattern a été répété 8 fois de suite à trois vitesses différentes (3 PB x 8 répétitions x 3 vitesses = 72 PB). Les fichiers sons ont été ré-échantillonnés à 32kHz dans le but d'observer les plus hautes fréquences.

2.2 Analyse

Nous avons utilisé Praat (Boersma & Weenink, 2006) pour segmenter les données et extraire le CdG et le coefficient d'asymétrie. Sur le textgrid nous avons noté le son tel qu'il a été produit et la cible qui était attendue. Pour chaque son, nous avons extrait les spectres FFT sur des fenêtres de 25ms afin d'observer les changements spectraux au cours du temps pour chaque son. À partir des spectres extraits, nous avons relevé les valeurs de Centre de Gravité spectral (CdG) et le coefficient d'asymétrie (i.e. Skewness) dans le but d'analyser la composition spectrale de chaque son en fonction de la position et de la vitesse de production. La durée (ms) des PB a été calculée à partir du début du premier son jusqu'à la fin du dernier son. Étant donné que la durée des PB diminue lorsque la vitesse de production augmente (Dehais Underdown, 2020) nous utiliserons cette valeur pour évaluer l'influence de la vitesse de production sur nos mesures spectrales. Dans notre analyse nous avons enlevé toutes les erreurs de production (e.g. substitutions, omissions) dans les 3 patterns à l'étude, c'est-à-dire lorsqu'un son produit ne correspond pas à la cible attendue, pour ne pas biaiser notre analyse. Nous n'analyserons pas les erreurs de production (4,3%).

Le logiciel R (Team R core, 2005) a été utilisé dans le but de générer des nuages de points et des coefficients de corrélations de Pearson des mesures spectrales (variable dépendante) et de la durée des PB (variable indépendante).

3 Résultats

3.1 Grosse caisse

Concernant le CdG des grosses caisses [p'] [↑Θ], sur la figure 1 (gauche) les valeurs se situent en dessous de 500 Hz, pour [t'] les valeurs se trouvent en dessous de 2,5 kHz. Cela suggère qu'il y a une concentration de l'énergie qui prédomine dans les basses fréquences dans le but d'imiter au mieux le timbre grave d'une grosse caisse. Nous avons d'un côté les sons [p'] et [↑Θ] qui, dans toutes les positions (i.e. position 1, 5 et 6), montrent peu de dispersion des valeurs de CdG alors que la vitesse augmente. Il n'y a pas de corrélation entre la vitesse et le CdG. De plus les positions 1, 5 et 6 ne semblent pas influencer la production des grosses caisses labiales. La position 6 est celle qui montre le moins de variation et suggère donc que la production est stable. De l'autre côté, [t'] montre plus de dispersion des valeurs de CdG suggérant des variations subtiles de lieu d'articulation dans la zone dentale-alvéolaire.

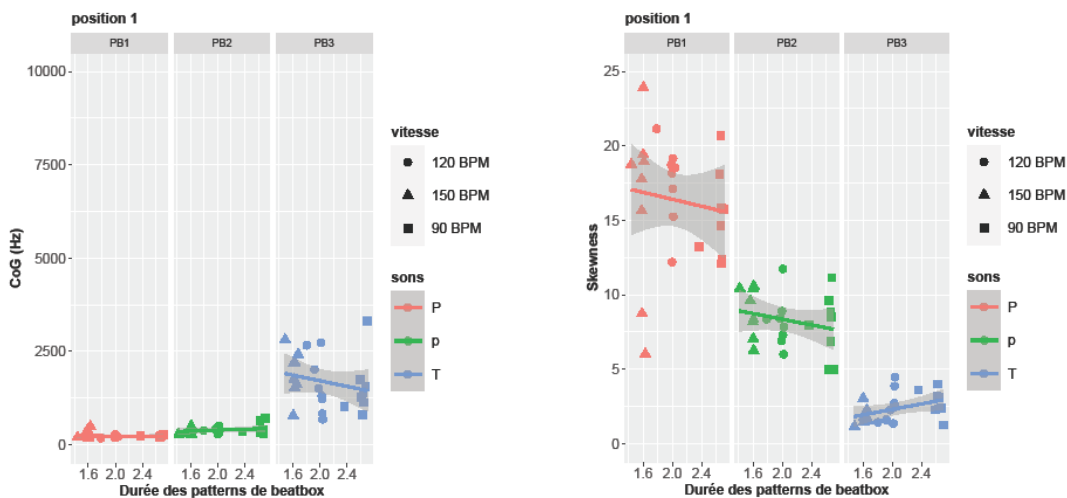


FIGURE 1 : Nuages de points du CdG (gauche) et coefficient d'asymétrie (droite) en fonction de la durée des PB de différentes imitations de grosses caisses en position 1. P = [p'], p = [↑Θ], T = [t']. Les zones grisées indiquent l'intervalle de confiance.

L'analyse du coefficient d'asymétrie nuance quelque peu notre analyse. [p'] montre plus de dispersion des valeurs, ce qui suggère des variations de composition spectrale dues à des différences de configuration du conduit vocal lors de la production de cette labiale. Ces variations de composition ne sont pas surprenantes pour une labiale car les lèvres sont l'articulateur actif, laissant les articulateurs buccaux libres de leurs mouvements. Il est possible que ces variations soient dues à une anticipation des gestes suivants ou bien à la fin du geste précédent. La grosse caisse [↑Θ] suit cette

tendance dans une moindre mesure car on observe moins de dispersion des valeurs du coefficient. Cela s'explique par le fait que le mécanisme d'initiation vélique de ce pattern contraint plus la configuration du conduit afin de maintenir l'occlusion postérieure du dos de la langue. Aucune corrélation entre la vitesse de production et le coefficient d'asymétrie n'a été trouvée pour les labiales. À l'inverse, [t'] montre moins de dispersion du coefficient sur les nuages de points générés. Cela suggère donc que la configuration du conduit vocal reste similaire d'une répétition à l'autre. Cette grosse caisse se trouve dans un PB uniquement composé de sons coronaux et implique donc seulement des changements de position et de la forme de l'apex et la lame de la langue dans la zone dentale-alvéolaire. Il est possible que ce pattern contraigne aussi les variations de configuration du conduit. La corrélation entre la vitesse et le coefficient d'asymétrie de la coronale est très faible ($r > 0.48$). Les positions 1, 5 et 6 ne semblent pas avoir d'influence sur la production des grosses caisses.

3.2 Charleston

À l'inverse des grosses caisses, les charlestons sont des instruments dont le timbre est aigu, les imitations de charleston produites par le participant ont des CdG allant jusqu'à 10.5 kHz et des coefficients d'asymétrie négatifs ou proches de 0 (i.e. répartition dans les hautes fréquences). Nous avons exclu de notre analyse tous les [ts'] du PB3 (i.e. pattern de coronales) en position 4 car ils étaient réalisés entre un [ts'] et un [tʃ₁'].

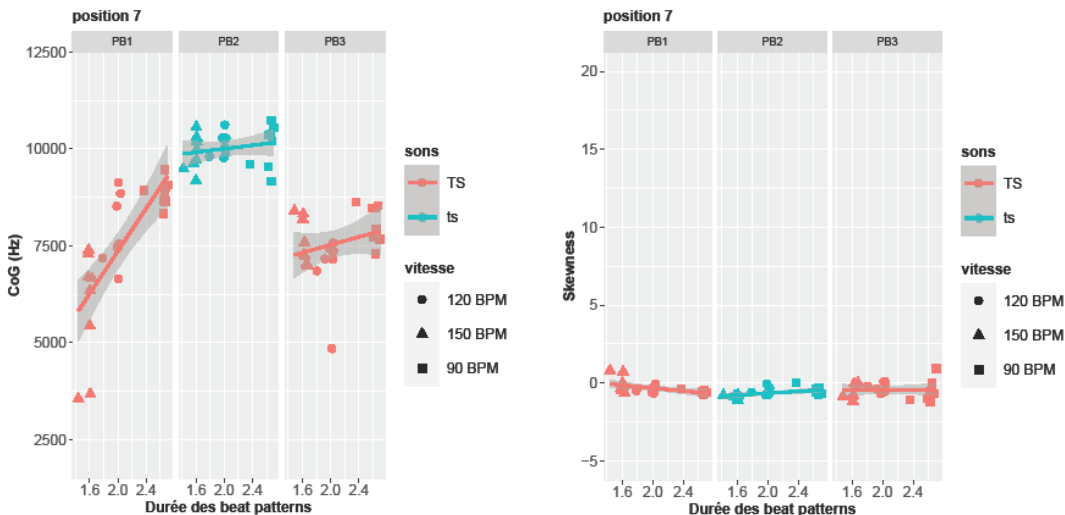


FIGURE 2 : Nuages de points du CdG (gauche) et coefficient d'asymétrie (droite) en fonction de la durée des PB de différentes imitations de Charleston en position 7. TS = [ts'], ts = [tʃ₁']. Les zones grisées indiquent l'intervalle de confiance.

L'analyse des charlestons montre que ces sons sont les plus variables. De plus, selon la position ou le PB où ils se trouvent, l'impact de la perturbation induite par la vitesse est différent. Les charlestons [ts'] du PB1 sont les plus affectés (figure 2). En effet, on observe une importante dispersion des valeurs de CdG et une forte corrélation entre le CdG et la vitesse pour les positions 2 ($r=0.70$), 4 ($r=0.87$), 7 et 9 ($r=0.76$). Ces résultats reflètent des variations articulatoires dues à un possible recul du lieu d'articulation. Cela supposerait donc une possible interaction avec les sons adjacents. En revanche, le même son dans le 3^{ème} pattern se comporte légèrement différemment. On observe moins

de dispersion des valeurs de CdG ce qui souligne plus de stabilité articuloire. Nous avons trouvé une corrélation du même ordre de grandeur en position 2 ($r=0.62$), plus faible en position 7 ($r=0.26$) et plus forte en position 9 ($r=0.83$). Ces résultats reflètent aussi des variations articuloires quant au lieu d'articulation. Enfin, [ʃ] montre moins de dispersion de CdG suggérant moins de variation de lieu d'articulation. La position la plus corrélée à l'augmentation de vitesse est la position 4 ($r=0.66$) et plus faiblement la position 9 ($r=0.46$). Ces résultats supposent encore une fois que la nature des contraintes articuloires propres à chaque PB influence la production des sons.

Concernant le coefficient d'asymétrie on observe peu de dispersion des valeurs du Skewness quelle que soit le type de charleston. Cela indique moins de variation dans la configuration du CV. On observe cependant des corrélations entre la vitesse et le skewness. Le skewness du PB1 est négativement corrélé à la vitesse, particulièrement pour la position 4 ($r=-0.77$) et la position 9 ($r=-0.62$). Le spectre se compose donc de fréquences plus hautes à vitesse rapide. De même, dans le PB3 il est négativement corrélé ; encore une fois cette corrélation est plus forte pour les position 4 ($r=-0.76$) et 9 ($r=-0.79$). À l'inverse, on observe le phénomène contraire pour le PB2 avec une corrélation positive.

3.3 Caisse Claire

Les caisses claires de notre corpus sont des sons longs. Il y a trois fenêtres d'analyse pour ces sons (Figure 3, chaque colonne est une fenêtre de 25 ms). Pour les affriquées, la première fenêtre correspond à la partie occlusive, inversement les fenêtres 2 et 3 correspondent à la partie fricative ; pour [ʃ*] les fenêtres 2 et 3 correspondent à du bruit. Les analyses de CdG montrent des valeurs comprises entre 1 et 5 kHz pour [ʃkʃ], entre 3 et 4 kHz pour [ʃ*] et entre 4 et 8 kHz pour [tʃʃ']. Quant aux valeurs du coefficient d'asymétrie elles sont comprises entre 1 et 6 (prédominance de moyennes et hautes fréquences) pour [ʃkʃ] et entre -0,5 et 1,5 (prédominance de hautes fréquences) pour [tʃʃ'] et [ʃ*]. La caisse claire ingressive [ʃkʃ] se compose d'une occlusive vélaire ingressive sourde et d'une fricative latérale ingressive sourde. La position dans le pattern n'influence pas le CdG de ces deux composantes. Le peu de dispersion des valeurs sur le nuage de points (Figure 3) et l'absence de relation forte entre la vitesse de production et le CdG ($r=-0.1$) montre que phase occlusive est stable. On observe un peu plus de variabilité en ce qui concerne la fricative latérale [ʃ] (figure 3) en terme de dispersion des valeurs des CdG. Il semble que la phase fricative soit tout aussi stable. Le coefficient d'asymétrie montre peu de dispersion ce qui nous laisse penser que la configuration du conduit vocal varie peu. Aucune corrélation n'a été faite entre les paramètres spectraux et la vitesse.

La caisse claire est en fait un clic [ʃ*] dont le relâchement se fait au niveau de la constriction postérieure (cf. Tuhuse & Traill, 1999 pour une description). Acoustiquement il se compose d'un relâchement avec la présence d'un burst acoustique et suivi de bruit. Sur la (Figure 3) on observe que la caisse claire vélique montre plus de dispersion dans les valeurs de CdG lors de la phase de relâchement (colonne 1) de la constriction vélaire que lors du bruit suivant le relâchement (colonne 2 et 3). Il n'y a pas de dispersion du coefficient d'asymétrie. Le relâchement sur la figure 3 est négativement corrélé à la vitesse concernant le CdG et positivement à la vitesse concernant le Skewness. Cependant cette corrélation vaut seulement pour ce son et la position 8 ; cela se traduit par une augmentation du CdG et une diminution du skewness à vitesse rapide.

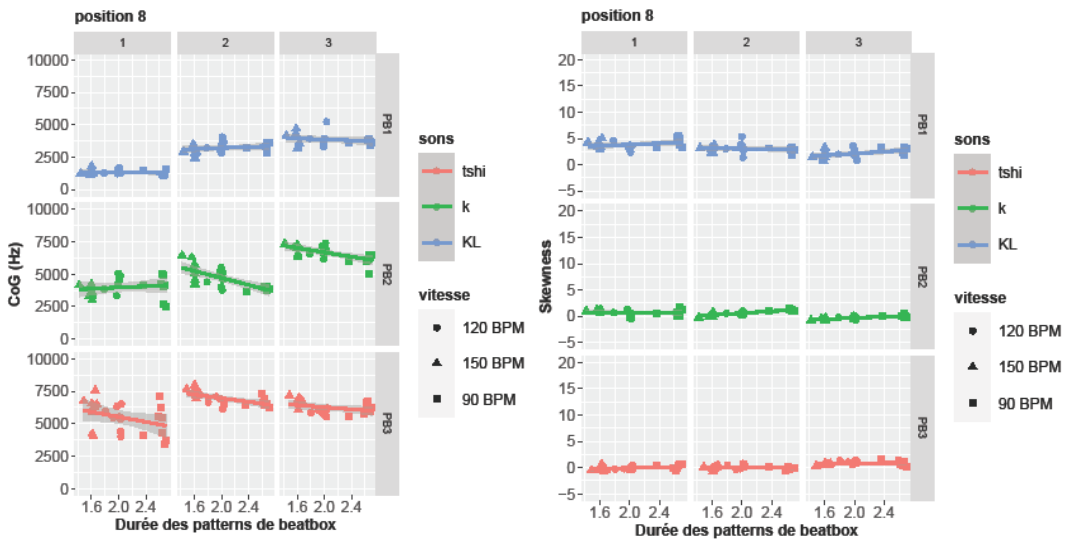


FIGURE 3 : Nuages de points du CdG (gauche) et coefficient d'asymétrie (droite) en fonction de la durée des PB de différentes imitations de caisse claire en position 8. Chaque colonne correspond à une fenêtre de 25ms. tshi = $[t_{\text{f}}^{\text{h}}]$, k = $[k^*]$, KL = $[k_{\text{L}}^{\text{h}}]$. Les zones grisées indiquent l'intervalle de confiance.

Enfin la caisse claire coronale $[t_{\text{f}}^{\text{h}}]$ est plus variable concernant le relâchement de la phase occlusive (Figure 3, fenêtre 1). Il y a une variabilité de positionnement articulaire. Le passage à la fricative post-alvéolaire est intéressant à analyser. Le CdG sur la 2^{ème} fenêtre d'analyse montre moins de dispersion du CdG que la partie occlusive. L'absence de relation entre la vitesse et le CdG suggère que la production de la phase fricative est plus stable que la phase occlusive. Conjointement, le coefficient d'asymétrie montre une stabilité de la configuration articulaire. Encore une fois on note l'absence de relation entre la vitesse et le coefficient d'asymétrie. Cela nous amène à penser que l'occlusive pourrait faire office de transitoire impulsif par analogie à la baguette de batterie qui bat la peau de l'instrument et que la friction ferait office de transitoire d'extinction qui porte les indices du timbre acoustique. Il se peut que pour des raisons perceptives le contrôle de la production de la phase fricative soit accru.

4 Discussion

Nous avons émis l'hypothèse que la vitesse entrainerait des perturbations du CdG, du coefficient d'asymétrie en fonction des positions. Les grosses caisses labiales, quelle que soit leur position dans le pattern, montrent une grande stabilité concernant le geste d'occlusion labiale, reflété par l'analyse du CdG. En revanche les variations de skewness laissent penser que la configuration du conduit vocal diffère d'une répétition à l'autre et d'une position à l'autre. Ces variations peuvent être dues à l'anticipation du geste suivant ou bien de la fin du geste précédent. À l'inverse la grosse caisse linguale montre la tendance inverse : des variations du lieu d'articulation et une stabilité de la configuration du CV. Le CdG des charlestons est plus impacté par le changement de vitesse comme le montre les résultats. Cela suppose que le geste lingual n'atteint pas le même lieu dans la zone dentale-alvéolaire. En revanche le peu de variation de skewness indique que la configuration du CV reste similaire d'une répétition à l'autre et d'une position à l'autre. Nous pensons initialement que

les sons vélares montreraient plus de variations articulatoires que les labiales, or les données ne vont pas dans ce sens. Le peu de variabilité du CdG de [ɰkɛ̃] et [ɰ*] suggèrent un bon contrôle de production des labiales et des vélares, et à moindre mesure de la caisse claire linguale [tʃĩ]. Concernant cette dernière, notre hypothèse prédisait qu'en tant que coronale, elle serait plus sujette à la variation articulatoire, mais les données nous racontent une autre histoire. Bien que la phase occlusive soit quelque peu plus variable, l'ensemble de l'affriquée est stable ce qui suggère un bon contrôle de cette caisse claire.

La position dans le pattern pourrait bien être un facteur influençant la production. Les positions les plus variables sont celles occupées par les charlestons, plus particulièrement les positions 4 et 9. C'est en effet sur ces positions que nous avons trouvé le plus de variation en terme de stabilité articulatoire. La vitesse agit comme une perturbation sur la production et le beatboxeur doit donc faire preuve d'adaptabilité pour respecter les contraintes rythmiques des structures à produire. Dehaïs Underdown et al. (2020) ont montré que pour s'adapter à la perturbation induite par la vitesse, le beatboxeur réduisait la durée des pauses entre chaque son ainsi que la durée des sons en eux-mêmes, plus particulièrement celle des caisses claires. Cela implique donc que le timing entre les gestes se réduit lorsqu'on augmente la vitesse. Une possible stratégie d'adaptation, ici, serait de réduire la précision articulatoire des charlestons afin de conserver une fluidité dans sa production et garder un meilleur contrôle articulatoire. Pour renforcer cette interprétation il serait bien d'analyser les erreurs de production ; cela permettrait d'alimenter la question du chevauchement des gestes articulatoires chez les beatboxeurs et du contrôle articulatoire. Il est évidemment nécessaire d'obtenir des données articulatoires pour confirmer ou réfuter cela.

5 Conclusion

L'étude du rythme et de son interaction avec des différences de vitesse de production est une question très intéressante en ce qu'elle permet d'étudier le contrôle articulatoire chez des experts de l'articulation. Dans de futures études (acoustiques et articulatoires) il faudra inclure plus de participants pour observer les tendances générales et les stratégies individuelles pour renforcer notre analyse sur les questions de contrôle et de précision articulatoire.

Remerciements

Ce travail est soutenu par le Labex EFL (ANR-10-LABX-0083)

Références

- BALL, M. J., HOWARD, S. J., & MILLER, K. (2018). Revisions to the extIPA chart. *Journal of the International Phonetic Association*, 48(2), 155-164. doi: [10.1017/S0025100317000147](https://doi.org/10.1017/S0025100317000147)
- BOERSMA, P. & WEENINK, D. (2006). Praat: doing phonetics by computer. Version 6.0.21, récupéré le 25 Septembre 2016 sur <http://www.praat.org/>
- CHO, T. & LADEFOGED, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics* 27, 207-229.
- DEHAIS UNDERDOWN, A., CREVIER BUCHMAN, L. & DEMOLIN, D. (2019). Acoustico-Physiological coordination in the Human Beatbox: A pilot study on the beatboxed Classic Kick Drum. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp.

- 142-146). Canberra, Australia: Australasian Speech Science and Technology Association Inc.: Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.).
- DEHAIS UNDERDOWN, A., VIGNES, P., CREVIER-BUCHMAN, L., & DEMOLIN, D. (2019). Human beatboxing: A multi-instrumental pilot. *The Journal of the Acoustical Society of America*, *146*(4), (pp. 3082-3082).
- DEHAIS UNDERDOWN, A., VIGNES, P., CREVIER BUCHMAN, L. & DEMOLIN, D. (2020). Articulatory control and beatboxing rate: A preliminary study. *12th International Seminar on Speech Production*. Providence.
- 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.
- INTERNATIONAL PHONETIC ASSOCIATION. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- JONGMAN, A., WAYLAND, R., & WONG, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252-1263.
- KINGSTON, J. (1985). *The Phonetics and Phonology of the Timing of Oral and Glottal Events (PhD)*. Berkeley University.
- LOUSADA, M. L., JESUS, L. M., & PAPE, D. (2012). Estimation of stops' spectral place cues using multitaper techniques. *Documentação de Estudos em Lingüística Teórica e Aplicada*, *28*(1), 1-26.
- NITTROUER, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, *97*(1), 520-530.
- PICART, B., BROGNAUX, S., & DUPONT, S. (2015). Analysis and automatic recognition of human beatbox sounds: A comparative study. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4255-4259). IEEE. doi: [10.1109/ICASSP.2015.7178773](https://doi.org/10.1109/ICASSP.2015.7178773)
- STOWELL, D., & PLUMBLEY, M. D. (2008). *Characteristics of the beatboxing vocal style*. Technical Report, Centre for Digital Music C4DMTR-08-01, Queen Mary, University of London, Dept. of Electronic Engineering. Récupéré sur <http://c4dm.eecs.qmul.ac.uk/papers/2008/Stowell08-beatboxvocalstyle-C4DM-TR-08-01.pdf>
- STOWELL, D., & PLUMBLEY, M. D. (2010). Delayed decision-making in real-time beatbox percussion classification. *Journal of New Music Research*, *39*(3), 203-213. doi: [10.1080/09298215.2010.512979](https://doi.org/10.1080/09298215.2010.512979)
- SUSSMAN, H. M., MCCAFFREY, H. A., & MATTHEWS, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, *90*(3), 1309-1325. doi: [10.1121/1.401923](https://doi.org/10.1121/1.401923)
- TABAIN, M. (2001). Variability in fricative production and spectra: Implications for the hyper-and hypo-and quantal theories of speech production. *Language and speech*, *44*(1), 57-93.
- TEAM, R. D. (2005). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Récupéré sur <http://www.R-project.org>
- TUHUSE, B. IX. & TRAILL, A. (1999). |Hán-|Hánsè, the desert Cisticola, implements an acoustic target , . In *ICPhS-14*, (pp. 1041-1042). San Francisco.
- WRIGHT, R., HARGUS, S., & DAVIS, K. (2002). On the categorization of ejectives: data from Witsuwit'en. *Journal of the International Phonetic Association*, *32*(1), 43-77.