Production de sons plosifs : comparaison du beatbox et de la parole

Annalisa Paroni

Production de sons plosifs : Comparaison du beatbox et de la parole

PARONI
Annalisa

Sous la direction de Nathalie HENRICH et Maëva GARNIER

Laboratoire : GIPSA-lab, Département Parole et Cognition

UFR LLASIC

Mémoire de master 1 recherche - 21 crédits – Sciences du Langage

Spécialité : Linguistique

Année universitaire 2015-2016

Production de sons plosifs : Comparaison du beatbox et de la parole

PARONI
Annalisa

Sous la direction de Nathalie HENRICH BERNARDONI et Maëva GARNIER

Laboratoire : GIPSA-lab, Département Parole et Cognition

UFR LLASIC

Mémoire de master 1 recherche - 21 crédits – Sciences du Langage

Spécialité : Linguistique

Année universitaire 2015-2016

Remerciements

Tout d'abord, je remercie mes directrices de mémoire Nathalie Henrich Bernardoni et Maëva Garnier, sans lesquelles ce travail n'aurait pas pu voir la lumière. Merci Maëva pour avoir accepté d'encadrer ce travail au sein de ton projet StopNCo, merci pour tous tes conseils et ta patience : j'ai beaucoup apprécié travailler avec toi. Merci Nathalie pour m'avoir donné la possibilité de travailler de nouveau avec toi, pour tes enseignements et ton soutien tout au long de cette année de reprise d'études, qui n'a pas été facile.

Je remercie sincèrement Claire Pillot qui a accepté d'être membre du jury et qui partage avec nous la passion pour cet art merveilleux qu'est le Human Beatbox.

Je remercie Christophe Savarieux pour son implication dans l'expérience EMA dans le cadre de ce mémoire.

Un gros merci à HugoBox pour avoir accepté de participer à notre étude EMA, pour avoir supporté beaucoup de matériel dans sa bouche pendant deux heures et pour avoir partagé avec nous sa passion et sa pratique du Human Beatbox.

Je remercie tous ceux qui, d'une façon ou d'une autre, m'ont aidé dans la rédaction de ce mémoire dans une langue que je ne maîtrise pas encore bien.

Enfin, merci à mes parents qui me soutiennent toujours dans mes études prolongées. J'espère un jour repayer tous vos efforts.

UNIVERSITÉ Grenoble Alpes

Déclaration anti-plagiat
Document **à scanner** après signature
et **à intégrer** au mémoire électronique

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : Paroni PRENOM : Annalisa

DATE : 28/06/2016 SIGNATURE : [redacted]

Mise à jour avril 2016

Sommaire

Introduction....6
Partie 1 : Partie théorique....7
Chapitre 1. Contexte de recherche....8
1. Le GIPSA-lab....8
2. Le projet StopNCo....9
3. Les notions d'effort vocal et d'efficacité....11
Chapitre 2. Les consonnes....15
1. Les consonnes du français....15
2. Les consonnes plosives orales non voisées....15
3. Indices acoustiques de discrimination des consonnes plosives....17
4. Contrôle des indices acoustiques de discrimination des consonnes plosives....20
Chapitre 3. Une production experte : le cas du Human Beatbox....23
1. Intérêt pour l'étude du Human Beatbox....24
2. Cadre théorique....25
3. Le Human Beatbox comme paradigme expérimental....26
4. Problématiques et hypothèses....28
Partie 2 – Partie expérimentale....30
Chapitre 5. Étude acoustique....31
1. Matériel et méthodes....31
2. Résultats....38
3. Discussion....54
4. Conclusions....57
Partie 3 – Perspectives de recherche....58
Chapitre 6. Étude articulatoire....59
1. Problématiques et hypothèses....59
2. Matériel et méthodes....61
3. Résultats....65
4. Discussion et conclusions....66

Introduction

Bien que l'on puisse considérer le fait de parler comme une action naturelle, cela nécessite d'une précise coordination entre la respiration, les gestes laryngés et les gestes articulatoires.

En particulier, l'étude de la production des consonnes plosives s'avère intéressante pour mieux comprendre le contrôle de la parole, puisqu'elle nécessite la coordination des gestes non seulement en amplitude et force, mais aussi dans leur organisation temporelle. Dans le cas où cette coordination est déséquilibrée, il peut se produire des tensions et des efforts excessifs et le risque de développement d'une pathologie de la voix augmente.

A ce jour, de très nombreuses études ont été menées sur la caractérisation des consonnes plosives et sur les gestes articulatoires et laryngés permettant le contrôle de plusieurs de leurs caractéristiques (transitions formantiques, délai d'établissement du voisement, …). En revanche, de nombreuses questions demeurent encore quant aux gestes permettant de contrôler finement les caractéristiques acoustiques du bruit de plosion (intensité, spectre, …).

Ce type de sons plosifs n'est pas produit qu'en parole, mais également dans le cadre d'une forme d'expression vocale : le human beatbox. Dans cette pratique vocale, des sons plosifs sont produits par imitation ou référence à des sons de percussion (naturels/acoustiques ou synthétiques/électroniques), avec une grande variété de timbres et une efficacité sonore importante. Les beatboxers[1] semblent donc posséder une maîtrise particulière du contrôle et de la variation des sons plosifs. C'est pourquoi dans cette étude nous avons souhaité explorer et comparer la production de sons plosifs dans le beatbox avec des consonnes occlusives en parole, de façon à caractériser les différences acoustiques et physiologiques entre ces modes de production, et à mieux comprendre les gestes permettant de contrôler les caractéristiques acoustiques du bruit de plosion.

Dans un premier temps, nous présenterons brièvement le contexte de recherche et le projet au sein duquel nous avons réalisé ce travail, ainsi que les notions d'effort vocal et d'efficacité vocale. Nous donneront ensuite un aperçu des consonnes plosives non voisées du français, les indices acoustiques de discrimination et leur contrôle. Nous aborderons donc l'intérêt que le Human Beatbox suscite pour notre travail de recherche et les hypothèses de notre étude. Ensuite, nous détaillerons notre étude acoustique et les résultats obtenus. En conclusion, nous présenterons les perspectives de recherche.

[1]Soit les chanteurs de beatbox.

Partie 1 : Partie théorique

Chapitre 1. Contexte de recherche

1. Le GIPSA-lab

Le GIPSA-lab, Grenoble Images Parole Signal Automatique, est une unité mixte de recherche du CNRS, de l'université Grenoble Alpes et de GrenobleINP, UPMF, UJF). Le travail de recherche qui se déroule au sein du Département Parole et Cognition (DPC) du GIPSA-lab est centré sur l'étude de la parole, de la voix, des langues et du langage. Cet objet de recherche est envisagé de façon multidisciplinaire, d'un point de vue des sciences du langage (notamment phonétique, phonologique et dialectologique), des sciences physiques (notamment acoustique et mécanique), des neurosciences cognitives et des sciences de l'ingénieur. Il s'ensuit que quatre équipes composent ce département, chacune ayant un champ thémathique spécifique.

Dans ce contexte de recherche, certains chercheurs travaillent sur la voix de plusieurs points de vue : physiologique, pathologique et artistique. Parmi ces chercheurs, travaillent Nathalie Henrich Bernardoni et Maëva Garnier.

Les recherches de Nathalie Henrich Bernardoni portent sur la compréhension physiologique et physique de la voix humaine (parole et chant). Les champs disciplinaires qu'elle aborde sont l'acoustique, l'aérodynamique et l'aéroacoustique de la production vocale humaine, le traitement des signaux vocaux humains, la phonétique acoustique et clinique, la voix chantée. En particulier, ses travaux portent sur l'analyse acoustique et électroglottographique de la voix parlée et chantée, la modélisation du débit glottique et les techniques de filtrage inverse, l'estimation spectrale et l'évaluation perceptive des paramètres de source glottique, les interactions source-filtre, et la voix chantée comme aide à l'apprentissage d'une langue étrangère.

Les travaux de Maëva Garnier portent sur les troubles de la voix, en particulier chez les personnes qui utilisent leur voix comme outil de travail. Ces troubles, pouvant être associés à des lésions au niveau des cordes vocales, se manifestent par une altération de la qualité et/ou de la puissance de la voix. Grâce à ses recherches, Maëva Garnier cherche à comprendre pourquoi certaines personnes sont plus susceptibles que d'autres de développer ces troubles, de comprendre comment optimiser la production de la voix (coordination des efforts laryngés, respiratoires, articulatoires) et la communication en

général, de façon à développer des programmes de prévention et de rééducation des troubles de la voix. Pour cela, Maëva Garnier travaille sur la notion d'effort et d'efficacité vocale, et plus largement sur les efforts et l'efficacité de communication. Elle mène des travaux sur des personnes maîtrisant certaines techniques expertes d'optimisation de la phonation (chanteurs) et essaie de comprendre l'intérêt de ces techniques d'un point de vue acoustique. Elle conduit également des travaux sur l'adaptation de la parole en situations perturbées, telles que les environnements bruyants, en examinant les différentes stratégies mise en œuvre par différents individus pour rester intelligibles dans de tels environnements, et en comparant les efforts de production que ces différentes stratégies requièrent. À présent, elle dirige le projet de recherche StopNCo, qui s'intéresse aux efforts et à la coordination des gestes de parole lors de la production de consonnes occlusives.

2. Le projet StopNCo

La production de la parole nécessite une coordination très précise de gestes respiratoires, laryngés et articulatoires. Ce contrôle de la production de la parole s'appuie sur l'exploration de son instrument vocal et sur la construction de représentations cognitives des limites et des contraintes de ce système. Par conséquent, on apprend inconsciemment quels sont les gestes les plus extrêmes que l'on est capable de produire et l'espace acoustique correspondant des sons que l'on peut produire. On apprend que certains gestes peuvent être contrôlés indépendamment ou que certains paramètres de production co-varient systématiquement. On apprend également les associations entre les variations des gestes respiratoires, laryngés et articulatoires et les variations résultantes dans le son produit. De plus, on apprend que certains gestes peuvent avoir des conséquences limitées dans le domaine acoustique (ce qui permet une plus ample variabilité du geste), tandis que d'autres gestes, bien que modestes, peuvent modifier le son résultant de manière significative (voir la Théorie Quantique de Stevens, 1972 ; 1989). Enfin, on apprend que certains gestes requièrent plus d'effort que d'autres et que certains gestes peuvent être même douloureux.

En ce qui concerne la parole, les locuteurs apprennent les différentes catégories de sons qui composent une langue, les caractéristiques acoustiques qui les distinguent et, par conséquent, la coordinations des gestes qui contrôle la variation de ces caractéristiques acoustiques. Les conditions et les limites qui permettent de discriminer les sons dépendent du contexte de communication. Ainsi, certains indices acoustiques peuvent être altérés ou

manquants, comme c'est le cas en environnements bruyants ou pour le chuchotement. Dans de telles circonstances, les locuteurs doivent modifier ou renforcer certains aspects de leur production afin de préserver leur intelligibilité. Pour toutes ces raisons, la production de la parole peut être considérée comme une action très complexe.

La production des consonnes plosives (/p/, /b/, /t/, /d/, /k/, /g/ en ce qui concerne le français) est particulièrement intéressante afin de mieux comprendre le contrôle de cette production. Premièrement, la discrimination entre les différentes plosives dépend de multiples indices, dont la hiérarchie n'est pas encore complètement comprise. Deuxièmement, la production des plosives requiert la coordination entre des gestes articulatoires en termes de timing (dont la littérature s'est bien occupée), mais aussi en termes de force et d'amplitude. Peu d'études s'intéressent à la coordination de force et amplitude entre tous les niveaux de production de la parole (souffle, comportement laryngé, articulation) et sur la relation entre variations articulatoires (position et force) et les variations spectro-temporelles des caractéristiques distinctives (bruit d'explosion, transitions formantiques). Ce mapping entre les domaines articulatoire et acoustique est non linéaire : de légers déplacements du point de constriction ou de faibles changements dans la coordination globale peuvent résulter en de grandes changements acoustiques. Au contraire, des faibles et imperceptibles modifications acoustiques peuvent résulter de grandes variations d'effort et d'efficacité de la coordination du geste. De plus, la relation entre variations acoustiques et catégories de consonnes plosives perçues est fortement non lineaire : les variations acoustiques de la réalisation standard d'une catégorie de plosive peuvent rester acceptable et la catégorie peut être perçue correctement jusqu'à un plafond à partir duquel l'articulation sera perçue en tant que déviante et non intelligible. Par conséquent, la caractérisation et une plus profonde compréhension de cette relation non linéaire entre gestes articulatoires, résultants en indices acoustiques et finalement en catégories perçues peuvent être d'un grand intérêt pour le diagnostic et le traitement de patients qui présentent une articulation déviante et non intelligible ou des locuteurs montrant un effort de production de la parole excessif et inefficace.

Le projet StopNCo vise à améliorer la compréhension du contrôle de la parole dans le cadre de la production des consonnes occlusives, en cherchant à répondre à quatre questions :

1. Quelles sont les caractéristiques acoustiques cruciales pour l'intelligibilité des consonnes plosives ?

2. Quelle coordination du souffle, de gestes laryngés et articulatoires permet la variation de ces caractéristiques acoustiques ? Dans quelle mesure cette coordination est-elle influencé par des contraints physiques ou par le contrôle du locuteur ?
3. Comment ce contrôle se développe-t-il chez les enfants et pourquoi dysfonctionne-t-il chez certains d'entre eux ?
4. Comment la coordination des gestes de parole peut-elle varier par rapport à l'efficacité, c'est-à-dire en ce qui concerne le rapport entre le niveau d'intelligibilité et les efforts physiologiques dépensés ?

À travers ces questionnements, le projet StopNCo vise à construire un modèle fonctionnel le plus complet possible de la coordination et du contrôle de la production des consonnes plosives, en tenant compte des différents niveaux physiologiques et des résultats perceptifs, de la variabilité intra-individuelle et individuelle, ainsi que de l'acceptabilité des variations dans cette coordination, en termes d'intelligibilité perçue, mais également en termes d'effort de production et de préservation de la santé vocale.

3. Les notions d'effort vocal et d'efficacité

Lors de la production vocale, l'énergie aérodynamique fournie par le souffle expiratoire est convertie en énergie acoustique grâce à la vibration des plis vocaux.

Dans des conditions normales, ce processus se déroule avec souplesse. En revanche, dans des conditions non optimales (comme, par exemple, en situation d'environnement bruyant ou de incoordination pneumo-phono-articulatoire), cette transformation d'énergie est plus difficile : plus d'énergie aérodynamique est nécessaire et les tensions musculaires augmentent, premièrement au niveau des muscles laryngés intrinsèques, mais aussi plus globalement au niveau des muscles posturaux et respiratoires, déclenchant une situation d'effort vocal.

D'après Giovanni et al. (2007, cité par Bourdin & Navion, 2013, p.12), « l'effort vocal correspond à une augmentation de la production d'énergie du sujet pour réaliser un son, accompagné par un contrôle constant de l'effort musculaire. Cette augmentation de la production d'énergie se caractérise par une élévation des tensions musculaires de tout l'appareil vocal.»

Une production vocale forcée provoque des adaptations à plusieurs niveaux. Au niveau acoustique, le changement plus important est l'augmentation de l'intensité vocale. À cela sont corrélées l'augmentation de la fréquence fondamentale (F0), un aplatissement de la pente spectrale et un renforcement spécifique de l'énergie entre 2 et 4 kHz. Au niveau articulatoire, les mouvements sont amplifiés (hyperarticulation), la compression des lèvres augmente sur les consonnes bilabiales, certains formants sont renforcés, la durée des voyelles est augmentée, tout comme la durée des consonnes selon certains auteurs (Bourdin & Navion, 2013), tandis que selon d'autres la durée des consonnes est réduite (Schulman, 1989).

Ces modifications, si persistantes dans le temps, peuvent avoir des conséquences sur les tissus laryngés et, par conséquent, sur la production de l'onde muqueuse. D'après Giovanni, Aumelas, Chapus, Lasalle, Remacle et Ouaknine (2004), l'augmentation de l'intensité vocale, de la F0 et de la durée des sons voisés provoque une augmentation des microtraumatismes subis par les plis vocaux.

Dans le cas où l'effort vocal est prolongé dans le temps et les tissus laryngés n'ont pas la possibilité de récupérer leur état physiologique, il peut s'établir une situation de fatigue vocale, dont un effort vocal accru perçu par le locuteur est un des symptômes plus importants (Solomon, 2008). D'après Vilkman (2004, cité par Solomon, 2008), un effort vocal accru est une parmi les trois variables représentant le risque plus élevé pour la fonction vocale, les autres étant un seuil de phonation et une F0 élevés.

L'effort vocal et le forçage vocal représentent donc un risque pour la santé vocale du locuteur. La littérature est riche d'études à tel sujet, pourtant il n'existe pas une définition de référence de forçage vocal ou d'effort vocal. Les deux sont également difficiles à mesurer et souvent ils sont étudiés indirectement par le biais de la mesure de l'efficacité vocale.

Si l'effort vocal correspond à une dépense d'énergie supérieure au nécessaire, une voix efficace est une voix qui produit « le maximum de résultat avec un minimum d'effort, de moyen et de dépense » (Pillot, 2004, p.8). De plus, « l'efficacité d'une voix se traduit non seulement par la qualité du geste phonatoire, mais aussi par la résistance aux sollicitations prolongées » (ibid., p. 21). Une voix efficace est donc une voix à la fois performante et endurante.

Plusieurs études qui s'intéressent à l'efficacité vocale utilisent des mesures de rendement, défini comme le rapport entre l'énergie utilisable et l'énergie totale dépensée

par le système (ibid.). Notamment, en ce qui concerne la production vocale, l'énergie utilisable correspond à la puissance acoustique rayonnée aux lèvres (proportionnelle à l'intensité) et l'énergie totale dépensée correspond à l'énergie consommée par le corps pour la phonation, c'est-à-dire la puissance aérodynamique (ibid.). Donc, Schutte (1980) fournit la formule suivante pour le calcul du rendement vocal :

$$\mathrm{RV} = \frac{I}{Psub \cdot q}$$

où I est l'intensité sonore, Psub la pression sous-glottique et q le débit d'air moyen. Donc, le rendement vocal est une grandeur physique calculable à partir de mesures aérodynamiques.

Or, un même locuteur peut mettre en place plusieurs stratégies d'adaptation en fonction de l'environnement où a lieu sa production vocale ou bien en fonction de la modalité d'expression vocale qu'il utilise. Chaque stratégie montre un degré d'efficacité différent, c'est-à-dire une consommation d'énergie différente (Garnier, 2007). De façon générale, plus la respiration, la phonation et l'articulation sont en équilibre, plus une voix est efficace (Bourdin & Navion, 2013).

Dans cette perspective, Bourdin & Navion (ibid.) décrivent l'efficacité de quatre types d'expression vocale : la parole, le cri, la voix projetée et le chant. L'efficacité de la parole spontanée physiologique dériverait du fait que « les forces de fermeture des plis vocaux et la pression sous-glottique s'équilibrent », permettant ainsi que « l'énergie transformée au niveau laryngé [soit] entièrement dédiée au signal acoustique » (ibid.). En revanche, en voix criée, le renforcement du signal sonore est obtenu grâce à l'augmentation de l'intensité et de la F0, ce qui provoque une dépense d'énergie supérieure. De plus, la respiration, la phonation et l'articulation ne sont pas en équilibre optimal, ce qui fait augmenter les tensions musculaires au niveau laryngé (Giovanni et al., 2007). Il en résulte une voix efficace d'un point de vue acoustique, mais tout-a-fait fatigable, qui, au cours du temps, peut créer des altérations de la muqueuse laryngée (Le Huche & Allali, 2010). La voix projetée est décrite comme ayant le rendement plus élevé et donc comme étant le type d'expression plus efficace. Cela est obtenu grâce à un équilibre optimal entre la phonation (intensité non accrue par rapport à la parole), l'articulation (précision articulatoire accrue et débit de parole ralenti), la respiration (abdominale) et la posture du corps (Le Huche & Allali, 2010). En voix chantée, deux stratégies principales permettent d'atteindre une bonne efficacité : l'allongement des voyelles et du temps de phonation entre deux inspirations, qui permet aux tensions de s'équilibrer, et la présence du formant du chanteur, qui amplifie les

fréquences entre 2 et 4 kHz, permettant un audibilité supérieure de la voix sans augmenter l'intensité (Sundberg et al., 1999).

À notre connaissance, les notions d'effort vocal, d'efficacité vocale et de forçage vocal n'ont été explorées jusqu'au présent qu'au niveau général des phrases ou des voyelles soutenues (voir Garnier, 2009), pour lesquelles certains comportements respiratoires et configurations glottiques apparaissent plus efficaces que d'autres. En revanche, les efforts impliqués dans la production des consonnes et, en particulier, des plosives n'ont pas encore été étudiés en détail. Il est connu que, à travers l'articulation des consonnes, le locuteur peut équilibrer les pressions des cavités de son appareil phonatoire et obtenir une intensité suffisante pour un minimum d'effort (Bourdin & Navion, 2013). Cependant, si le locuteur ne gère pas correctement les forces articulatoires, un effort excessif pourrait mener au développement de pathologies vocales. Dans la production des consonnes plosives, plusieurs gestes sont susceptibles de provoquer des tensions au niveau laryngé ou bien un effort excessif. Une Pio accrue pendant la phase d'occlusion peut avoir plusieurs conséquences. Entre autres, le larynx peut être poussé et subir des déplacements verticaux et les forces d'adduction des plis vocaux peuvent augmenter afin de s'opposer à l'augmentation de Pio et maintenir la fermeture. Toutes ces modifications peuvent être gérées plus ou moins efficacement par le locuteur. Dans le cas où le contrôle n'est pas efficace, au cours du temps le locuteur peut ressentir de la fatigue vocale (Morrison, 1997). De plus, la reprise de voisement après le relâchement de l'occlusion peut être difficile à contrôler. En fait, il est commun chez les sujets dysphoniques de commencer le voisement par un « coup de glotte » (Andrade et al., 2000), qui est un geste produit avec de l'effort au niveau laryngé.

Le projet StopNCo vise à mieux comprendre le rôle de l'effort vocal dans la production notamment des consonnes plosives. Cela a le but de dévoiler des stratégies de contrôle efficaces, qui permettent d'obtenir une coordination adéquate des gestes, grâce à laquelle la parole puisse être le plus intelligible possible avec le plus faible effort de production possible.

Chapitre 2. Les consonnes

1. Les consonnes du français

En français, les sons consonantiques peuvent être regroupés en trois classes :

1. les occlusives ou plosives, caractérisées par la présence d'un silence dû à la fermeture complète du conduit vocal ;
2. les constrictives ou fricatives, caractérisées par un bruit de friction dû à un fort rétrécissement en un point du conduit vocal ;
3. les sonantes, caractérisées par la présence d'une structure formantique.

Comme mentionné plus haut, le projet StopNCo s'intéresse aux consonnes occlusives et en particulier dans ce travail nous traitons une sous-catégorie de tels sons : les consonnes occlusives orales sourdes ou non voisées.

2. Les consonnes plosives orales non voisées

Comme tout son de parole, les plosives sourdes peuvent être décrites de deux façons, du point de vue articulatoire ou acoustique. Selon le point de vue, ces sons prennent deux noms différents, d'après l'événement qui les caractérise : occlusives du point de vue articulatoire, dû à l'occlusion complète du conduit vocal et plosives du point de vue acoustique, dû à l'explosion qui suit le relâchement. Étant donné que dans ce travail nous nous positionnons dans une perspective acoustique, dans la suite nous adopterons le terme plosives.

D'un point de vue articulatoire, on peut identifier trois phases principales :

1. *occlusion* : un articulateur mobile (les lèvres ou la langue) se déplace pour créer une fermeture complète du conduit vocal au niveau de la cavité buccale ;
2. *tenue de l'occlusion* : l'articulateur reste en place pendant que la pression augmente derrière la fermeture. Les plis vocaux cessent de vibrer pendant cette phase ;
3. *relâchement* : l'articulateur mobile se déplace en ouvrant le conduit vocal et l'air est rapidement expulsé. Après le relâchement, les plis vocaux recommencent à vibrer.

D'un point de vue acoustique, on peut identifier une suite de quatre événements (Calliope, 1989) (Fig. 3.1) :

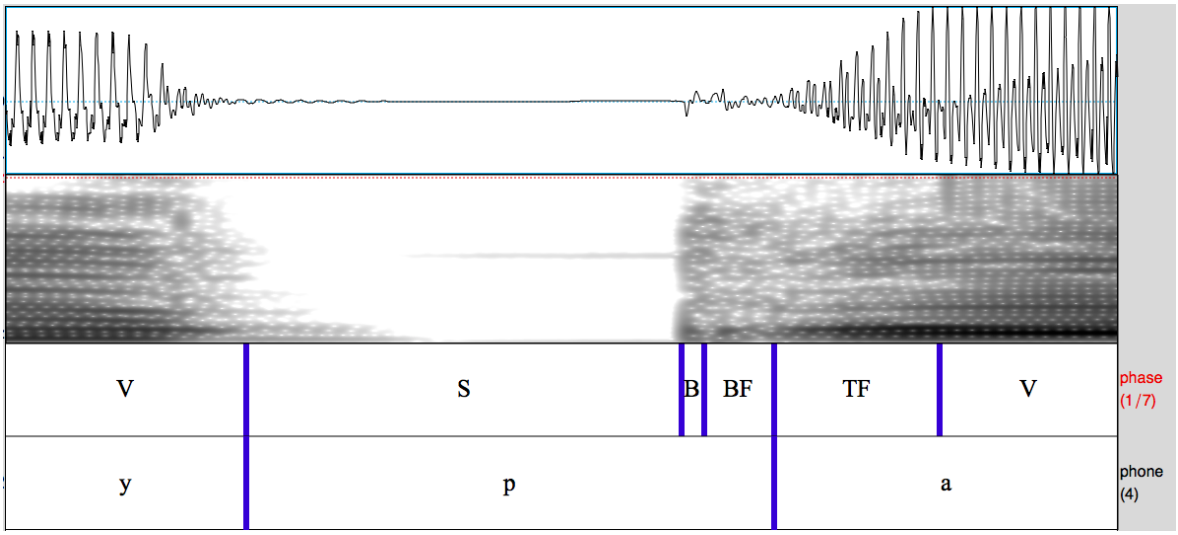

Figure 3.1: Signal audio, spectrogramme et phases de la production du segment /ypa/. V : voyelle ; S : silence ; B : « burst » ; BF : bruit de friction ; TF : transitions formantiques. Spectre : 0-12 kHz.

1. *silence* : il est dû à la tenue articulatoire de l'occlusion complète du conduit vocal. Une telle occlusion empêche l'air de s'écouler et provoque une augmentation de la pression intra orale derrière la fermeture. Sa durée est brève, 100 ms au maximum (Shadle, 1997) ;
2. *barre d'explosion* ou « burst » : au moment du relâchement soudain de l'occlusion, l'air comprimé derrière ce barrage est rapidement expulsé, ce qui provoque une perturbation acoustique impulsive (10 ms environ) et d'intensité variable. La forme spectrale de ce bruit est caractéristique du lieu d'articulation, car elle dépend du volume de la portion de cavité orale devant le lieu d'occlusion (voir §3.2) ;
3. *bruit de friction* : pendant la phase de relâchement de l'occlusion, les articulateurs se déplacent vers la position d'articulation de la voyelle suivante. Ce mouvement n'est pas instantané et dépend de la vitesse de déplacement des articulateurs. Par conséquent, dans le point du conduit vocal où a eu lieu l'occlusion l'air rencontre un rétrécissement qui provoque une turbulence. Naturellement, la durée du bruit de friction dépend de la vitesse de déplacement des articulateurs qui est notamment plus élevée pour les lèvres, intermédiaire pour la pointe de la langue et plus faible pour le dos de la langue. Par conséquent, le bruit de friction sera plus court dans le premier cas, de durée intermédiaire dans le deuxième cas et plus long dans le troisième cas. Néanmoins, le phone qui suit la plosive influence le bruit de friction : celui-ci sera plus long et intense si la plosive est suivie d'une voyelle antérieure fermée (notamment [i, e, y]) ou d'une palatale (e.g. [j]). Ce phénomène est connu sous le nom d'affrication et peut être expliqué par le fait que la langue n'a qu'un

court chemin à parcourir pour atteindre la cible vocalique. De ce fait, sa vitesse de déplacement est faible et la constriction du conduit vocal est de longue durée. Evidemment, ce phénomène affecte en priorité les occlusives où la langue joue un rôle articulatoire fondamental. Donc, une articulation affriquée est une articulation intense dont le bruit de friction est perceptible et peut durer de 20 ms à 40 ms ou plus ;

4. *transitions formantiques* : en cette phase, les formants sont présents, mais ne sont pas stabilisés, car les articulateurs sont encore en train de bouger vers la cible vocalique.

Les plosives sourdes du français sont au nombre de trois : [p, t, k]. Elles se distinguent par le lieu d'articulation, c'est-à-dire le point du conduit vocal où a lieu la fermeture (Fig. 3.2). [p] est une plosive bilabiale produite par l'occlusion du conduit vocal au niveau des lèvres ; [t] est une plosive apico-alvéolaire où l'occlusion est produite par l'accolement de l'apex de la langue à la partie alvéolaire du palais ; enfin, [k] est une occlusive dorso-vélaire et l'occlusion a lieu par l'accolement du dos de la langue au vélum.

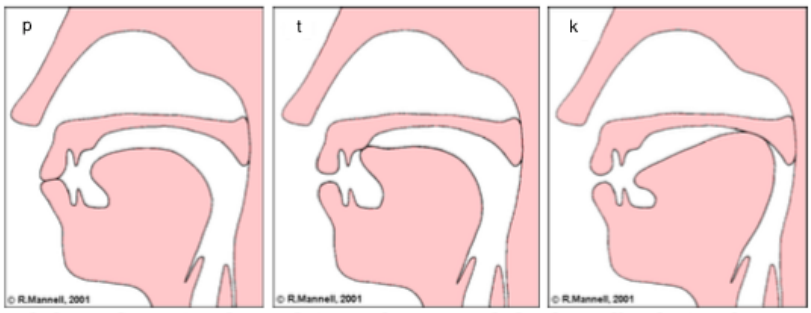

Figure 3.2: Représentation schématique de la configuration du conduit vocal lors de la phase d'occlusion des consonnes plosives orales du français /p/, /t/, /k/.

Image modifiée repérée à
http://clas.mq.edu.au/speech/phonetics/consonants/oral_stops.html

3. Indices acoustiques de discrimination des consonnes plosives

La perception et la discrimination des différentes catégories de plosives (voisées vs. non voisées, labiales, apico-alvéolaires ou palatales) se base sur plusieurs indices acoustiques (voir Calliope, 1989).

3.1. Indices de voisement

Pendant la phase de tenue de l'occlusion et donc pendant la phase de silence des consonnes plosives non voisées, les plis vocaux ne vibrent pas. Ils recommencent à vibrer après le bruit de friction. De façon générale, le délai d'établissement du voisement (ou VOT pour « Voice Onset Time » en anglais) a été classiquement défini de deux façons :

1. la durée (mesurée en ms) entre le relâchement de l'occlusion (la barre d'explosion ou « burst » en termes acoustiques) et la reprise de vibration des plis vocaux (Lisker et Abramson, 1964). Par conséquent, le VOT est toujours positif pour les plosives non voisées et a une durée de 30 ms environ (Serniclaes, 1984). Cette définition est valable uniquement pour les consonnes plosives.
2. la durée entre le relâchement de l'occlusion et le début de la structure formantique (Klatt, 1975, cité par Calliope, 1989). D'après cette définition, le VOT est toujours positif. Selon la définition de Klatt, il est possible identifier un VOT pour les plosives et les fricatives.

Le VOT et l'intensité du « burst » ne sont que deux des multiples indices de voisement (F0 initiale, F1 initiale, longueur de transition de F1, durée de l'occlusion et de la voyelle précédente) impliqués dans la discrimination des plosives. Néanmoins, ils représentent les deux indices les plus importants en situation normale (Lisker, 1975 ; Francis, Kaganovich & Driscoll-Huber, 2008).

3.2. Indices de lieu d'articulation

Le lieu d'articulation (pour rappel, au niveau des lèvres, des alvéoles et du vélum en ce qui concerne les plosives du français) configure le volume des cavités situées devant et derrière le point d'occlusion du conduit vocal. Comme mentionné plus haut, la taille de la cavité avant joue un rôle déterminant sur l'enveloppe spectrale du bruit de plosion créé au relâchement de l'occlusion (« burst »).

L'enveloppe spectrale du « burst » est caractérisée en la considérant comme une distribution de fréquence de 0 à 8 kHz et en calculant ses quatre premiers moments spectraux (Forrest, Weismer, Milenkovic & Dougall, 1988).

1. le **Centre de Gravité spectral** (**CDG**). Exprimé en Hz, il caractérise le barycentre spectral de l'énergie, c'est-à-dire, il indique de façon générale si le spectre est plutôt riche en basses fréquences, ce qui fait que le bruit est perçu comme grave ou en hautes fréquences, ce qui fait que le bruit est perçu comme aigu ;
2. l'**écart type du CDG**. Il permet de différencier entre formes spectrales, mais il ne permet pas de distinguer les spectres produits par des lieux d'articulation différents ;
3. l e **coefficient d'asymétrie** (« **skewness** » en anglais). Sans unité de mesure et normalisé, il informe sur le degré d'asymétrie de la distribution, par comparaison avec une distribution normale ou gaussienne. Ce coefficient est inférieur à 0 lorsque l'énergie augmente dans les hautes fréquences (le spectre est alors dit montant), supérieur à 0 lorsque l'énergie est renforcée dans les basses fréquences (le spectre est alors dit descendant) ;
4. l e **coefficient d'aplatissement** (« **kurtosis** » en anglais). Egalement normalisé et sans unité de mesure, il décrit le degré d'aplatissement de la distribution. Ce coefficient est supérieur à 3 dans le cas où la distribution est plus pointue et acute qu'une distribution gaussienne (la distribution est alors dite leptokurtique), inférieur à 3 dans le cas où la distribution est plus plate qu'une gaussienne (la distribution est alors dite platikurtique).

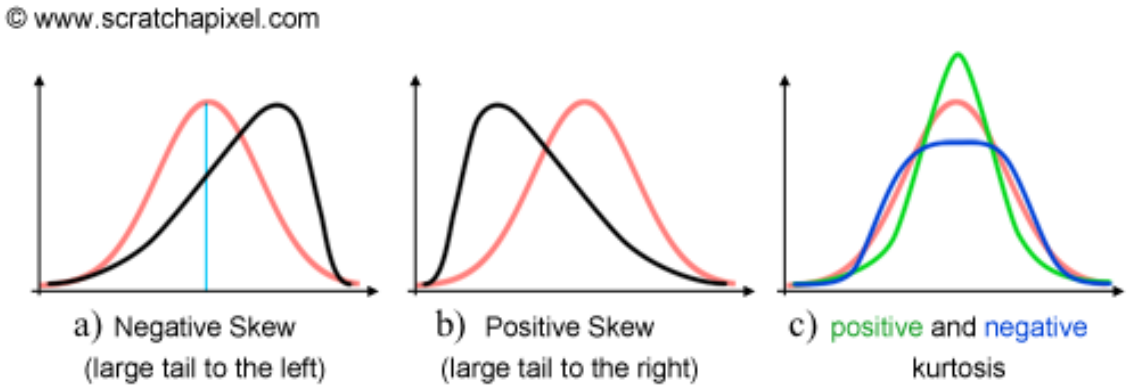

Figure 3.3: Illustration du coefficient d'asymétrie (a et b) et du coefficient d'aplatissement (c).

En ce qui concerne la plosive bilabiale [p], généralement il n'y a presque pas de cavité avant, sauf si la voyelle suivante est arrondie (c.à.d. [y, u]). Dans ce cas, la durée et l'intensité du « burst » sont faibles, son spectre est diffus et descendant avec une

amplification des basses fréquences. Cela se traduit par un faible CDG (~ 350 Hz, plus élevé pour les voyelles antérieures), un coefficient d'asymétrie positif et un faible degré d'aplatissement (Lousada, Jesus & Pape, 2012 ; Forrest et al., 1988).

Pour la plosive alvéolaire [t], la cavité avant est étroite, ce qui fait que l'intensité du « burst » est moyenne, son spectre est diffus et ascendant avec une amplification dans les hautes fréquences. Le CDG est élevé (~ 3-4 kHz), le coefficient d'asymétrie est négatif et le coefficient d'aplatissement est faible (Forrest et al., 1988).

En ce qui concerne la vélaire [k], la cavité avant dépend fortement de la voyelle qui suit la plosive : elle est étroite pour les voyelles antérieures (par exemple [i]) et plus grande pour les voyelles postérieures (par exemple [u]). L'intensité du « burst » est élevée et le spectre du bruit est compact.

La littérature scientifique a montré que la variation de ces indices acoustiques influence la perception du lieu d'articulation (Cooper et al., 1952 ; Gravel, 1983).

4. Contrôle des indices acoustiques de discrimination des consonnes plosives

4.1. Contrôle du VOT

Le voisement est conditionné par l'existence d'une différence de pression positive entre les cavités sous-glottiques et supra-glottique (van den Berg, 1958 ; Ohala & Riordan, 1979). Le temps minimum à partir duquel les plis vocaux peuvent recommencer à vibrer après le relâchement de l'occlusion correspond à l'instant où la différence de pression (orale et sous-glottique) atteint le seuil de pression phonatoire. Plus vite ce seuil est atteint, plus le VOT est court. Par conséquent, la durée minimale du VOT devrait pouvoir être contrôlée dans une certaine mesure par trois facteurs principaux :

1. la Pio maximale pendant l'occlusion : théoriquement, une plus grande Pio devrait impliquer un plus long VOT. En effet, le VOT augmenterait avec l'effort vocal (Arkebauer et al., 1987 ; Revis, Marques, Fredouille, Ghio, Giovanni, 2009);
2. la vitesse articulatoire : plus le relâchement est rapide, plus le VOT devrait être court ;
3. la surface de contact articulatoire : une large zone de contact peut induire une plus grande force de Bernoulli, qui fait rapprocher les articulateurs. En conséquence, les articulateurs s'écartent plus lentement au relâchement de l'occlusion et il faut plus

de temps pour que le seuil de pression phonatoire soit atteint (Stevens, Manuel & Matthies, 1999). Osfar (2011) a observé que les consonnes chuchotées présentent des plus faibles pourcentages de zone de contact et des durées plus grandes que les consonnes normales.

La durée minimale du VOT est fortement contrainte par la physique. En revanche, le VOT n'est pas toujours produit avec cette valeur minimale et peut tout-a-fait être plus long. Dans ce cas, il est plutôt relié à un mouvement volontaire d'adduction glottique.

En conclusion, les contraintes principales sur le contrôle du VOT sont :

- le lieu d'articulation, qui lui-même influence le max de Pio : plus la cavité derrière l'occlusion est étroite (comme par exemple dans le cas des plosives vélaires), plus la Pio augmente dans cette cavité (Hardcastle, 1973 ; Maddieson, 1997) ;
- la vitesse articulatoire : les lèvres et l'apex de la langue bougent à une vitesse plus élevée que le dos de la langue (Hardcastle, 1973) ;
- la quantité de contact articulatoire : la surface de contact entre la langue et le palais est plus grande au niveau du dos de la langue qu'au niveau de la pointe (Stevens et al., 1999) ;
- la durée de la tenue de l'occlusion.

4.2. Contraintes et contrôle de l'énergie du « burst »

Les différentes caractéristiques d'intensité et de spectre observées pour les plosives vélaires, alvéolaires et labiales peuvent être liées au volume de la cavité avant l'occlusion du conduit vocal (Kuhn, 1975). En effet, un volume plus élevé de la cavité antérieure (comme c'est le cas pour les plosives vélaires) résulte en une intensité plus élevée du « burst » et une enveloppe spectrale compacte, tandis qu'un volume plus faible de la cavité antérieure (comme c'est le cas pour les alvéolaires et les labiales) est lié à une intensité moins élevée et une enveloppe spectrale diffuse. Il s'ensuit que le locuteur peut être en mesure de varier l'intensité du « burst » et son enveloppe spectrale même au sein de la même catégorie de plosives en déplaçant légèrement le lieu d'occlusion du conduit vocal. C'est le cas notamment pour les consonnes chuchotées, dont la modification du lieu d'articulation provoque une augmentation du CDG spectral du « burst » (Osfar, 2010).

Toutefois, la relation exacte entre le déplacement de l'occlusion et les variations des paramètres acoustiques n'a pas encore été déterminée.

Plusieurs auteurs ont montré qu'il existe une forte corrélation entre l'activité musculaire des lèvres et la vitesse articulatoire impliquée (McClean & Tasko, 2003 ; Wolhert & Hanmen, 2000) ou entre l'augmentation de Pio et la force de compression des lèvres (Lubker & Parris, 1970), mais à notre connaissance, aucune étude n'a encore examiné l'influence que la Pio, la force et la vitesse articulatoires peuvent avoir sur l'intensité et l'enveloppe spectrale du « burst ».

Chapitre 3. Une production experte : le cas du Human Beatbox

Le human beatbox est un art vocal urbain relativement nouveau qui appartient à la culture Hip-Hop, né aux Etats Unis pendant les années 80 (Martino, 2009) dans le but de reproduire les sons produits par les machines rythmiques tels que ceux des séries Roland TR Rhytm, spécialement la Roland TR-808 (Tyte & White Noise, 2005). Puisque de tels machines étaient très chères et inabordables pour la plupart des artistes, les 'emcees'[2] ont commencé à imiter ces sons en utilisant un instrument beaucoup moins cher : leur bouche, ou mieux leur conduit vocal. Au tout début, la 'old skool' reproduisait des sons qui étaient pour la plupart percussifs : le beatbox était technique et son but était de « faire le maximum de bruit en un minimum de temps[3] ». Plus récemment la discipline a été étendue, elle est devenue plus complexe (Ojamaa & Ross, 2009) par imitation de toutes sortes de sons et son défi est de « produire des sons de plus en plus durs à produire[4] ». Le Human Beatbox est ainsi devenu, comme le dit le beatboxer français Ezra, « l'art de reproduire toutes les sons possibles avec la bouche » (cité par Clouet & de Torcy, 2010, p.25), en exploitant, comme le suggère un célèbre beatboxer australien, « la capacité innée pour la production de bruit non humain[5] » [traduction libre] (Tom Thum, TEDx talk, Sydney 2013). Cependant, les sons plosifs non voisés restent les sons plus importants et plus utilisés de cet art.

Le human beatbox est une forme artistique extrêmement personnelle, dans le sens qu'il n'y a pas une pédagogie commune : très souvent les beatboxers apprennent cet art tout seuls en imitant soit des sons instrumentaux ou environnementaux, soit d'autres beatboxers. Néanmoins, ces dernières années, la communauté du beatbox a montré un intérêt croissant vers l'échange d'idées et de techniques, et de nombreux congrès ont été organisés (Tyte, 2005). L'Internet est devenu un outil fondamental pour la communauté des beatboxers (Martino, 2009), pour le partage des savoirs et des techniques. Afin de faciliter les échanges, Splinter & Tyte (2002) ont développé une méthode de représentation des sons et des « beat patterns[6] », le sytème de notation standard (« Standard Beatbox Notation » ou SBN). Ce système utilise les caractères de l'alphabet standard anglais et il combine « la

[2]« Dérivé de l'abréviation originelle 'MC' (Master of Ceremonies – Maître de Cérémonie), désormais utilisé comme le terme générique pour désigner quelqu'un qui parle sur un beat, ou exécute des chansons que l'on peut nommer 'hip-hop'. » Source : www.urbandictionary.com (date de dernière consultation janvier 2014)
[3]Conversation privée avec Hugo Box, avril 2016.
[4]Conversation privée avec Hugo Box, avril 2016.
[5]« the innate hability for inhuman noise making ». Tom Thum, TEDx talk, Sydney, 2013.
Source: https://www.youtube.com/watch?v=GNZBSZD16cY
[6]Une combinaison de sons en Beatbox.

typographie et la phonétique afin d'utiliser les formes des lettres comme image de leurs sons[7] » [traduction libre] (ibid.).

1. Intérêt pour l'étude du Human Beatbox

Le human beatbox s'avère intéressant pour la recherche scientifique dans de nombreux domaines.

En ce qui concerne la phonétique expérimentale en générale, le human beatbox se montre intéressant puisqu'il offre la possibilité d'explorer le potentiel de fonctionnement du conduit vocal, car « les détails de la performance phonétique articulatoire du beatbox vont au-delà des combinaisons et des éventails de sons typiques que l'on trouve dans la production langagière de la plupart des langues » [traduction libre] (de Torcy, Clouet, Pillot-Loiseau, Vaissiere, Brasnu & Crevier-Bushman, 2013, p.10), jusqu'à employer des conformations extrêmes et rares (de Torcy et al., 2013 ; Proctor, Bresch, Byrd, Nayak & Narayanan, 2013 ; Sapthavee, Yi & Sims , 2014 ; Paroni, 2014) et des lieux et des mécanismes d'articulation non attestés en parole (Proctor et al., 2013 ; Paroni, 2014). Cependant, Proctor et collègues observent que « même quand les buts de la production sonore humaine sont extra-linguistiques, les locuteurs typiquement utilisent des modes de coordination articulatoire qui sont exploités dans les phonologies des langues humaines » [traduction libre] (2013, p. 1050). En outre, les études menées par Clouet & de Torcy (2010), de Torcy et al. (2013), Bourdin & Navion (2013), Septhavee et al. (2014) font apercevoir un scénario très intéressant : les beatboxers apprennent à exploiter leurs structures articulatoires au maximum, à les utiliser plutôt indépendamment les unes des autres, à les contrôler via un retour proprioceptif très développé et à mettre en place des mécanismes de protection contre les lésions glottiques. De plus, les caractéristiques articulatoires, acoustiques et aérodynamiques des sons du beatbox apparaissent plus marquées que celles des sons de parole, ce qui suggère que l'étude du human beatbox peut apporter des contributions précieuses afin de parvenir à des modèles fonctionnels plus complets de l'articulation, de la coordination articulatoire et phonatoire, et du contrôle de la parole.

[7]« […] combining typography and phonetics to use the shapes of the letters as pictures of their sounds ». Splinter & Tyte, 2002.

2. Cadre théorique

Dans l'étude scientifique de la voix, il est courant de comparer plusieurs types différents de production vocale, comme par exemple la voix parlée et la voix chantée, ou la voix parlée et la voix criée. Les structures utilisées sont les mêmes et les similitudes entre les divers types de production sont évidentes : en effet, lorsque nous chantons, crions ou parlons, nous produisons des mots, des phrases et ce qui change est la façon de les produire. En ce qui concerne le human beatbox, son lien avec d'autres formes de production vocale peut sembler moins évident, tout d'abord parce que des sons remplacent des mots, ensuite parce que le beatboxer cherche à dissimuler la provenance humaine du son et, par conséquent, toute ressemblance avec les sons langagiers, ce qui fait qu'il met en place un certain nombre de stratégies et de modifications articulatoires, acoustiques et aérodynamiques que l'on ne trouve pas dans d'autres formes d'expression vocale.

Néanmoins, toutes les études concernant la production vocale du human beatbox menées jusqu'à présent s'inscrivent, plus ou moins explicitement, dans l'hypothèse qu'il existe un continuum entre la production des sons langagiers et des sons du beatbox (cf. Clouet & de Torcy, 2010 ; Paroni, 2014 ; Septhavee, 2014). Les principaux arguments en faveur de cette hypothèse sont de deux natures : anatomo-physiologique et comportementale. Premièrement, comme c'est le cas pour la voix chantée, criée, ou projetée, les structures anatomiques de phonation et d'articulation sont les mêmes que pour la voix parlée. Dans tout type de production vocale, y compris le human beatbox, on utilise les poumons et les muscles respiratoires, le conduit vocal et les structures connexes pour gérer les aspects aérodynamiques, acoustiques et articulatoires. Deuxièmement, en ce qui concerne strictement le beatbox, la pratique de cet art est commencé presque systématiquement à partir de sons langagiers (cf. Clouet & de Torcy, 2010 ; Paroni, 2014), qui peu à peu sont déformés et modifiés, jusqu'à ne pas être perçus comme tels. Il s'agit de la procédure la plus utilisée dans le cadre de la pédagogie du beatbox, notamment dans les tutoriels que l'on retrouve en ligne, par exemple sur YouTube, où des beatboxers experts partagent leurs savoirs et compétences. Ceci est la méthode pédagogique utilisée également dans l'application française d'apprentissage du human beatbox, Beatbox Maker, récemment publiée. Très souvent, les beatboxers suggèrent de penser à un son langagier et ils donnent ensuite des astuces pour le modifier et le transformer en un son de beatbox. Par exemple, pour enseigner le son qui reproduit la grosse caisse ('kick drum' en anglais) de la batterie, ils suggèrent de penser à un [b] et expliquent la dynamique articulatoire à utiliser et comment le modifier pour faire en sorte que le son ressemble plus à un son produit par une

grosse caisse qu'à un [b]. Cette approche didactique est utilisée non seulement pour les sons isolés, mais aussi pour les suites de sons, comme, par exemple, [b d b], suite utilisée pour enseigner à alterner entre un son de grosse caisse et un son de charleston « hit-hat » ('high-hat' en anglais).

En conclusion, tout cela permet d'inscrire la production du human beatbox dans le cadre de la production para-langagière et de justifier l'intérêt dans la comparaison entre parole en tant que production langagière et le human beatbox en tant que production para-langagière (Lederer, 2005 ; Proctor et al., 2013 ; Paroni, 2014 ; Garrigues, 2015). Telle est la perspective théorique que nous adoptons dans notre travail.

3. Le Human Beatbox comme paradigme expérimental

Lorsque l'on parle avec des beatboxers de leur apprentissage, une caractéristique commune émerge de ces conversations : même avant de s'être intéressé au human beatbox, ils ont commencé à « jouer » et expérimenter avec « les sons et la bouche[8] ». Plusieurs d'entre eux commencent ces expérimentations depuis leur enfance, d'autres plus tard. Ce comportement provient du désir et du plaisir d'imiter des sons instrumentaux ou bien environnementaux.

Le but principal du human beatbox est précisément d'imiter et reproduire des sons instrumentaux, principalement percussifs. Les beatboxers sont capables de faire cela avec une grande précision. En comparant trois sons percussifs produits électroniquement et leurs contreparties beatboxés, Lederer (2005) a montré que certaines caractéristiques acoustiques étaient reproduites avec précision. Toutefois, en général, l'exactitude avec laquelle son beatboxer reproduisait les sons dépendait de la nature du son à imiter et non pas du degré de contrôle que le chanteur avait sur le son lui-même.

Comme nous l'avons mentionné plus haut, la pratique de cet art est basée elle-même sur ce mécanisme de manipulation de sons, notamment langagiers. Cela fait que, pour des fins artistiques, les beatboxers développent une maîtrise du contrôle de certains paramètres du conduit vocal que l'on ne retrouve pas en parole. Proctor et al. (2013) remarquent que leur beatboxer était capable de manipuler de façon fine des détails phonétiques tels que des paramètres de résonance et de débit d'air afin d'obtenir des effets sonores différentes sans pour autant changer le pattern articulatoire de base associé à une

[8]Conversation privée avec le beatboxer français Ezra, novembre 2013.

catégorie de sons percussifs donnée. Les beatboxers peuvent donc utiliser la même stratégie articulatoire générale, mais grâce à des modifications mineures au niveau des lèvres, de la langue et du larynx, ils arrivent à produire des effets différents.

Ces artistes sont capables d'exploiter des sons et des mécanismes articulatoires qui n'appartiennent pas à la phonologie de leurs langues natives (Proctor et al., 2013 ; Lederer, 2005 ; Garrigues, 2015). En fait, d'après Lederer (2005), plus le son est éloigné de la phonologie de la langue native, mieux il est reproduit (en termes acoustiques). Garrigues (2015) a mené des analyses acoustiques et aérodynamiques sur huit effets musicalement saillants produits par un beatboxer professionnel qui ont permis de confirmer l'utilisation par le chanteur de sons non appartenant au système phonologique de sa langue (le français). Notamment les analyses acoustiques portaient sur l'étude du signal audio, du spectrogamme et du signal EGG (électro-glottographique), afin de mesurer les paramètres d'intensité, la durée du son, le CDG, le coefficient d'asymétrie et le coefficient d'aplatissement. Pour les analyses aérodynamiques, le débit d'air oral a été également pris en compte. Les sons (trois pulmonaires, trois éjectifs, un implosif et un click) ont été regroupés en trois catégories, selon le lieu d'articulation de la plosive initiale : [p, t, k]. Les mesures acoustiques ont montré que l'intensité du « burst » est plus élevée pour les éjectives [p’] et [k’] que pour les autres sons appartenants à la même classe respective. En ce qui concerne les pulmonaires, l'intensité des sons bilabiaux est moindre que pour les effets alvéolaires et vélaires. La durée des sons semble dépendre du lieu d'articulation. Il nous semble néanmoins nécessaire de prendre en compte séparément les plosives des affriquées pour ne pas fausser les mesures. De façon générale, par rapport à leurs contreparties pulmonaires, les éjectives sont plus courtes et plus graves (moindre CDG), l'implosive vélaire plus intense et aigüe (CDG plus élevé) et le click dental plus court et aigüe (CDG plus élevé).

Comme mentionné plus haut, les beatboxers utilisent des comportements vocaux à risque (Clouet & de Torcy, 2010 ; de Torcy et al., 2013 ; Septhavee et al., 2014) sans pourtant montrer de signes permanents de pathologie. Bourdin & Navion (2013) ont avancé l'hypothèse que le manque de pathologie soit expliqué par le fait que les beatboxers développent « une utilisation spécifique et optimale de leur appareil phonatoire » (ibid., p. 25) qui leur permettrait d'être vocalement efficaces. Plus spécifiquement, ils développeraient « une maîtrise des paramètres acoustiques et aérodynamiques lors de la production, tels que le débit, l’intensité, la répartition de l’énergie et la pression sous-glottique » (ibid., p. 25). Afin de répondre à leur questionnement, les auteurs ont conduit

des analyses acoustiques et aérodynamiques sur la production vocale de quatre beatboxers, dont deux professionnels et deux amateurs, en comparant cinq modalités d'expression vocale différentes : la parole, le chant, le cri, la voix projetée et le human beatbox. C'est sur le même corpus que nous avons mené nos analyses. Les résultats ont montré que les valeurs des paramètres acoustiques et aérodynamiques du human beatbox se rapprochaient plus des valeurs de la voix criée par rapport aux autres modalités d'expression vocale. Cela suggère que les beatboxers exercent leur pratique dans une situation de forçage vocal. Néanmoins, le rendement vocal du human beatbox atteint des valeurs importantes (entre 41 et 53 dB), supérieures aux autres modes de production pour les deux beatboxers professionnels et du même ordre que les valeurs obtenus en voix projetée. Les auteurs ont conclu que les beatboxers arrivent à obtenir une gestion efficace de leur voix malgré l'utilisation de comportements à risque, caractéristiques du forçage vocal.

En conclusion, ces études montrent que le human beatbox est un bon paradigme expérimental pour l'étude des sons consonantiques, en tant que production experte et efficace.

4. Problématiques et hypothèses

Cette étude vise à traiter la problématique suivante : nous cherchons à mieux comprendre les mécanismes qui permettent de contrôler et d'augmenter l'efficacité sonore de la production de ces consonnes en cherchant à mieux comprendre les mécanismes qui permettent le contrôle des caractéristiques acoustiques du bruit de plosion des consonnes plosives, tels que l'intensité est les moments spectraux (CDG, coefficient d'asymétrie, coefficient d'aplatissement).

Pour cela, nous avons choisi de nous placer dans le cadre de production vocale du human beatbox, afin d'analyser les caractéristiques acoustiques de trois sons plosifs non voisés (/p/, /t/, /k/) et de les comparer aux consonnes correspondantes produites en mode parole conversationnelle, parole criée, parole projetée et chant.

Comme nous l'avons mentionné plus haut, le human beatbox est un bon paradigme expérimental pour l'étude des sons plosifs, puisque les beatboxers montrent un contrôle particulièrement fin de leur production et peuvent réaliser des gestes légèrement différents de la parole, tout en restant dans la même catégorie de son. De plus, l'efficacité vocale de

leurs productions plosives est accrue par rapport à d'autres types d'expression, notamment la parole.

Dans ce cadre, nous émettons les hypothèses suivantes :

1. H1 : les réalisations du beatbox montrent des différences acoustiques, aérodynamiques et articulatoires par rapport aux réalisations de la parole. En particulier, l'intensité du « burst », la Pio et la vitesse articulatoire seraient plus importantes dans le beatbox qu'en parole ; ils seraient plus grands dans le cri et la voix projetée par rapport à la parole. Le spectre des bruits de plosion dans le beatbox serait légèrement différent comparé à celui des consonnes parlées et chantées ;
2. H2 : l'intensité du « burst » serait corrélée à la Pio et à la vitesse articulatoire de relâchement de l'occlusion (mesurée à travers la dérivée temporelle du débit d'air oral) ;
3. H3 : le spectre du « burst » serait influencé principalement par le lieu d'articulation, mais pourrait être aussi influencé dans une moindre mesure par les caractéristiques aérodynamiques de la production, tels que la Pio et la vitesse du débit d'air oral ; notamment, nous nous attendons à que le CDG s'élève en fréquence, conjointement à l'augmentation de la Pio et de la vitesse du débit oral .

Partie 2 – Partie expérimentale

Chapitre 5. Étude acoustique

Pour nos analyses acoustiques, nous avons exploité un corpus déjà existant : il s'agit du corpus constitué par Bourdin & Navion en 2013 lors de leur étude sur l'efficacité vocale chez les chanteurs de human beatbox. Pour plus d'informations, voir Bourdin & Navion (2013, p. 28 – 39).

1. Matériel et méthodes

1.1. Sujets

Les données ont été recueillies sur quatre sujets décrits sur la Fig. 4.2.1.

Nom	**BB1**	**BB2**	**BB3**	**BB4**
Sex	H	H	H	H
Age	30	28	23	20
Profession	Beatboxer	Beatboxer	Animateur	Gendarme
Nombre d'années de pratique	12	12	8	2
Travail de la voix	- voix - rythme - scratchs - basses - puissances - finesse - texture des sons	- voix - rythme - scratchs - basses - puissances - finesse - texture des sons	- endurance - jeux de scène - adaptabilité à l'ambiance générale	- rythme - basses
Autre instrument	Basse, guitare	Guitare	Batterie	Non
Nombre d'heures de pratique sans ressentir de gêne	5 à 6h	2h et plus	5 à 6h	1h

Figure 4.2.1: Les sujets de l'étude acoustique.

Tous les beatboxers sont des hommes âgés entre 20 et 30 à l'époque des enregistrements. Deux d'entre eux sont des beatboxers professionnels qui pratiquent cet art vocal depuis 12 ans et vivent de cette activité, en tant que chanteurs et enseignants. Les

deux autres beatboxers sont des amateurs dont l'un avec déjà un grand nombre d'années de pratique (8 ans).

1.2. Protocole et corpus

Les enregistrements des données acoustiques et aérodynamiques ont eu lieu dans la chambre sourde du laboratoire GIPSA-lab à Grenoble, à l'aide de la station d'évaluation assistée EVA2® (SQLab-LPL, Aix en Provence, France).

> Pour recueillir les signaux, EVA2 dispose de plusieurs capteurs placés sur un ensemble appelé « pièce à main ». Celle-ci est disposée face au sujet, à hauteur de son visage, afin que celui-ci adopte une posture la plus naturelle possible au cours des mesures. (Bourdin & Navion, 2013, p. 31)

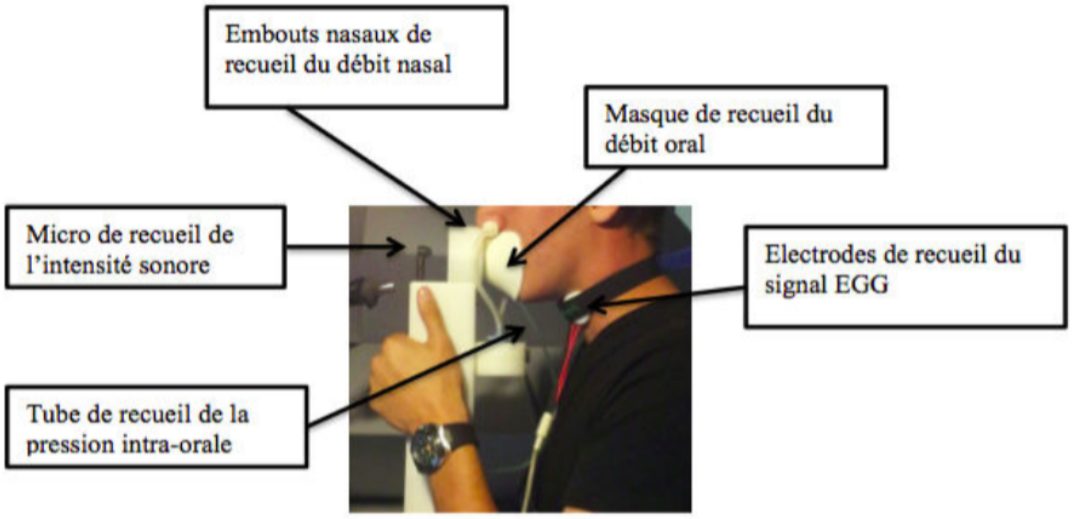

Figure 4.2.2: Dispositif EVA2 et EGG, au cours de la manipulation. D'après Bourdin & Navion, 2013, p. 32.

Le dispositif EVA2 (Fig. 4.2.2) a donc permis d'obtenir les mesures suivantes :

- l'intensité vocale en décibels (dB), acquise via le microphone intégré dans le dispositif EVA2, à la fréquence d'échantillonnage de 12500 Hz sur 16 bits ;
- la Pio en hectopascals (hPa), acquise à la fréquence d'échantillonnage de 6250 Hz sur 16 bits ;
- le signal EGG, acquis à la fréquence d'échantillonnage de 25000 Hz sur 16 bits ;
- le débit d'air oral en millilitres par seconde (ml/s), acquis à la fréquence d'échantillonnage de 25000 Hz sur 16 bits ;

- le débit d'air nasal en millilitres par seconde (ml/s), acquis à la fréquence d'échantillonnage de 25000 Hz sur 16 bits.

Pendant les enregistrements, le beatboxer était debout dans la chambre sourde, assisté par deux expérimentateurs. Le sujet visualisait les consignes et phrases du protocole sur un écran placé en face de lui.

La partie du corpus originel utilisée pour notre étude est constituée de 32 phrases rangées en quatre séries (voir Annexe 1). Ces phrases sont utilisées pour la réhabilitation orthophonique afin de travailler l'articulation des consonnes plosives principalement sourdes. Chaque phrase a été produite et enregistrée dans cinq modes vocales différents :

1. en parole conversationnelle ;
2. en parole criée ;
3. en parole projetée ;
4. en voix chantée ;
5. en beatbox.

En ce qui concerne les phrases beatboxées, BB1 avait auparavant réalisé des correspondances écrites, lisibles par le sujet tout au long des enregistrements. Bourdin & Navion rapportent que « deux des sujets ont montré des difficultés à respecter la consigne et en particulier la structure de phrase en beatbox » (ibid. p. 35).

1.3. Méthodologie et outils d'analyse

L'intérêt de l'étude portant sur l'analyse acoustique des « bursts » des consonnes non voisées, il a fallu repérer les segments correspondant dans les fichiers audio. Nous avons segmenté manuellement deux des quatre séries de phrases du corpus (la première et la troisième série), à l'aide du logiciel Praat (Boersma & Weenink, 2006) de façon à explorer un nombre suffisant d'occurrences, au moins 80 réalisations pour chaque cible (/p, t, k/) et pour chaque mode d'expression.

De façon générale, seules les données audio et EGG ont été observées pendant la phase de segmentation.

Pour l'annotation, nous avons utilisé un TextGrid composé de six Tiers (Fig. 4.2.2 et Fig. 4.2.3) :

- le premier Tier était dédié à l'annotation de l'expression vocale : parole, chant, cri, voix projetée ou beatbox ;
- le deuxième Tier contenait les phrases entières produites par le sujet ;
- le troisième Tier reportait la transcription des phrases en beatbox ;
- le quatrième Tier était dédié aux cibles consonantiques attendues ;
- le cinquième Tier reportait les consonnes effectivement produites par le sujet ;
- le sixième Tier contenait les voyelles qui suivaient les plosives. En ce qui concerne le beatbox où les sons plosifs ne sont pas suivis de sons voisés, la voyelle annotée sur ce Tier faisait référence à la voyelle correspondante attendue dans la phrase en parole.

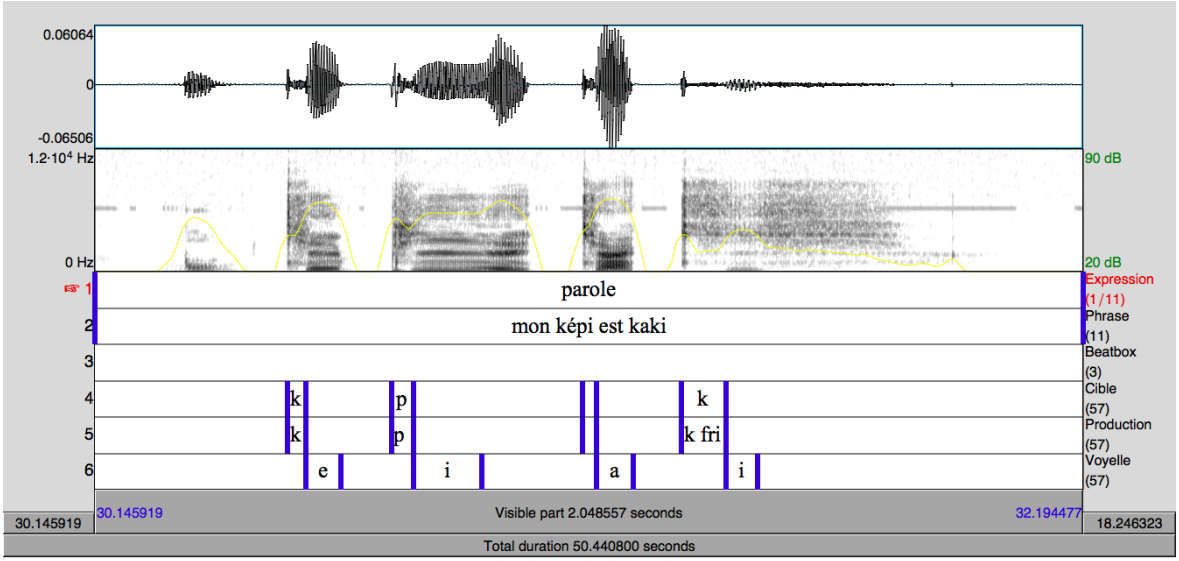

Figure 4.2.3: Segmentation sous Praat d'une phrase parlée de façon conversationnelle.

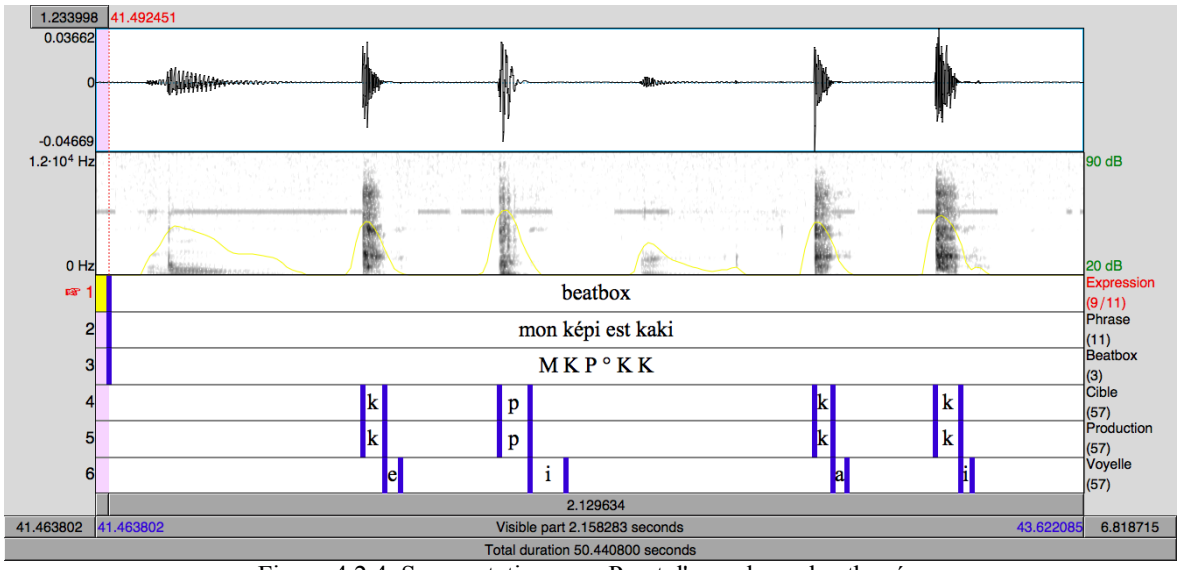

Figure 4.2.4: Segmentation sous Praat d'une phrase beatboxée.

En ce qui concerne la segmentation des « bursts » pour toutes les types d'expression, la borne gauche correspondait à l'instant d'explosion détecté sur le signal audio et sur le spectrogramme ; pour les trois types de voix parlée et pour la voix chantée, la borne droite correspondait à l'instant de reprise du voisement, détecté grâce au signal audio et sur le spectrogramme et parfois dans les cas ambigus en recourant au signal EGG (voir Figure 8). Dans le cas du beatbox où il n'y a pas de voyelle et où les sons sont isolés, la borne droite correspondait à la fin de la forme d'onde du signal audio (Fig. 4.2.4). Dans le cas des consonnes parlées et chantées, suivies d'une voyelle, cette durée du bruit correspondait également au VOT.

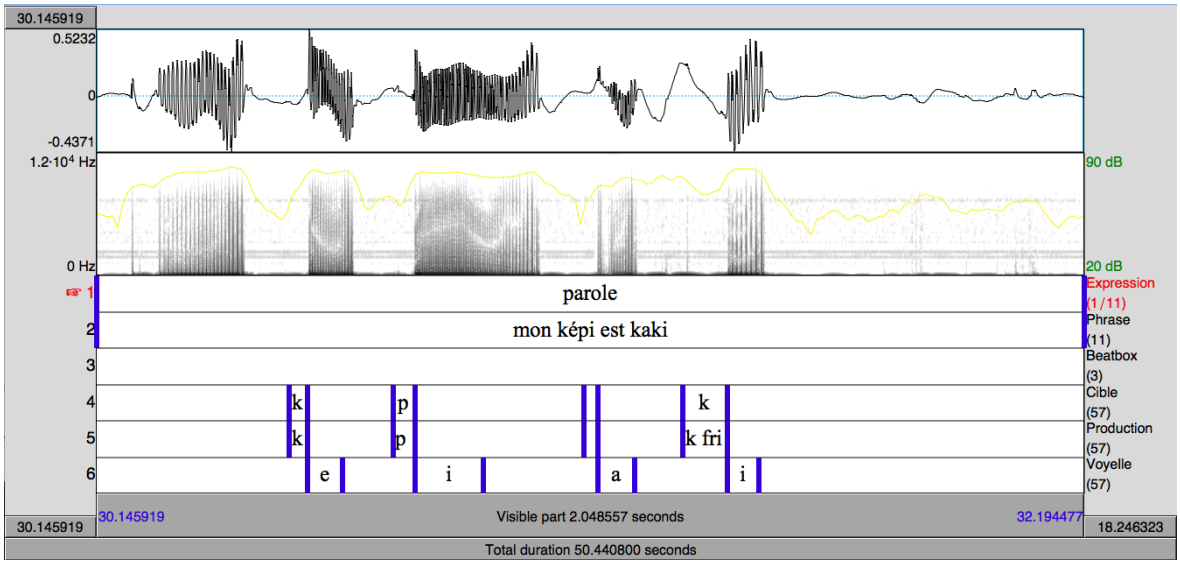

Figure 4.2.5: Illustration de l'usage du signal EGG pour la segmentation d'une phrase parlée.

Pour les annotations, nous n'avons pas pu utiliser l'API, car tels caractères ne sont pas compatibles avec le logiciel MATLAB (voir §1.3.1). Nous avons donc utilisé les caractères de l'alphabet standard.

Comme mentionné plus haut, deux beatboxers n'ont pas pu respecter la structure des phrases en beatbox prévue par le protocole, c'est-à-dire il n'ont pas pu respecter l'ordre et le nombre de sons prévus. En ces cas, la correspondance cible-production a été accordée si le son produit faisait partie du groupe de sons attendus, sans tenir compte de la structure de la phrase beatboxée attendue.

1.3.1. Statistique descriptive

Une fois la phase de segmentation et d'annotation terminée, des programmes MATLAB ont été écrits pour extraire les informations contenues dans les fichiers TextGrid, et les signaux d'intensité, de Pio, de débit d'air oral et calculer les paramètres acoustiques d'intérêt pour cette étude. Un fichier Excel était généré contenant les informations suivantes :

1. le nom du beatboxer ;
2. son code (BB1, BB2, BB3, BB4) ;
3. son âge à l'époque des enregistrements ;
4. les années de pratique de beatbox ;
5. l'activité (professionnel ou amateur) ;
6. la phrase du corpus à laquelle appartenait la cible consonantique ;
7. la phrase correspondante en beatbox (seul pour le mode d'expression beatbox) ;
8. le mode d'expression ;
9. la consonne cible ;
10. la consonne effectivement produite ;
11. la voyelle suivante la consonne cible ;
12. le temps de début du « burst », en ms ;
13. le temps de fin du « burst », en ms ;
14. la durée du « burst », en ms ;
15. le CDG, en Hz ;

16. le coefficient d'asymétrie (« skewness ») ;

17. le coefficient d'aplatissement (« kurtosis ») ;

18. le max de Pio, en hPa ;

19. l'intervalle de temps entre le début du « burst » et le max de Pio, en ms ;

20. le max de la vitesse du débit oral, en ml/s^2;

21. la valeur moyenne d'intensité pour la consonne produite, en dB ;

22. la valeur maximale d'intensité pour la consonne produite, en dB.

A partir de ce fichier, pour chaque paramètre d'intérêt acoustique (c'est-à-dire à partir du point 14 de la liste ci-dessus) un autre programme MATLAB a calculé les moyennes et les écarts types. En outres, les pourcentages de correspondance entre cible attendue et consonne produite pour chaque chanteur et chaque mode d'expression ont été calculés sur une feuille Excel et les graphes correspondants générés.

Nous n'avons pas pu utiliser toutes les données concernant la Pio, car non fiables. En effet, chez trois sujets, la salive a obstrué le tube de mesure. Les seules données de Pio exploitables sont celles de BB1.

1.3.2. Statistique inférentielle

Nous avons effectué des analyses statistiques à l'aide du logiciel R afin de comparer les moyennes obtenues pour chaque paramètre et dégager les corrélations significatives entre les paramètres étudiés.

Nous avons d'abord cherché un modèle qui s'adapte au mieux à nos données et en suite conduit des test statistiques.

Pour nos analyses statistiques, nous nous sommes basée sur un modèle mixte, où le type d'expression et la consonne cible sont considérés en tant que facteurs à effet fixe et le sujet en tant que facteur à effet aléatoire.

2. Résultats

2.1. Correspondance cible-production

2.1.1. Structure de la phrase

En ce qui concerne les phrases beatboxées, les deux beatboxers professionnels montrent une bonne capacité à reproduire correctement leur structure, c'est-à-dire, globalement ils respectent l'ordre et le nombre de sons prévu par le protocole. En revanche, les deux amateurs ont des difficultés à respecter cette structure : les sons sont produits dans un ordre qui ne correspond pas à celui attendu ou bien un ordre aléatoire; en plus, le nombre de sons produits est supérieur au nombre de sons de la phrase correspondante du protocole. Toutefois, les sons eux-mêmes ne sont pas au hasard : généralement les typologies de sons produits correspondent aux typologies de sons attendus (voir §2.1.2). De plus, souvent les consonnes sont suivies d'une voyelle chuchotée correspondante à la voyelle attendue dans les autres modes d'expression, alors qu'en Beatbox aucun son n'est prévu par le protocole après la consonne. Cette difficulté à se détacher d'une modalité expressive de parole est notamment présente chez BB2, tandis que chez BB3 et BB4 ce phénomène est moindre et il ne se présente pas chez BB1.

Quant aux autres types d'expression, la structure des phrases est exactement respectée, sauf dans un cas, où BB1 rajoute un mot monosyllabique dont la consonne est une cible prévue dans la phrase originelle. Les modalités expressives sont toujours bien respectées.

2.1.2. Correspondance consonne cible-consonne produite

Globalement, les catégories des consonnes sont bien respectées dans toutes les modalités d'expression vocale : les cibles /p/ sont produites en tant que sons commençant par [p], les cibles /t/ sont produites en tant que sons commençant par [t] et également les cibles /k/ sont produites en tant que sons commençant par [k].

Plus spécifiquement (voir Annexe 2), en ce qui concerne le beatbox, un pourcentage négligeable (inférieur à 5%) de cibles est produit en tant que sons appartenant à une catégorie différente (par exemple, une cible /p/ est produite [t]).

La cible /p/ est produite en plusieurs variantes par tous les sujets, sauf BB1. Les variantes sont : une plosive bilabiale non voisée [p], une plosive bilabiale non voisée suivie d'un trille bilabiale non voisé [pʙ̥], un trille bilabiale non voisé [ʙ̥]. Toutes ces réalisations représentent des variantes d'un même effet, la grosse caisse (Paroni, 2014). C'est pourquoi

nous les avons regroupés dans une catégorie unique de sons produits, qui rend compte d'un pourcentage entre 90% et 99% de correspondance avec la cible.

La cible /t/ est également produite en plusieurs variantes chez tous les chanteurs. Les variantes sont : une plosive alvéolaire non voisée [t], une plosive alvéolaire non voisée suivie d'une fricative alvéolaire non voisée [ts], une plosive alvéolaire ou post-alvéolaire non voisée suivie d'une fricative post-alvéolaire non voisée [ʧ]. Les deux sons affriqués ne sont pas toujours aussi nets que leurs contreparties parlées et chantées et les différences entre les trois sons [t], [ts], [ʧ] sont parfois subtiles, pourtant il est généralement possible de les distinguer d'un point de vue perceptif et spectrographique. En effet, chacune de ces réalisations représente un effet différent en beatbox : [t] et [ts] reproduisent deux types différent de charleston et [ʧ] reproduit un type particulier de cymbale. Chez tous les chanteurs sauf BB3, le son présent en pourcentage plus élevé est [t] (entre 40% et 70%). Le son [ts] est aussi beaucoup répandu (entre 20% et 55%). Enfin, [ʧ] est le son le moins produit, sauf que chez BB2 (entre 10% et 25%).

La cible /k/ est la mieux réalisée : la correspondance entre consonne cible et consonne produite est supérieure à 90% chez tous les beatboxers et atteint 100% chez BB1.

En ce qui concerne les autres quatre modalités d'expression (voir Annexe 2), la correspondance entre consonne cible et consonne produite dépend du contexte vocalique : la correspondance est presque parfaite si la plosive n'est pas suivie d'une voyelle antérieure fermée. Toutefois, si la voyelle suivant la consonne est antérieure fermée, le bruit de friction de la plosive est allongé et sa production est affriquée. Ce phénomène est globalement présent chez tous les sujets de notre étude de façon comparable et dans tous les types d'expression. Plus précisément, la cible /t/ est réalisée avec un allongement du bruit de friction entre 25% et 40% des fois en voix parlée et projetée chez tous les chanteurs et environ 30% des fois en voix criée et chantée chez BB2, BB3, BB4. Quant à BB1, il n'y a aucun allongement du bruit de friction en voix criée et le phénomène est presque négligeable en voix chantée. Le bruit de friction des réalisations de la cible /k/ est allongé entre 15% et 30% des fois en voix parlée et projetée, et globalement moins de 10% des fois en voix criée et chantée chez tous les sujets. En ce qui concerne la cible /p/, le phénomène est négligeable (<10%) chez tous les sujets.

2.2. Caractéristiques acoustiques des bruits de plosion

2.2.1. Durée du bruit

La durée du bruit désigne ici dans tous les cas l'ensemble du bruit de plosion (le « burst ») et du bruit de friction.

A noter que pour les plosives suivies d'une voyelle, c'est-à-dire les plosives produites en mode parole (conversationnelle, criée, projetée) et chant l'annotation du bruit se termine lorsque le voisement reprend. La durée du bruit correspond donc au VOT. Pour les sons plosifs du beatbox, non suivis d'une voyelle, l'annotation du bruit se termine lorsque la forme d'onde acoustique cesse. La durée du bruit peut donc être de fait légèrement différente.

Comme le montrent les Figg. 4.2.6, 4.2.7 et 4.2.8, la durée des sons plosifs du beatbox est toujours significativement supérieure à la durée du bruit des consonnes plosives dans les autres types d'expression, quelle que soit le lieu d'articulation (/p/ : 29.619 ± 2.004 ms, $p<0.001$; /t/ : 22.297 ± 2.537 ms, $p<0.001$; /k/ : 36.432 ± 2.106 ms, $p<0.001$) et le chanteur.

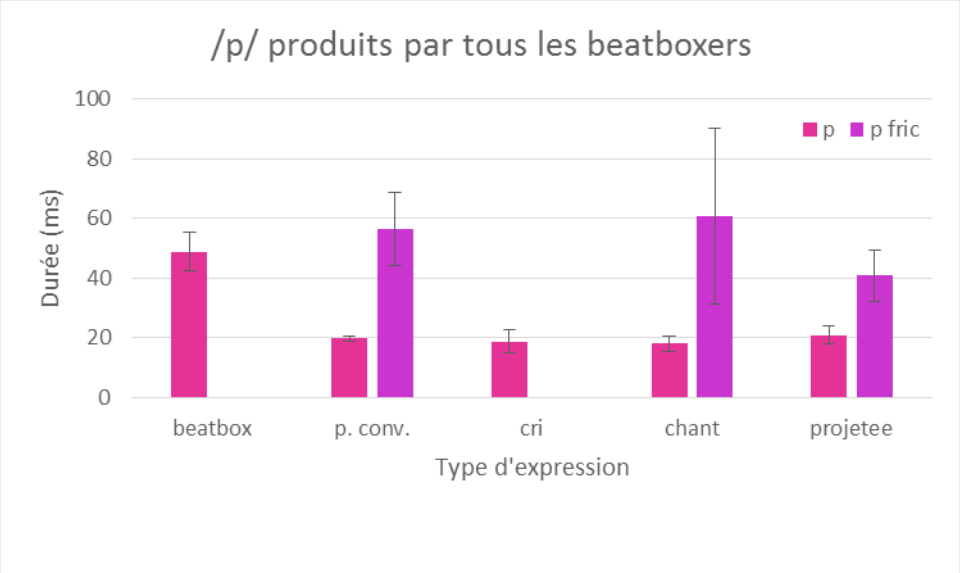

Figure 4.2.6: Durée des réalisations de la cible /p/. « p fric » désigne la réalisation d'une plosive bilabiale parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).

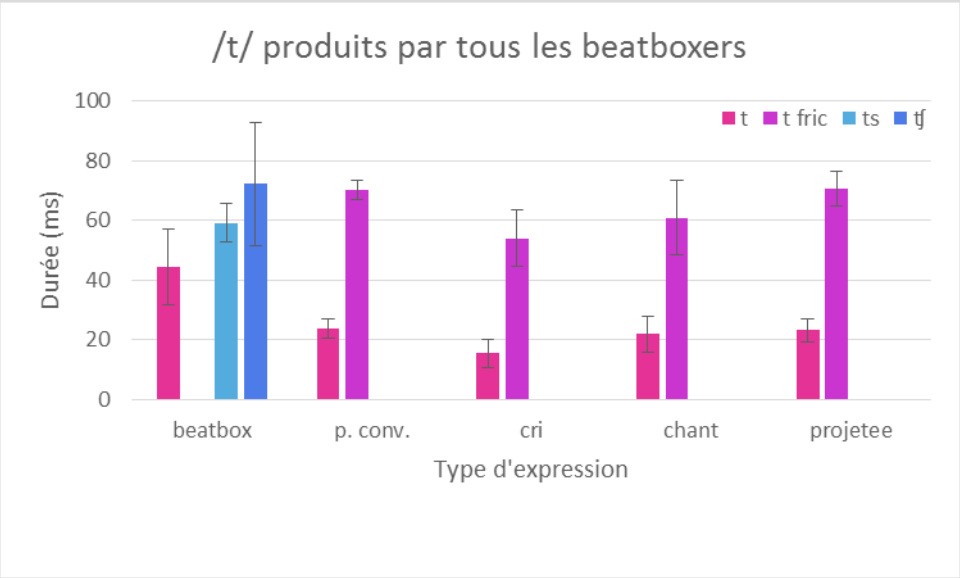

Figure 4.2.7: Durée des réalisations de la cible /t/. « t fric » désigne la réalisation d'une plosive alvéolaire parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).

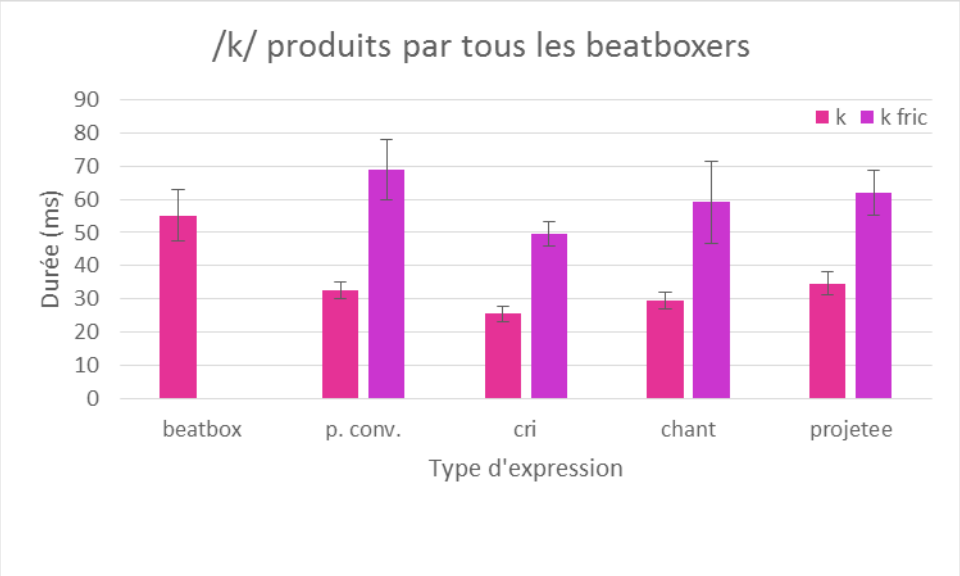

Figure 4.2.8: Durée des réalisations de la cible /k/. « k fric » désigne la réalisation d'une plosive vélaire parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).

Ainsi, le bruit de plosion des consonnes plosives réalisées de façon purement plosive (barres rose en Fig. 4.2.6, 4.2.7, 4.2.8) est d'environ 20 ms pour les consonnes antérieures /p/ et /t/ et 30 ms pour les consonnes postérieures /k/, tandis que les sons plosifs en beatbox montrent des durées presque deux fois supérieures : 40 ms environ pour /p/ et

/t/, 55 ms environ pour /k/. En revanche, les sons plosifs du beatbox ont des durées comparables aux bruits des consonnes plosives réalisées de façon affriquée, pour toutes les autres modalités vocales en ce qui concerne les consonnes /p/ et /k/, seulement pour le cri et le chant en ce qui concerne les consonnes /t/.

Dans les quatre types d'expression autres que le beatbox, le bruit des consonnes plosives est deux à trois fois plus long lorsque la consonne est réalisée avec un bruit de friction prolongé (barres violettes en Fig. 4.2.5, 4.2.6, 4.2.7), comparé à lorsque la consonne est produite de façon purement plosive. En revanche, la durée des sons affriqués en beatbox (barres bleues en Fig. 4.2.6) n'est que légèrement supérieure à celle des sons non affriqués. Leur durée est comparable à celle des bruits de consonnes produites dans les autres modes d'expression avec un bruit de friction allongé.

2.2.2. Intensité de la consonne

Nous avons analysé ici (Fig. 4.2.9) le niveau maximal d'intensité du bruit de la consonne pour chaque cible et chaque mode d'expression.

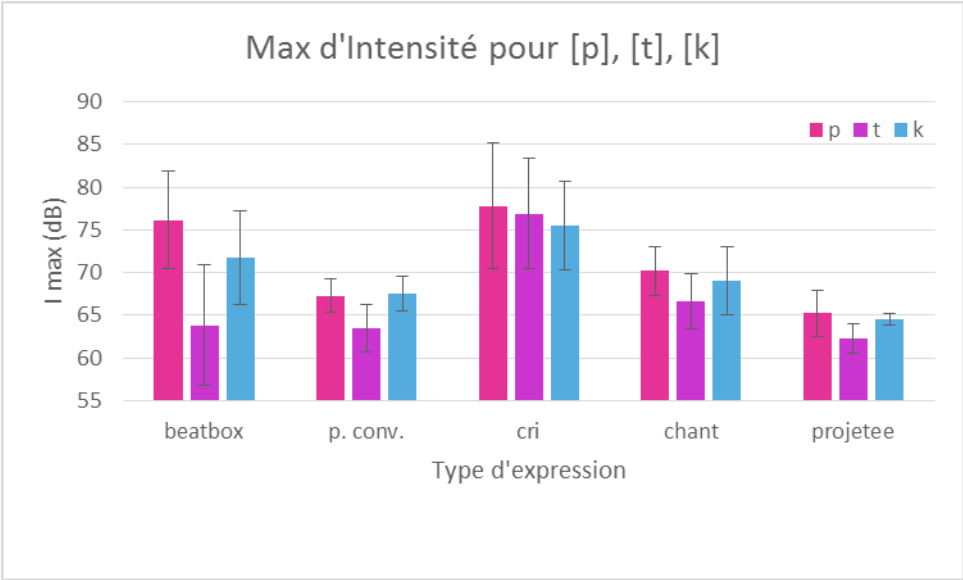

Figure 4.2.9: Max d'intensité pour les réalisations purement plosives des cibles /p/, /t/, /k/.

Contrairement à nos hypothèses, le beatbox n'est pas le mode d'expression pour lequel les bruits de plosion sont produits avec la plus forte intensité : l'intensité des bruits plosifs est maximale en parole criée, d'environ 80 dB pour les réalisations purement plosives des consonnes /p/ et /t/ et d'environ 75 dB pour la consonne /k/. Dans le beatbox

l'intensité maximale des sons plosifs est significativement inférieure au cri : environ 75 dB pour une articulation bilabiale (-2.23 ± 0.53 dB, $p<0.001$), environ 72 dB pour une articulation vélaire (-4.27 ± 0.58 dB, $p<0.001$) et environ 65 dB (-13.86 ± 0.65 dB, $p<0.001$) pour une articulation apico-alvéolaire. Les plosives du chant atteignent des valeurs d'intensité maximale plus faibles que les plosives du beatbox, à l'exception des réalisations des cibles /t/, pour lesquelles les max d'intensité sont légèrement supérieurs pour le chant. Les intensités maximales plus faibles sont produites en modalité de voix parlée et projetée pour toutes les cibles. Concernant la cible /t/, les valeurs d'intensité maximale des réalisations purement plosives sont comparables entre les modalités beatbox, parole et projetée.

Globalement, les consonnes dont le bruit de friction est allongé sont produites avec des intensités plus faibles que les sons purement plosifs en cri, chant, parole et voix projetée, tandis que les sons affriqués du beatbox atteignent les mêmes valeurs d'intensité que les sons purement plosifs. En ce qui concerne l'intensité maximale, cette différence est significative en cri (-3.95 ± 1.17 dB, $p<0.05$) et en chant (-3.22 ± 1.09 dB, $p<0.01$).

De façon générale, les réalisations des cibles /t/ sont moins intenses que les réalisations des autres deux cibles. Les différences d'intensité entre les réalisations plosives de la cible /t/, de la cible /p/ et de la cible /k/ sont supérieures pour le beatbox par rapport aux autres types d'expression et sont moindres pour le cri. Le cri est le seul mode d'expression pour lequel l'intensité des [t] est équivalente à celle des [k]. De manière générale, le paramètre d'intensité maximale distingue les plosives /p/, /t/ et /k/ entre elles en beatbox, tandis qu'en parole ce paramètre permet la discrimination entre /p/ et /k/ versus /t/, mais ne permet pas la discrimination entre /p/ et /k/. Concernant le cri, le paramètre d'intensité ne permet pas de distinguer une plosive des autres.

En beatbox et pour chaque lieu d'articulation, l'intensité maximum du bruit de plosion est significativement corrélée avec sa durée (0.103 ± 0.017, $p<0.001$) (voir Annexe 4 et Fig. 4.2.10 pour un exemple). Cette corrélation n'est pas toujours significative pour les autres modes d'expression.

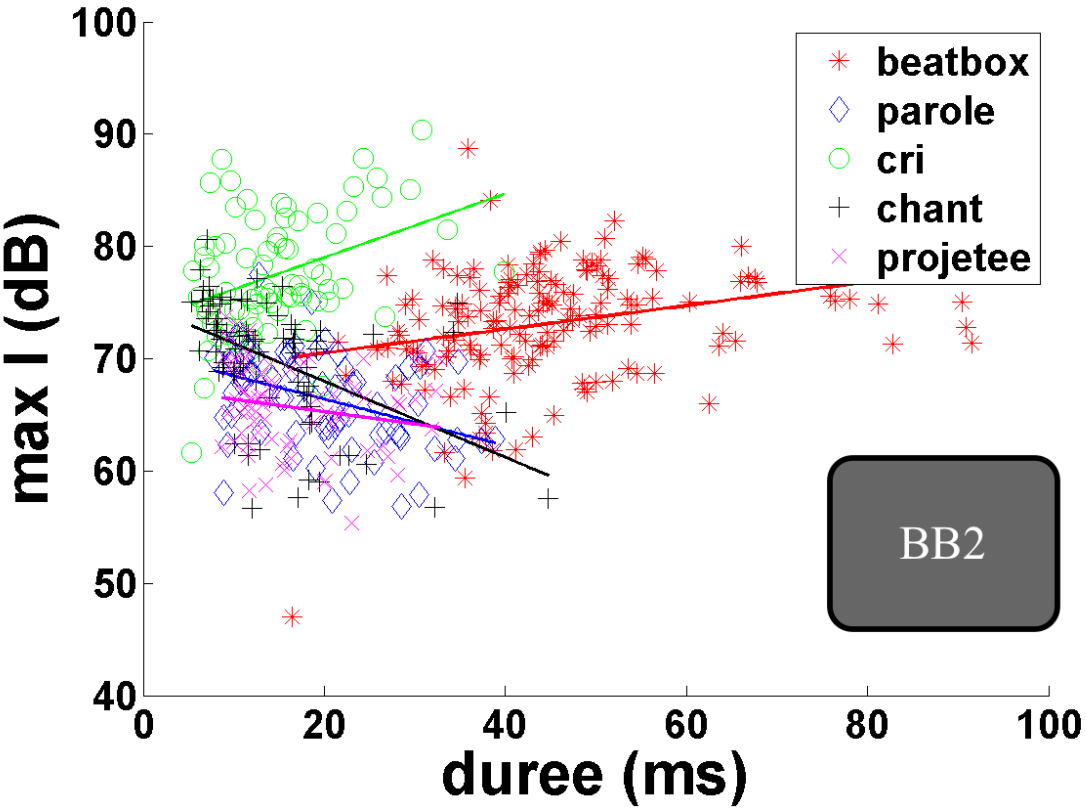

Figure 4.2.10: Corrélation entre l'intensité maximale et la durée du bruit de la cible /p/ dans le cas de BB2.

Contrairement à nos hypothèses, l'intensité maximale des plosives bilabiales ne montre pas de corrélation avec la Pio, sauf que pour le chant (1.03 ± 0.23, $p<0.001$).

En général elle corrèle avec la dérivé temporelle du débit oral (3205.1 ± 454.7, $p<0.001$), mais cette corrélation n'est pas significative en beatbox et en parole criée.

2.2.3. CDG

La Fig. 4.2.11 présente les résultats obtenus concernant le CDG.

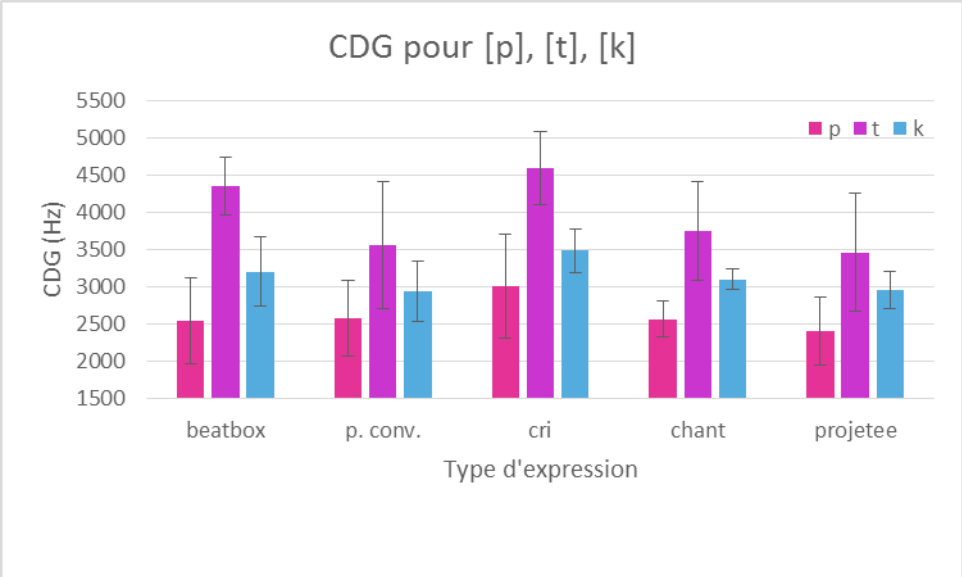

Figure 4.2.11: CDG pour les réalisations purement plosives des cibles /p/, /t/, /k/.

Les plosives [p] montrent le CDG le plus bas pour tous les types d'expression, globalement entre 2500 Hz et 3000 Hz. Cela montre que l'énergie des plosives bilabiales est concentrée dans les basses fréquences. Les plosives [k] ont un CDG moyen, entre 3000 Hz et 3500 Hz. Cela rend compte du fait que l'énergie des plosives vélaires est concentrée à fréquences moyennes. Le CDG des plosives [t] est le plus variable à travers les différents types d'expression : la voix projetée montre le plus bas CDG, avec une valeur légèrement inférieure à 3500 Hz, tandis que la voix criée montre la valeur plus importante, légèrement supérieure à 4500 Hz. Le CDG des plosives alvéolaires montre les valeurs plus élevées, ce qui rend compte du fait que l'énergie est concentrée à plus hautes fréquences pour ces sons. En conclusion, le CDG permet de bien distinguer les trois plosives, c'est-à-dire le lieu d'articulation bilabiale, alvéolaire et vélaire.

Le CDG des plosives bilabiales du beatbox est comparable à celui de la parole conversationnelle, projetée et du chant, alors qu'il est significativement plus faible que celui de la voix criée (-485.02 ± 84.41 Hz, $p<0.01$). En ce qui concerne les plosives vélaires, le CDG tend à être plus élevé en beatbox qu'en voix parlée, chantée et projetée, mais plus faible qu'en voix criée, toutefois sans différence significative. En revanche, le CDG des plosives alvéolaires apparaît comparable entre beatbox et le cri et significativement plus élevé par rapport aux autres modes d'expression (parole

conversationnelle : 871.06 ± 104.73 Hz, $p<0.001$; voix projetée : 892.63 ± 110.03 Hz, $p<0.001$; chant : 676.37 ± 104.04 Hz, $p<0.001$).

Les réalisations affriquées de chaque mode d'expression montrent un CDG significativement plus élevé (1410.87 ± 75.48 Hz, $p<0.001$) que les réalisations purement plosives.

Globalement, le CDG montre une corrélation significative avec la durée de la consonne dans certains cas seulement : pour /p/ en beatbox (8.42 ± 2.70, $p<0.05$), parole conversationnelle (46.59 ± 8.50, $p<0.001$) et voix projetée (32.08 ± 8.23, $p<0.01$) ; /t/ en beatbox (-18.62 ± 2.52, $p<0.001$) et chant (33.04 ± 9.63, $p<0.05$) ; /k/ en modalité beatbox (3.74 ± 0.56, $p<0.001$), chant (24.46 ± 5.25, $p<0.001$), parole conversationnelle (29.57 ± 3.79, $p<0.001$) et projetée (29.45 ± 4.31, $p<0.001$). En général, la pente de cette régression linéaire est significativement plus faible dans le beatbox, comparé aux autres modes d'expression (parole conversationnelle : -29.44 ± 4.57, $p<0.001$; parole projetée : -27.75 ± 5.31, $p<0.001$; chant : -25.69 ± 6.67, $p<0.001$).

Le beatbox (voir Annexe 5 et Fig. 4.2.12 pour un exemple) est le seul type d'expression à montrer pour chaque lieu d'articulation une corrélation significative entre le CDG et l'intensité maximale de la consonne (bilabiale : -47.795 ± 7.086, $p<0.001$; apico-alvéolaire : -34.635 ± 8.122, $p<0.001$; vélaire : -22.164 ± 6.837, $p<0.05$).

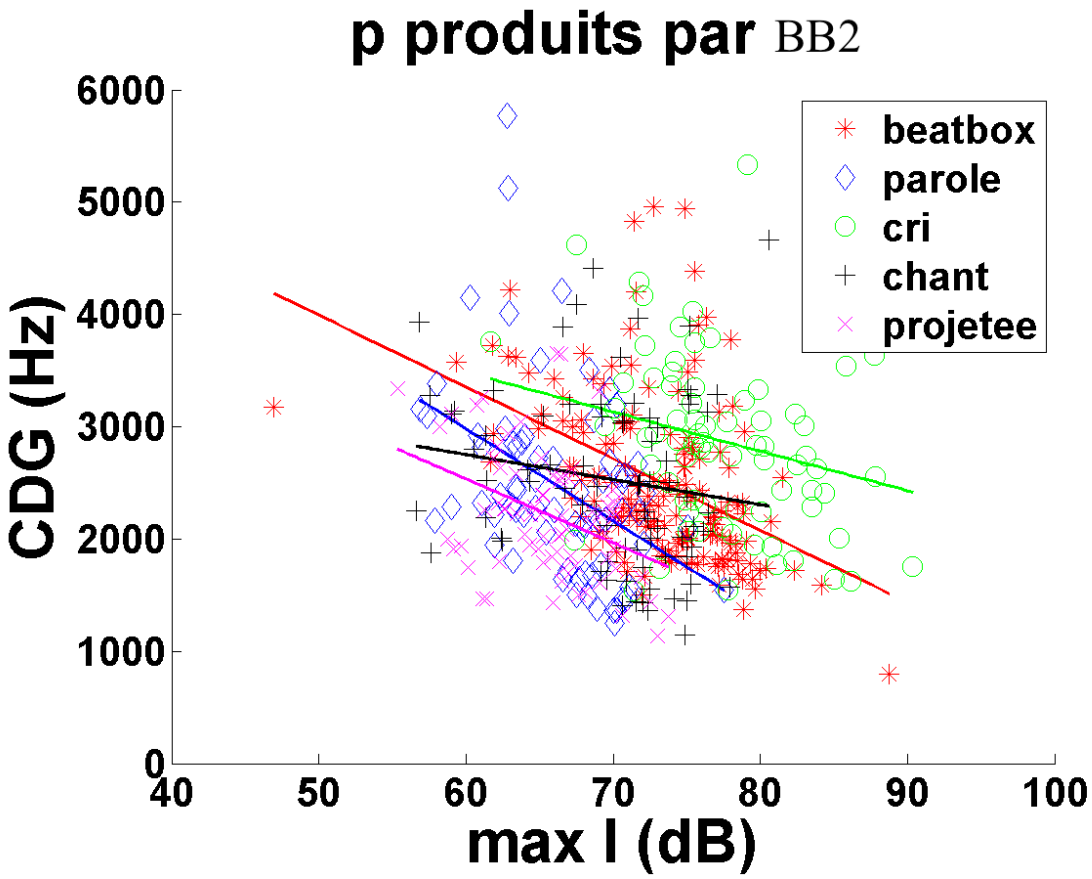

Figure 4.2.12: Corrélation le CDG et l'intensité maximale du son plosif /p/ dans le cas de BB2.

2.2.4. Coefficient d'asymétrie

La Fig. 4.2.13 présente les résultats obtenus concernant le coefficient d'asymétrie.

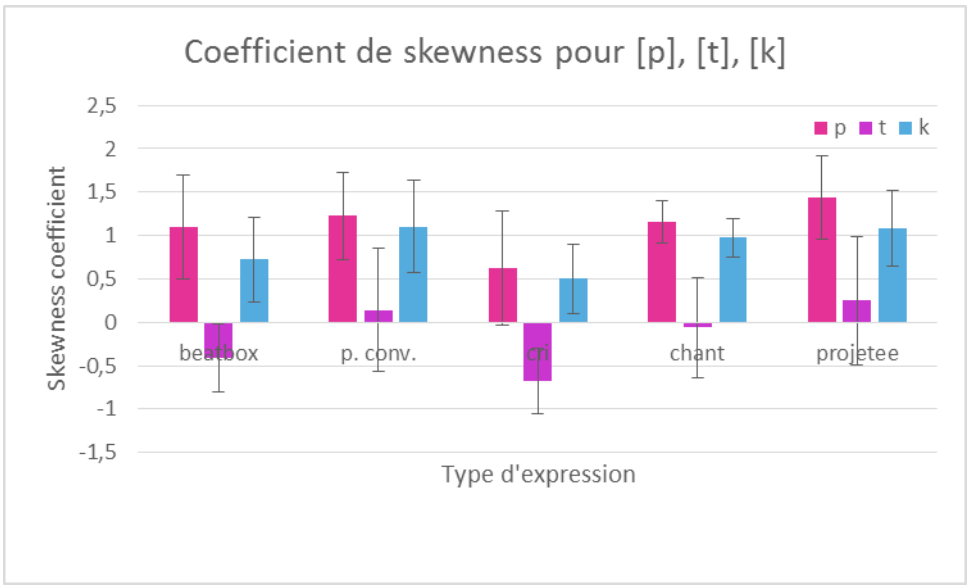

Figure 4.2.13: Coefficient d'asymétrie pour les réalisations purement plosives des cibles /p/, /t/, /k/.

Les plosives bilabiales /p/ montrent les valeurs de coefficient d'asymétrie toujours plus élevées par rapport aux autres lieux d'articulation, quel que soit le mode d'expression. Ce coefficient est comparable entre beatbox, parole conversationnelle et voix projetée, alors qu'il est significativement supérieur dans le beatbox comparé au cri (0.51 ± 0.09, $p<0.001$) et significativement inférieure dans le beatbox comparé au chant (0.39 ± 0.09, $p<0.001$). Le coefficient d'asymétrie des plosives vélaires atteint valeurs moyennes. Le coefficient d'asymétrie des vélaires du beatbox est en moyenne légèrement supérieur à celui du cri et légèrement inférieur à celui des autres types d'expression, mais la différence n'est pas significative. Le coefficient d'asymétrie des plosives alvéolaires est proche de 0 en parole conversationnelle et projetée (positif) et en voix chantée (négatif), tandis que pour le beatbox et le cri ce coefficient est négatif et supérieur, en valeur absolue, à celui des autres modes d'expression. Le coefficient d'asymétrie des alvéolaires du beatbox est significativement inférieur à celui de la voix parlée (-0.64 ± 0.11, $p<0.001$), projetée (-0.67 ± 0.12, $p<0.001$) et chantée (-0.45 ± 0.11, $p<0.01$), alors qu'il est comparable à celui de la voix criée.

Globalement, le cri est le type d'expression qui montre les valeurs les plus basses du coefficient d'asymétrie selon le lieu d'articulation. Le beatbox montre des coefficients d'asymétrie intermédiaires, entre le cri et les autres types d'expression, surtout en ce qui concerne les alvéolaires et les vélaires.

Contrairement à nos hypothèses, les moments spectraux ne montrent pas des corrélations significatives avec la Pio et la vitesse du débit oral, quoi que ce soit le mode d'expression.

2.2.5. Coefficient d'aplatissement

La Fig. 4.2.14 présente les résultats obtenus concernant le coefficient d'aplatissement.

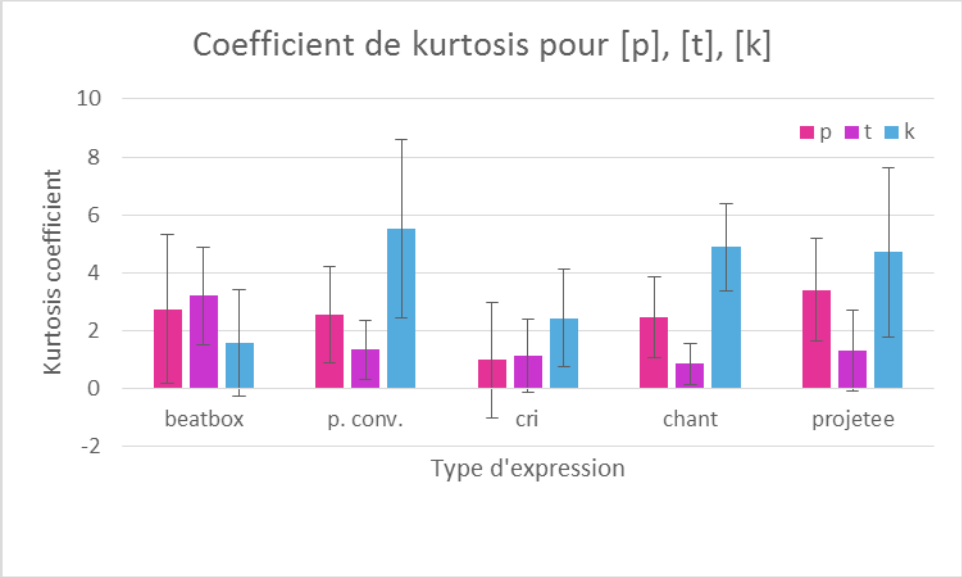

Figure 4.2.14: Coefficient d'aplatissement pour les réalisations purement plosives des cibles /p/, /t/, /k/.

Le coefficient d'aplatissement semble avoir un comportement différent pour les trois catégories de plosives beatboxées par rapport aux autres modes d'expression. [k] montre des valeurs plus faibles par rapport à [p] et [t], ce qui signifie que la distribution d'énergie de la plosive vélaire est plus diffuse que la distribution d'énergie pour les plosives plus antérieures dans le cas du beatbox. En revanche, pour toutes les autres modalités [k] atteint les valeurs les plus élevées, c'est-à-dire sa distribution d'énergie est plus pointue et l'énergie est concentrée autour du CDG.

En général, le coefficient d'aplatissement de la plosive bilabiale en beatbox est comparable au même coefficient en voix parlée, chantée et projetée et significativement supérieur à la voix criée (1.90 ± 0.59, $p<0.01$). Le coefficient d'aplatissement de la plosive alvéolaire est significativement supérieur (1.87 ± 0.39, $p<0.01$) en beatbox par rapport à tous les autres types d'expression (énergie plus concentrée). Le coefficient d'aplatissement de la plosive vélaire est significativement inférieur (2.36 ± 0.32, $p<0.01$) en beatbox par rapport à tous les autres types d'expression (énergie plus diffuse).

Le coefficient d'aplatissement montre une corrélation significative avec l'intensité pour la cible /p/ produite en modalité beatbox (0.177 ± 0.032, $p<0.001$), parole projetée (0.266 ± 0.049, $p<0.001$), et parole conversationnelle (0.159 ± 0.048, $p<0.05$) et chant

(0.133 ± 0.043, p<0.05), alors qu'aucune corrélation significative n'est observée en parole criée.

En conclusion, nos analyses confirment que les trois moments spectraux permettent une bonne discrimination des consonnes /p/, /t/ et /k/ indépendamment du mode d'expression. En revanche, contrairement à nos hypothèses, le beatbox ne se démarque pas nettement des autres modes d'expression, en particulier de la parole conversationnelle, par des valeurs significativement différentes du CDG, du coefficient d'asymétrie et d'aplatissement. Le seul son plosif qui montre en général des différences significatives de ces paramètres avec les autres types d'expression est l'apico-alvéolaire : le CDG et le coefficient d'aplatissement sont plus élevés (CDG : 555.21 ± 72.24 Hz, p<0.001, coefficient d'aplatissement : 1.871 ± 0.393, p<0.001) et le coefficient d'asymétrie est plus bas (-0.391 ± 0.079, p<0.001). En ce qui concerne les différences entre beatbox et parole conversationnelle, le CDG est significativement plus élevé et le coefficient d'asymétrie est significativement plus bas pour l'apico-alvéolaire et le coefficient d'aplatissement est significativement plus élevé pour la vélaire.

2.3. Caractéristiques aérodynamiques des bruits de plosion

2.3.1. Pression intra-orale (Pio)

La Fig. 4.2.15 présente les résultats obtenus concernant la Pio des sons plosifs bilabiaux produits par BB1.

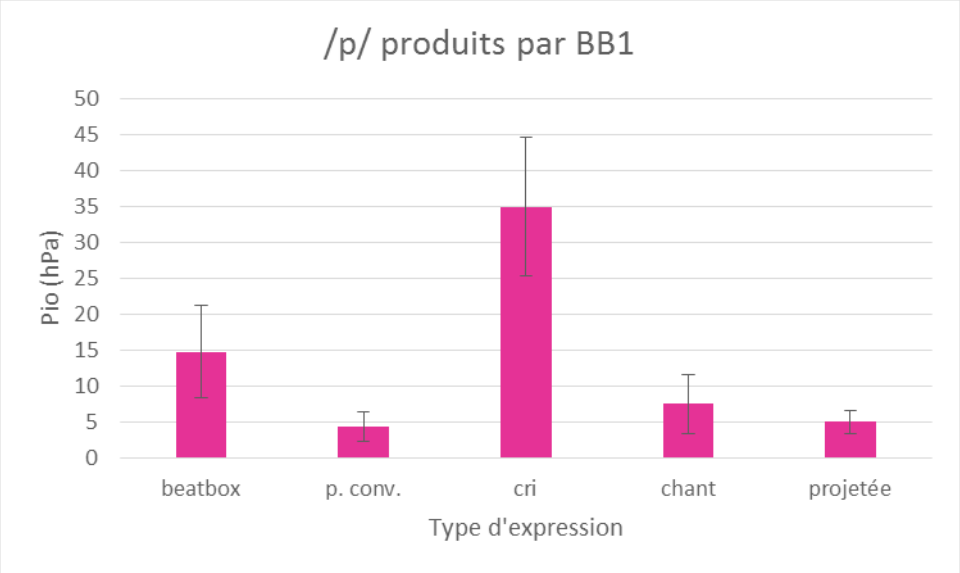

Figure 4.2.15: Pio pour les sons plosifs bilabiaux produits par BB1.

En partiel accord avec nos hypothèses, les sons plosifs bilabiales du beatbox montrent des valeurs de Pio significativement plus élevées qu'en parole conversationnelle (5.38 ± 0.53 cmH_2O, $p<0.001$), en parole projetée (5.66 ± 0.55 cmH_2O, $p<0.001$) et en chant (5.07 ± 0.54 cmH_2O, $p<0.001$). En revanche, contrairement à nos hypothèses, ces valeurs sont significativement inférieurs en beatbox qu'en parole criée (-2.91 ± 0.54 cmH_2O, $p<0.001$). En fait, les valeurs de Pio du cri sont environ deux fois supérieures aux valeurs de Pio du beatbox.

En ce qui concerne la parole, en accord avec notre hypothèse, les valeurs de Pio les plus élevées sont celles du cri, plus grandes que les valeurs en parole conversationnelle et projetée. En revanche, la voix projetée ne montre pas des valeurs de Pio plus élevées que celles de la parole conversationnelle.

La Pio montre une corrélation significative avec l'intensité seulement pour le chant (1.03 ± 0.23, $p<0.001$) (Fig. 4.2.16), tandis que elle n'a pas d'effet significatif sur les caractéristiques spectrales du bruit de plosion.

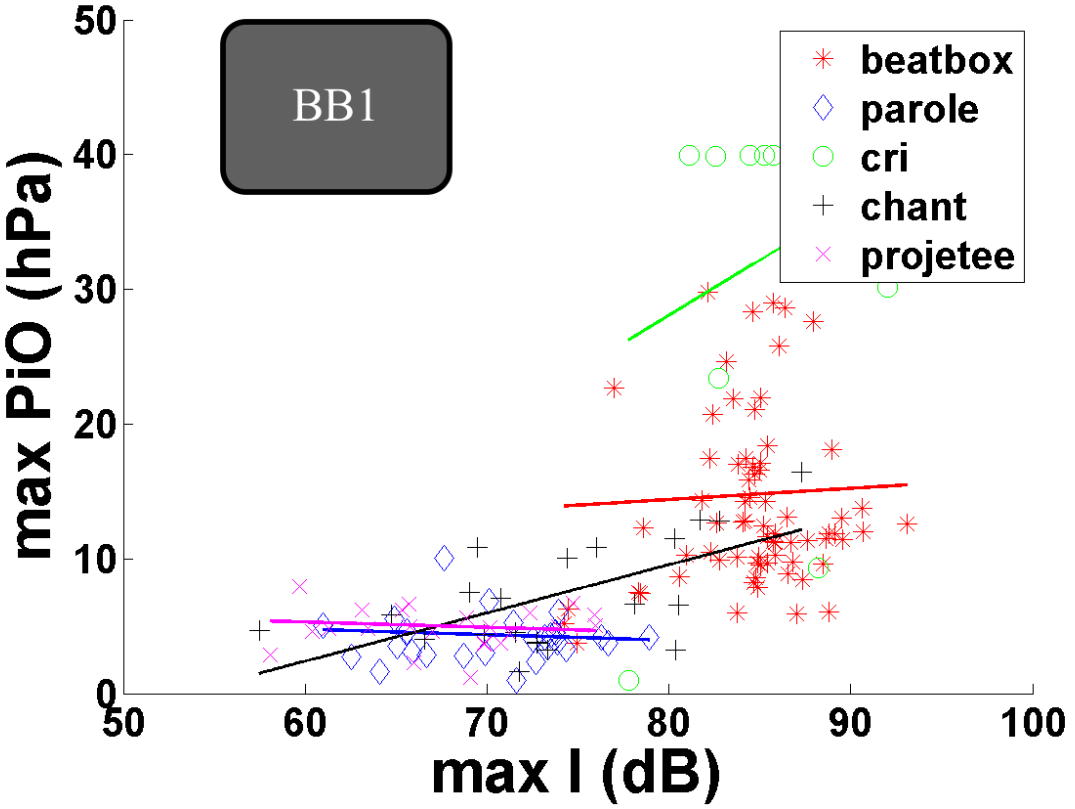

Figure 4.2.16: Corrélation entre la Pio et l'intensité maximale du bruit de la cible /p/ dans le cas de BB1.

2.3.2. Vitesse du débit d'air oral

La Fig. 4.2.17 présente les résultats obtenus concernant la vitesse du débit d'air oral.

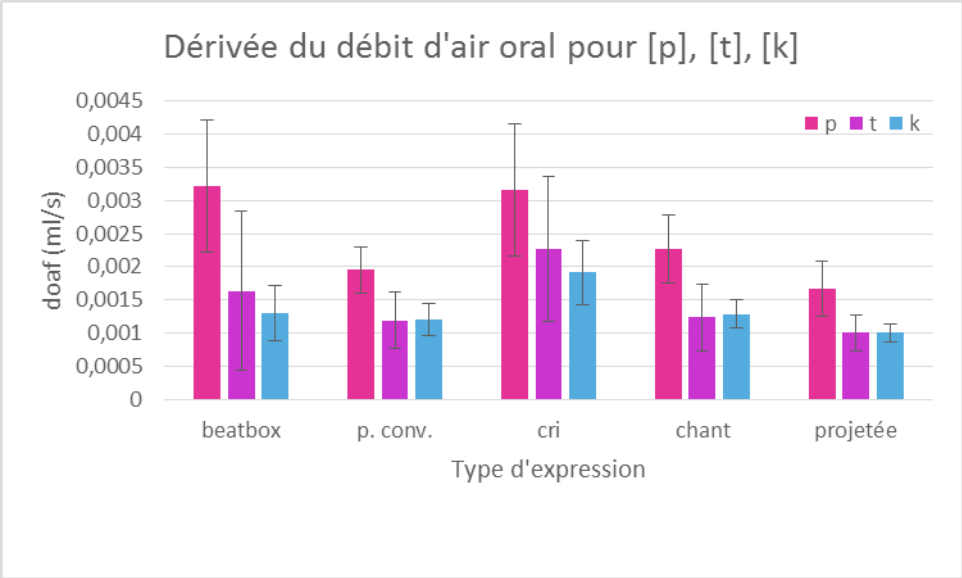

Figure 4.2.17: Dérivée temporelle du débit d'air oral pour les réalisations purement plosives des cibles /p/, /t/, /k/.

Partiellement en accord avec nos hypothèses, le beatbox montre les valeurs les plus élevées de vitesse du débit d'air oral pour la cible /p/, significativement supérieurs aux autres modes d'expression ($7.475e^{-04} \pm 0.520e^{-04}$ ml/s, $p<0.001$) . Globalement, le beatbox montre des valeurs de vitesse du débit oral significativement supérieures aux autres modes d'expression (parole conversationnelle : $5.261e^{-04} \pm 0.490e^{-04}$ ml/s, $p<0.001$; parole projetée : $7.235e^{-04} \pm 0.509e^{-04}$ ml/s, $p<0,001$; chant : $3.697e^{-04} \pm 0.490e^{-04}$ ml/s, $p<0.001$), sauf que pour la parole criée ($-5.986e^{-04} \pm 0.493e^{-04}$ ml/s, $p<0.001$).

La vitesse du débit d'air oral corrèle significativement avec l'intensité du bruit de plosion dans tous les modes d'expression sauf le beatbox et la parole criée (chant: 3389.0 ± 937.9 ; $p<0.01$; parole conversationnelle : 3808.0 ± 1021.6, $p<0.01$; parole projetée 6349.2 ±1 445.6, $p< 0.001$) (Fig 4.2.18).

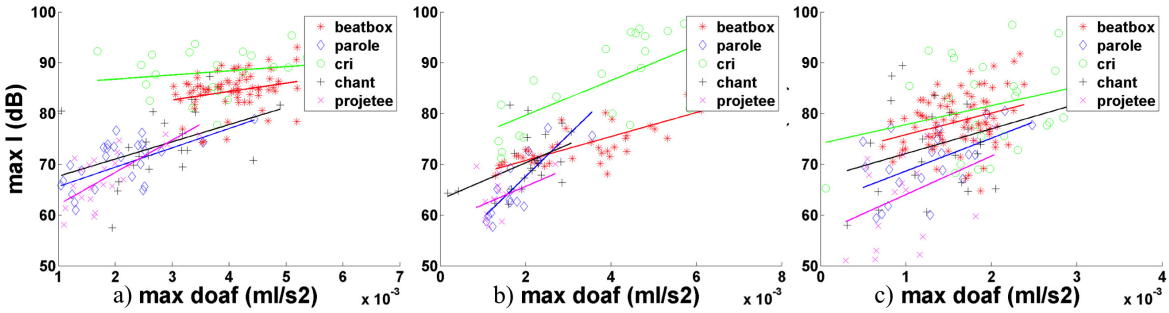

Figure 4.2.18: Corrélation entre l'intensité maximale et la vitesse du débit d'air oral pour la cible /p/ (a), /t/ (b) et /k/ (c) produites par BB1.

La vitesse du débit d'air oral corrèle significativement avec le coefficient d'aplatissement pour le chant seulement (6349.2 ± 1445.6, $p < 0.001$) (Fig. 4.2.19).

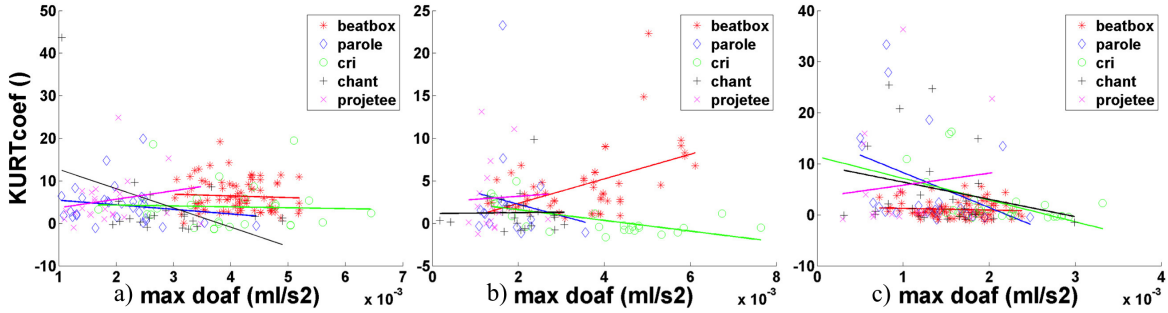

Figure 4.2.19: Corrélation entre le coefficient d'aplatissement et l'intensité maximale pour la cible /p/ (a), /t/ (b) et /k/ (c) produites par BB1.

Ce paramètre ne montre pas de corrélation significative avec les autres caractéristiques spectrales du bruit de plosion (CDG et coefficient d'asymétrie).

3. Discussion

Nous allons discuter les résultats obtenus en cherchant à les comparer aux données de la littérature.

3.1. H1 : différences acoustiques, aérodynamiques et articulatoires

Nous avons émis l'hypothèse que les réalisations du beatbox montreraient des différences acoustiques, aérodynamiques et articulatoires par rapport aux réalisations de la parole.

Le résultats que nous avons obtenus confirment partiellement cette hypothèse.

La réalisation du protocole est plus variée dans le beatbox que dans les trois types de parole et le chant, concernant les sons plosifs, le mode d'expression et la structure des phrases.

En comparant le beatbox à tous les autres formes d'expression, nous remarquons une variabilité plus importante en ce qui concerne les réalisations d'une même cible, notamment par rapport aux cibles /t/ et /p/. Les réalisations de la cible /t/ sont triples, [t] [ts] [ʧ]. Ces trois sons reproduisent trois effets différents en beatbox, à savoir le charleston fermé, le charleston ouvert et le clash. Cette observation de nature acoustique est en accord avec les observations articulatoires de Proctor et collègues (2013) et de notre étude précédente (Paroni, 2014). Les deux études ont montré que les beatboxeurs exploitent des subtiles différences articulatoires, et, par conséquent, acoustiques et aerodynamiques, pour différencier des effets proches entre eux et ainsi élargir leur répertoire. Concernant la cible /p/, la variabilité des réalisations est moindre par rapport à la cible /t/ : parfois une trille bilabiale accompagne ou remplace la plosive bilabiale. Nous retrouvons ici un résultat obtenu lors de notre exploration articulatoire (Paroni, 2014). Cette différenciation n'apparaît pas de façon uniforme chez un même beatboxer, du moins chez nos sujets. Cela peut indiquer que le trille ne soit pas complètement volontaire, mais qu'au contraire il puisse résulter d'une force articulatoire plus grande au niveau des lèvres, d'une Pio augmentée et/ou d'un débit d'air oral plus important. Ces hypothèses aérodynamiques restent à vérifier.

Nous observons une proportion non négligeable de réalisations affriquées des consonnes parlées et chantées suivies d'une voyelle antérieure fermée. Cela est valable pour chaque lieu d'articulation, en proportion différente. En revanche, dans les sons de beatbox les bruits restent purement plosifs pour les lieux d'articulation bilabiale et vélaire,

ou alors ils changent de catégorie ([t] [ts] [ʧ]) et d'effet (charleston fermé, charleston ouvert, clash) pour le lieu apico-alvéolaire.

En beatbox, nous avons observé des difficultés à se détacher d'une modalité expressive parlée : dans de nombreuse cas, lorsque des voyelles chuchotées suivent les sons plosifs. Cette difficulté, ne dépend pas du niveau de maîtrise du beatbox, est induite par le protocole (voir Annexe 1). En effet, le sujet doit produire la phrase beatboxée tout de suite après l'avoir produite en plusieurs modes de parole et chant.

La structure de la phrase beatboxée n'est pas toujours respectée, alors qu'elle est toujours reproduite correctement en parole et chant. En général, cela semble dépendre du niveau de maîtrise du beatbox et du niveau de familiarisation avec la SBN, le langage utilisé pour l'écriture des phrases beatboxées.

Le beatbox se démarque par rapport aux autres modes d'expression principalement par une durée significativement plus importante (environ deux fois supérieure) des bruits de plosion. Cette durée est comparable aux plosives affriquées parlées et chantées. Une durée supérieure pourrait provenir d'une définition différente de la segmentation des bruits de plosion dû au fait que les sons plosifs en mode beatbox sont produits isolés, alors qu'en modes parlé et chanté ils sont enchaînés avec des voyelles. Toutefois, dans les cas où le son plosif est suivi par une voyelle chuchotée, la durée de la plosive ne semble pas être différente par rapport au cas où le son plosif beatboxé est isolé. Par ailleurs, Proctor et al. (2013) remarquent le même phénomène, à savoir la durée de certains sons plosifs du beatbox est supérieure à la durée du VOT de leurs contreparties langagières. En comparant nos valeurs de durée avec les valeurs trouvées par Garrigues (2015), nous remarquons des différences importantes surtout concernant les plosives alvéolaires : nous retrouvons des valeurs plus grandes. Cela peut être expliqué par le fait que Garrigues a mesuré seulement la durée du bruit de plosion, alors que nous avons mesuré l'ensemble du bruit de plosion et du bruit de friction.

Conformément à nos attentes, le beatbox se démarque de la parole conversationnelle par des valeurs significativement plus importantes en ce qui concerne l'intensité du « burst » et la Pio. Cependant, cette caractéristique n'est pas une spécificité du beatbox en soi. En fait, à effort vocal comparable, la parole criée présente également des valeurs d'intensité des bruits de plosion et de Pio très élevées, voir plus élevées que le beatbox. En revanche, en ce qui concerne la plosive bilabiale, le beatbox est le mode d'expression qui montre les valeurs les plus élevées de dérivée du débit oral, significativement supérieures aux autre types d'expression. Une spécificité ultérieure du

beatbox est que celui-ci est le seul mode d'expression où le paramètre intensité distingue touts les trois lieux d'articulation entre eux.

Contrairement à nos attentes, les caractéristiques spectrales des sons plosifs du beatbox ne sont pas fondamentalement différentes dans le beatbox comparé à la parole conversationnelle et aux autres modes d'expression. Lorsque les moments spectraux du beatbox sont significativement différents de la parole conversationnelle, ils sont comparables au moments spectraux du cri et, lorsqu'ils sont significativement différents du cri, ils sont comparables au moments spectraux de la parole conversationnelle.

3.2. H2 : corrélations concernant l'intensité du « burst »

Nous avons émis l'hypothèse que l'intensité du « burst » serait corrélée à la Pio et à la dérivée temporelle du débit d'air oral.

Nos résultats ne confirment pas cette hypothèse.

Nous rappelons ici que nous avons pu utiliser les données d'un seul sujet afin de répondre aux questions concernantes la Pio.

L'intensité du bruit de plosion ne montre aucune corrélation avec la Pio, excepté pour le chant.

La corrélation entre l'intensité et la dérivée du débit oral n'est significative qu'en certains cas (parole conversationnelle, parole projetée et chant) et les pentes de corrélation linéaire sont variables en fonction du mode d'expression.

3.3. H3 : corrélations concernant le spectre du « burst »

Nous avons émis l'hypothèse que le spectre du bruit de plosion serait influencé principalement par le lieu d'articulation, mais aussi par les caractéristiques aérodynamiques de la production (la Pio et la dérivée du débit oral).

Nos résultats ne confirment que partiellement cette hypothèse.

Les moments spectraux sont influencés par le lieux d'articulation, spécialement en beatbox. Cependant, les moments spectraux ne semblent pas être influencés par la Pio et/ou la dérivée du débit d'air oral, indépendamment du mode d'expression.

En revanche, le CDG des sons plosifs semble être influencé par la durée du son et par l'intensité du bruit. Le beatbox est le seul mode d'expression où cela est significatif pour chaque lieux d'articulation.

Enfin, les questions concernants l'articulation des sons plosifs seront explorées à l'aide du corpus articulatoire EMA (voir Chapitre 6) dans une prochaine étude.

4. Conclusions

Nous avons cherché à analyser et comparer certains caractéristiques acoustiques de trois sons plosifs non voisés produits en mode beatbox avec les consonnes correspondantes produites en mode parole, cri, chant et voix projetée.

Tout d'abord, en comparant la production du human beatbox aux autres formes d'expression, la variété de sons produits est plus ample pour une même consonne cible, surtout pour la plosive alvéolaire. Ces sons sont tous des effets qui appartiennent à la production artistique du human beatbox.

De plus, en accord avec la littérature, la durée des consonnes plosives beatboxée est significativement plus grande que leurs contreparties parlées et comparable à la durée des plosives parlées et chantées dont le bruit de friction est allongé (affriquées).

En ce qui concerne l'intensité et la Pio, les plosives du beatbox atteignent des valeurs supérieures par rapport à la parole conversationnelle, toutefois cela n'est pas une spécificité du beatbox. En revanche, la dérivée du débit oral est plus importante en beatbox qu'en parole et chant pour la plosive bilabiale.

Les moments spectraux sont influencés par le lieu d'articulation et non pas par les paramètres aérodynamiques. En revanche, le CDG en beatbox est influencé par la durée et l'intensité du son.

En conclusion, les plosives du human beatbox montrent des valeurs des paramètres acoustiques qui se situent généralement entre celles du cri et celles des autres modes d'expression.

Partie 3 – Perspectives de recherche

Chapitre 6. Étude articulatoire

1. Problématiques et hypothèses

1.1. Problématiques générales

Une difficulté majeure de l'étude articulatoire du human beatbox porte sur la mesure des mouvements articulatoires.

Les travaux menés jusqu'à présent dans ce domaine ont utilisés trois approches expérimentales, à savoir l'Imagerie par Résonance Magnétique ou IRM (Proctor *et al.*, 2013), l'imagerie par ultrasons (Paroni, 2014) et la vidéo-fibroscopie (Clouet & de Torcy, 2010 ; de Torcy *et al.*, 2013 ; Septhavee *et al.*, 2014), chacune présentant des avantages et des inconvénients.

Tout d'abord, en ce qui concerne la fréquence d'échantillonnage, l'IRM permet d'atteindre 24 images par second (24 fps), ce qui permet une analyse visuelle satisfaisante de la plupart des gestes articulatoires du beatbox. Dans notre étude précédente (Paroni, 2014), nous avons utilisé la technologie des ultrasons qui nous a permis d'obtenir une fréquence d'échantillonnage de 60 fps. Toutefois, cette résolution temporelle n'a pas été suffisante pour décrire les détails de certains gestes articulatoires du beatbox, telles que des roulements de langue.

Quant aux structures visualisées, la technique de visualisation articulatoire par IRM permet d'obtenir des images qui montrent ce qui se passe au niveau du plan médio-sagittal du conduit vocal entier (comprenant les niveaux laryngé, pharyngé, le palais, ainsi que la langue). En fibroscopie, les images obtenues ne montrent que les parties laryngée et pharyngée du conduit vocal, mais avec l'avantage d'une vue presque tridimensionnelle.

Pendant le recueil des données en IRM le sujet doit rester allongé dans la machine IRM et produire les sons dans une position inhabituelle. Cela a des conséquences sur le mouvement de la langue qu'il faut prendre en compte et également sur le confort du chanteur. De plus, la machine IRM produit un bruit de forte intensité, ce qui détériore la qualité des enregistrements audio et peut perturber le sujet dans sa production par effet Lombard. Avec la fibroscopie, le fibroscope est introduit par les cavités nasales et positionné le long de la paroi pharyngée du sujet, ce qui peut gêner la production du

beatboxer. La technique de visualisation par ultrason est moins invasive. Le beatboxer peut rester assis sur une chaise pendant le recueil des données dans une chambre sourde. La sonde échographique doit rester attachée sous le menton du sujet. Pour faire en sorte que le sujet n'ait pas à maintenir la sonde tout le temps et pour avoir une référence fixe pour le palais (structure invisible aux ultrasons), la sonde est attachée de façon rigide à un casque porté par le sujet. Dans notre précédent travail (Paroni, 2014), nous avons suivi cette procédure (qui est la procédure que l'on emploie pour l'étude articulatoire des sons langagiers), mais le casque rigide gênait la production de notre sujet, puisque celui-ci ne pouvait pas abaisser sa mandibule. En effet, dans le beatbox les mouvements de la mandibule et d'autres articulateurs sont beaucoup plus larges que dans la parole. De ce fait, nous avons dû utiliser un casque semi-rigide et nous avons de fait perdu la possibilité de référencer les informations à la position du palais.

Comme nous venons de le montrer, l'étude articulatoire du human beatbox est complexe. Cette branche de recherche est nouvelle, peu de travaux explorent les différentes techniques expérimentales de recueil de données, les avantages et inconvénients de chaque technique, leur faisabilité par rapport à la production du human beatbox et par rapport au coût et aux risques pour la santé du sujet (e.g. par l'emploi de rayonnements ionisants) et les types de données que ces différentes techniques permettent de récolter.

1.2. Problématique spécifique et hypothèses

Une autre technique qui semblerait prometteuse pour l'étude articulatoire du human beatbox est l'Articulographie ÉlectroMagnétique (EMA). Cette méthode, très utilisée pour l'étude articulatoire des sons langagiers, permet d'aborder l'étude de la dynamique de la langue, car elle mesure la position et la vitesse de mouvement de certains points de chair sur la surface de la langue, grâce à des bobines électro-magnétiques collées sur la langue et plongées dans un champ magnétique.

A ce jour, à notre connaissance aucune étude n'a encore exploité cette technique pour explorer la dynamique et les mécanismes articulatoires de la production du human beatbox. Effectivement, plusieurs problèmes peuvent se présenter en utilisant l'EMA pour recueillir des données de ce type de production vocale. Tout d'abord, les forces musculaires impliquées dans l'articulation des sons du beatbox sont beaucoup plus importantes que celles que l'on trouve dans l'articulation des sons de la parole et la langue bouge plus rapidement en accomplissant des mouvement plus amples. Par conséquent, il est

envisageable qu'il soit difficile de garder les bobines collées à la langue pendant la production vocale du beatboxer, d'autant plus que cela arrive parfois même en parole.

En plus, l'EMA est une technique plutôt invasive : plusieurs bobines sont collées sur la langue et sur les dents du sujet et chaque bobine est connectée au dispositif par un fil qui sort de la bouche. Tout ce matériel dans la bouche peut affecter la réalisation des tâches articulatoires demandées, surtout en ce qui concerne le human beatbox pour lequel sont requis des mouvements précis et fins.

En conclusion, nous émettons les deux hypothèses suivantes :

I. en raison des forces articulatoires en jeu, les bobines collées à la langue vont se décoller en cours d'expérience, ne permettant pas de parcourir tout le protocole expérimental ;

II. le matériel présent dans la bouche (notamment les bobines et les fils) peut gêner le beatboxer dans sa production, jusqu'à l'empêcher d'accomplir les tâches demandées.

Afin de tester ces deux hypothèses, nous avons conduit une étude de faisabilité pour recueillir des données articulatoires de la production d'un beatboxer amateur par le biais de la technologie EMA.

2. Matériel et méthodes

Nous présentons ci-dessous le sujet qui a participé à notre étude, les corpus que nous avons constitué et la procédure d'enregistrement que nous avons suivi.

2.1. Sujet

Le sujet est un homme de 28 ans, ex-fumeur, qui pratique le beatbox en tant qu'amateur depuis 9 ans. Il donne des concerts de façon occasionnelle et n'a jamais participé à des compétitions de human beatbox. Il s'entraine tous les jours entre 15 minutes et 1 heure. Il maîtrise plusieurs variantes de beatbox, les principales étant le Humming Beatbox et le Power Beatbox.

Au tout début de son apprentissage du human beatbox, il ressentait souvent de la fatigue vocale et de la gêne. Il a donc arrêté la pratique du beatbox pendant un an, au cours duquel il a appris la respiration diaphragmatique, qui l'a beaucoup aidé dans la pratique de cet art vocal.

2.2. Corpus et protocole

2.2.1. Corpus

Le corpus a été conçu afin de répondre aux questionnements suivants :

1. l'efficacité vocale observée dans la production vocale du human beatbox dérive-t-elle de la rapidité d'articulation des gestes de cet art et notamment de la rapidité de création de l'occlusion pour les sons plosifs ?
2. Comment la variabilité des sons produit en beatbox correspondant à une même cible est-elle réalisée d'un point de vue articulatoire ?

Le corpus est réparti en quatre parties, chacune correspondant à un questionnement expérimental.

La première partie est également répartie en deux sous-parties, dont la première explore les conséquences de la variation d'intensité sur l'articulation de cinq effets de beatbox, produits en 15 répétitions chacun, en suivant le rythme imposé par un métronime :

1. grosse caisse (« kick »), dans les variantes « humming » et « power » ;
2. charleston (« hi-hat »), dans les variantes « humming » et « power », ouvert et fermé ;
3. caisse claire, dans les variantes « humming », « power » et « power » inverse ;
4. « rimshot », dans les variantes « humming » et « power » ;
5. cymbale, dans un enchaînement (alternance de sons expirés et inspirés).

La deuxième sous-partie explore les conséquences de la variation de timbre sur l'articulation de trois des effets précédents, à savoir la caisse claire inspirée, le « rimshot » et le charleston, dans un enchaînement. Ces effets sont produits en 5 répétitions contenant chacune 10 variations de timbre, en suivant le rythme imposé par un métronome.

La deuxième partie du protocole explore les différences articulatoires entre beatbox, parole, cri et chuchotement sur un continuum d'intensité. Le sujet produit les syllabes /pu, pi, pa/, /tu, ti, ta/ et /ku, ki, ka/. En parole, chaque syllabe est produite sur 10 niveaux d'intensité, de faible à fort et de fort à faible. En ce qui concerne le chuchotement, les syllabes sont produites en modalité sans effort et en modalité avec effort. En beatbox, elles sont produites en modalité normale, faible et forte. Pour chaque condition, 5 répétitions sont produites, tout en suivant le rythme d'un métronome.

La troisième partie du protocole explore le continuum entre la production en modalité parlée et la production en beatbox. Des phrases sont répétées à partir d'une modalité parlée, jusqu'à une modalité beatboxée.

La quatrième partie du protocole est consacrée à l'expression libre du beatboxer.

2.2.2. Procédure d'enregistrement et dispositifs

Les enregistrements ont eu lieu dans la chambre sourde du laboratoire GIPSA-lab de Grenoble.

Une discussion avec le sujet en préalable de l'expérience a permis de recueillir des informations biographiques et le détail de sa pratique du human beatbox. Les effets appartenant au répertoire du sujet ont été également discutées et le protocole d'enregistrement a été adapté en conséquence.

En suite, après la signature du consentement informé, le sujet a été positionné dans la chambre sourde (Fig. 4.2.20), il a été habillé avec le gilet pour l'enregistrement des données respiratoires (VisuResp), et fait asseoir sur une chaise adaptée pour stabiliser la tête et empêcher que celle-ci puisse sortir du champ magnétique de l'EMA.

Figure 4.2.20: Positionnement du sujet à l'intérieur du champ magnétique de l'EMA

Les bobines de l'EMA ont été donc positionnées comme ceci (Fig. 4.2.21) :

- 3 bobines sur la langue, 1 dans la zone apicale, 1 dans la zone dorso-palatale et 1 dans la zone dorso-vélaire ;
- 1 bobine sur la machoîre ;
- 4 bobines sur les lèvres, 2 sur la lèvre supérieure et 2 sur la lèvre inférieure ;
- 2 bobines de référence, 1 sur l'incisive supérieure et 1 sur l'incisive inférieure ;
- 2 bobines de référence, 1 derrière l'oreille droite, 1 derrière l'oreille gauche ;
- 1 bobine de référence sur le front.

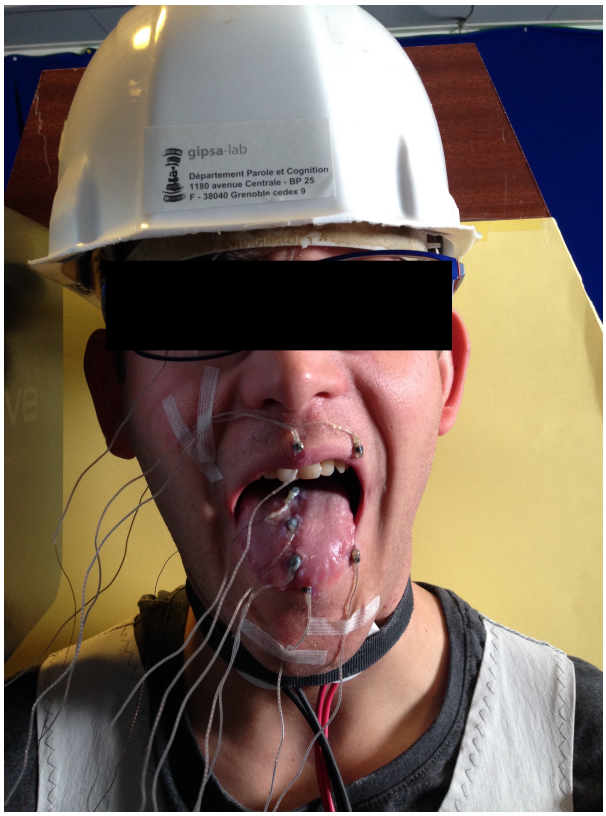

Figure 4.2.21: Positionnement des bobines de l'EMA.

Enfin, l'EGG a été positionné sur le cou du sujet au niveau du larynx. Un microphone AKG et un microphone BXR ont été positionnés à 23 cm de la bouche du sujet respectivement pour mesurer le signal audio et l'intensité. Une caméra a été placée en face du sujet pour les enregistrements vidéo. À côté de la caméra, était placé un écran sur lequel étaient affichées les représentations des points de chair.

Une fois terminé le positionnement, les dispositifs ont été branchés et les enregistrements ont commencé (Fig. 4.2.22).

Figure 4.2.22: Les conditions d'expérience à l'intérieure de la chambre sourde.

Un expérimentateur était dans la chambre sourde, assis à côté du sujet et lui annonçait à chaque fois la tâche à produire. Deux expérimentateurs étaient dans la chambre de contrôle, l'un s'occupant des enregistrements EMA, l'autre des autres signaux. Au début de chaque enregistrement, un signal marqueur était lancé et capté à la fois par l'EMA et par les autres signaux, afin de pouvoir synchroniser toutes les données en post-traitement. Suite à ce signal, l'expérimentateur à l'intérieur de la chambre sourde recevait via interphone la communication de début d'enregistrement et énonçait la tâche au sujet. Avant la fin de chaque enregistrement, un deuxième marqueur était lancé pour la synchronisation.

3. Résultats

Concernent l'hypothèse I : nous avons pu enregistrer les parties de 1 à 3 du protocole, sans que les bobines se détachent. En effet, toutes les bobines étaient bien collées jusqu'à la fin des enregistrements.

Concernant l'hypothèse II : au début de l'expérience, le sujets a affirmé ressentir de la gêne à cause des bobines et ne pas pouvoir produire les sons beatboxés requis. Après une brève période d'adaptation, la gêne a presque complètement disparu. Toutefois, le sujet a affirmé que les sons n'étaient pas totalement naturels. La quatrième partie du protocole (expression libre) n'a pas pu être enregistrée, car le matériel dans la bouche du beatboxer empêchait une telle production.

4. Discussion et conclusions

La colle utilisée pour cette expérience est une colle orthodontique spéciale et très efficace. Nos enregistrements ont duré environ une heure et, à la fin, les bobines étaient encore toutes bien collées. Donc, en choisissant la colle adaptée, il est possible de mener une expérience EMA afin d'étudier la production de sons plosifs (entre autres) produits en modalité beatboxée, où les forces articulatoires sont supérieures à celles de la parole. Notre hypothèse I est donc invalidée.

Les bobines et les fils causent de la gêne au beatboxer, surtout dans un premier temps. Cela est tout-à-fait normale et de telles situations se produisent aussi en situation d'étude de la parole. Cependant, cette gêne n'empêche pas le beatboxer de produire de sons isolés ou bien des simples enchaînements. En revanche, une production libre et naturelle de « beats » n'apparaît pas possible. En conclusion, notre hypothèse II est partiellement vérifiée : le matériel dans la bouche cause effectivement de la gêne dans une certaine mesure, toutefois il n'empêche pas complètement la production de sons de beatbox.

En conclusion, la technique de récolte de données par EMA s'avère prometteuse et faisable pour l'étude de la production de sons plosifs produit en mode human beatbox.

Bibliographie

Andrade, D.H., Heuer, R., Hockstein, N.E., Castro, E., Spiegel, J.R. & Sataloff, R.T. (2000). The frequency of hard glottal attacks in patients with muscle tension dysphonia, unilateral benign masses and bilateral benign masses. *Journal of Voice*, 14(2), 240-246.

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer [Logiciel]. Version 5.3.55, repéré le 2 Septembre 2013 à http://www.praat.org/.

Bourdin D. & Navion A., (2013). *Mesure de l'efficacité vocale au sein d'une population de chanteurs de humain beatbox: Analyse acoustique, aérodynamique et observation comportamentale* (Mémoire d'Orthophonie non publié). Université Claude Bernard Lyon1 - ISTR – Orthophonie.

Calliope (1989). *La parole et son traitement automatique*. Paris : Masson.

Clouet A. & de Torcy T., (2010). *Le Human Beatbox : études qualitatives acoustique en vidéo-fibronasoscopie* (Mémoire pour le certificat de capacité d'Orthophoniste non publié). Université Paris 6, France.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M. & Gerstman L.J. (1952). Some Experiments on the Perception of Synthetic Speech Sounds. *The Journal of the Acoustical Society of America*, 24, 597.

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

Forrest, K., Weismer, G., Milenkovich, 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.

Francis, A.L., Kaganovich, N. & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, 124(2), 1234-1251.

Garnier, M. (2009). Forçage vocal et efficacité de communication. Dans P. Gatignol, *La voix dans tous ses maux* (pp. 83-107). Paris : Ortho Edition.

Garnier, M. (2007). *Communiquer en environnement bruyant : de l'adaptation jusqu'au forçage vocal*. (Thèse de doctorat). Université Paris 6, France.

Garrigues L., (2015). *Gestion des sons non pulmonaires et de la phonation inversée en Human Beatbox* (Mémoire pour le certificat de capacité d'Orthophoniste). Université Paris 6, France.

Giovanni, A., Aumelas, E., Chapus, E., Lassalle, A., Remacle, M. & Ouaknine, M. (2004). Le forçage vocal et ses conséquences. *Annales d'Otholaryngologie et de Chirurgie Cervicofaciale*, 121, 187-196.

Giovanni, A., Sacre, J. & Robert, D. (2007). *Le forçage vocal*. Otho-rhino-laryngologie. Paris : Elsevier.

Gravel, J.S. & Ohde, R.N. (1983). Perception of stop place of articulation : Effects of stimulus amplitude. *American Speech-Language-Hearing Association*, 25, 101.

Hardcastle, W.J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics*, 1, 263-272.

Kuhn, G.M. (1975). On the front cavity resonance and its possible role in speech perception. The Journal of the Acoustical Society of America, 58(2), 428-433.

Lagier, A., Vaugoyeau, M., Legou, T., Ghio, A., Amy de la Bretèque, B., Assaiante, Ch. & Giovanni, A. (2010). Etudes expérimentales préliminaires de la voix chuchotée : pression sous-glottique et étude posturale. *Revue de Laryngologie, Otologie, Rhinologie*, 113(1), 1-4.

Le Huche, F. & Allali, A. (2010). *La voix* (éd. 4, Vol. 1, éd. 2, Vol. 2, éd. 3, Vol. 3). Paris : Masson.

Lederer K., (2005). *The phonetics of beatboxing* (Mémoire de Licence non publié). Leeds University, UK.

Lisker, L. (1975). Is VOT or a first formant transition detector?The Journal of the Acoustical Society of America, 57, 1547-1551.

Lisker, L. & Abramson, A.S. (1964). A Cross Language Study of Voicing in Initial Stops : Acoustical Measurements. *Word*, 20, 384-422.

Lousada, M., Jesus, L. & Pape, D. (2012). Estimation of stops' spectral place cues using multitaper techniques. *DELTA*, 28(1), 1-26.

Lubker, J.F. & Parris, P.J. (1970). Simultaneous measurements of intraoral preassure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *The Journal of the Acoustical Society of America*, 47, 625-633.

Martino R., (2009). *Le Human Beatbox et ses pratiquants* (Mémoire de Master 2 non publié, Université Pierre-Mendés-France, Grenoble). Repéré à http://robin.martino.perso.neuf.fr/Site/Robin_Martino_files/Le%20beatbox%20et%20ses%20pratiquants.pdf

MATLAB Version R2011b, The MathWorks, Inc., Natick, Massachussetts, United States.

Maddieson, I. (1997). Combining frication and glottal constriction : Two solutions to a dilemma. *The Journal of the Acoustical Society of America*, 102(5), 3135.

McClean, M.D. & Tasko, S.M. (2003). Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech, Language, and Hearing Research*, 46(6).

Morrison, M. (1997). Pattern recognition in muscle misuse voice disorders : How I do it. *Journal of Voice*, 11(1), 108-114.

Ohala, J.J. & Riordan, C.J. (1979). Passive vocal tract enlargement during voiced stops. Dans J. Wolf & D.H. Klatt, *Speech Communication Papers* (p.89-92). New York : Acoustical Society of America.

Ojamaa, T. & Ross, J. (2009). Sound and timing must be perfect. Production aspects of the human beatboxing. Dans *Proccedings of the Fifth Conference on Interdisciplinary Musicology* (CIM09).

Osfar, M.J. (2011). Articulation of whispered alveolar consonants (Thèse de doctorat). University of Illinois.

Paroni A., (2014). *How do beatboxers play with their tongue and lips? An ultrasound and high-speed imaging exploration* (Mémoire d'Orthophonie non publié). Università degli Studi di Padova, Italie.

Pillot, C. (2004). *Sur l'efficacité vocale dans le chant lyrique, Aspects physiologique, cognitif, acoustique et perceptif*. (Thèse de doctorat non publiée). Université Paris 3, France.

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.

R Development Core Team (2008). R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Revis, J., Marques, A., Fredouille, C., Ghio, A. & Giovanni, A. (2009). Influence du contexte phonétique sur la manifestation dysphonique : apport des nouvelles méthodologies d'analyse de la voix pathologique. Dans P. Gatignol, *La voix dans tous ses maux* (pp. 63-82). Paris : Ortho Edition.

Sapthavee, A., Yi, P. & Sims, H.S., (2014). Functionale Endoscopic Analysis of Beatbox Performers. *Journal of Voice, 28 (3)*, 328-331.

Sernicles, W. (1984). Fenêtre de prélèvement temporel des indices d'occlusives. Dans *Actes des Xièmes Journées d'Etudes sur la parole* (67-78). Williams.

Shadle, C.H. (1997). The Aerodynamics of Speech. Dans W. J. Hardcastle & J. Laver *Handbook of Phonetics* (pp.33-64). Blackwell Publishers.

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America, 85(1)*, 295-312.

Shutte, H. (1980). *The efficiency of voice production*. Groningen : Kemper.

Solomon, N.P. (2008). Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology*, 10(4), 254-266.

Stevens, K.N. (1972). The quantal nature of speech : Evidence from articulatory-acoustic data. Dans P.B. Denes, & E.E. David Jr., *Human communication : a unified view* (pp. 51-66). New York : McGraw Hill.

Stevens, K.N. (1989). On the quantal nature or speech. *Journal of Phonetics, 17*, 3-46.

Stevens, K.N., Manuel, S.Y. & Matthies, M. (1999). Revisiting place of articulation measures for stop consonants : Implications for models of consonant production. Dans *Proceeding of the International Congress of Phonetics Sciences* (pp.1117-1120).

Sundberg, J., Cleveland, T., Stone, R. & Iwarsson, J. (1999). Voice source carachteristics in six premier country singers. *Journal of voice, 13(2)*, 168-183.

Tyte, G. & Splinter, M. (2002). Standard beatbox notation. Repéré à https://www.humanbeatbox.com/articles/standard-beatbox-notation-sbn/

Tyte, G. (2005). The new skool. Repéré à https://www.humanbeatbox.com/articles/history-of-beatboxing-part-3/

Tyte, G. & White Noise (2005). The old skool. Repéré à https://www.humanbeatbox.com/articles/history-of-beatboxing-part-2/

Van den Berg, J.W. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech, Language, and Hearing Research*, 1, 227-244.

Wohlert, A.B. & Hammen, V.L. (2000). Lip muscle activity related to speech rate and loudness. Journal of Speech, Language, and Hearing Research, 43(5), 1229-1239.

Sigles et abréviations utilisés

BB1 :	beatboxer 1
BB2 :	beatboxer 2
BB3 :	beatboxer 3
BB4 :	beatboxer 4
CDG :	centre de gravité spectral
EGG :	electroglottographie
EMA :	electo-magnetic articulography
Pio :	pression intra-orale
IRM :	imagerie à résonance magnétique
SBN :	standard beatbox notation
VOT :	voice onset time

Table des illustrations

Figure 3.1: Signal audio, spectrogramme et phases de la production du segment /ypa/. V : voyelle ; S : silence ; B : « burst » ; BF : bruit de friction ; TF : transitions formantiques. Spectre : 0-12 kHz. 16

Figure 3.2: Représentation schématique de la configuration du conduit vocal lors de la phase d'occlusion des consonnes plosives orales du français /p/, /t/, /k/.....17

Figure 3.3: Illustration du coefficient d'asymétrie (a et b) et du coefficient d'aplatissement (c).....19

Figure 4.2.1: Les sujets de l'étude acoustique.....31

Figure 4.2.2: Dispositif EVA2 et EGG, au cours de la manipulation. D'après Bourdin & Navion, 2013, p. 32.....32

Figure 4.2.3: Segmentation sous Praat d'une phrase parlée de façon conversationnelle.....34

Figure 4.2.4: Segmentation sous Praat d'une phrase beatboxée.....35

Figure 4.2.5: Illustration de l'usage du signal EGG pour la segmentation d'une phrase parlée.....35

Figure 4.2.6: Durée des réalisations de la cible /p/. « p fric » désigne la réalisation d'une plosive bilabiale parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).....40

Figure 4.2.7: Durée des réalisations de la cible /t/. « t fric » désigne la réalisation d'une plosive alvéolaire parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).....41

Figure 4.2.8: Durée des réalisations de la cible /k/. « k fric » désigne la réalisation d'une plosive vélaire parlée ou chantée dont le bruit de friction est allongé (réalisation affriquée).....41

Figure 4.2.9: Max d'intensité pour les réalisations purement plosives des cibles /p/, /t/, /k/.....42

Figure 4.2.10: Corrélation entre l'intensité maximale et la durée du bruit de la cible /p/ dans le cas de BB2.....44

Figure 4.2.11: CDG pour les réalisations purement plosives des cibles /p/, /t/, /k/. 45

Figure 4.2.12: Corrélation le CDG et l'intensité maximale du son plosif /p/ dans le cas de BB2.....47

Figure 4.2.13: Coefficient d'asymétrie pour les réalisations purement plosives des cibles /p/, /t/, /k/. 47

Figure 4.2.14: Coefficient d'aplatissement pour les réalisations purement plosives des cibles /p/, /t/, /k/.....49

Figure 4.2.15: Pio pour les sons plosifs bilabiaux produits par BB1.....50

Figure 4.2.16: Corrélation entre la Pio et l'intensité maximale du bruit de la cible /p/ dans le cas de BB1.....51

Figure 4.2.17: Dérivée temporelle du débit d'air oral pour les réalisations purement plosives des cibles /p/, /t/, /k/.....52

Figure 4.2.18: Corrélation entre l'intensité maximale et la vitesse du débit d'air oral pour la cible /p/ (a), /t/ (b) et /k/ (c) produites par BB1.....53

Figure 4.2.19: Corrélation entre le coefficient d'aplatissement et l'intensité maximale pour la cible /p/ (a), /t/ (b) et /k/ (c) produites par BB1.....53

Figure 4.2.20: Positionnement du sujet à l'intérieur du champ magnétique de l'EMA.....63

Figure 4.2.21: Positionnement des bobines de l'EMA.....64

Figure 4.2.22: Les conditions d'expérience à l'intérieure de la chambre sourde.....65

Figure 1: Consonnes réalisées en Beatbox par BB1.....77

Figure 2: Consonnes réalisées en Beatbox par BB2.....77

Figure 3: Consonnes réalisées en Beatbox par BB4.....77
Figure 4: Consonnes réalisées en Beatbox par BB3.....77
Figure 5: Consonnes réalisées en parole par BB1.....78
Figure 6: Consonnes réalisées en parole par BB2.....78
Figure 7: Consonnes réalisées en parole par BB4.....78
Figure 8: Consonnes réalisées en parole par BB3.....78
Figure 9: Consonnes réalisées en voix criée par BB1.....79
Figure 10: Consonnes réalisées en voix criée par BB2.....79
Figure 11: Consonnes réalisées en voix criée par BB3.....79
Figure 12: Consonnes réalisées en voix criée par BB4.....79
Figure 13: Consonnes réalisées en voix projetée par BB1.....80
Figure 14: Consonnes réalisées en voix projetée par BB2.....80
Figure 15: Consonnes réalisées en voix projetée par BB3.....80
Figure 16: Consonnes réalisées en voix projetée par BB4.....80
Figure 17: Consonnes réalisées en voix chantée par BB4.....81
Figure 18: Consonnes réalisées en voix chantée par BB4.....81
Figure 19: Consonnes réalisées en voix chantée par BB4.....81
Figure 20: Consonnes réalisées en voix chantée par BB4.....81
Figure 21: Valeurs maximales d'intensité pour toutes les réalisations des cible /p/, /t/, /k/. « fric » désigne la réalisation d'une plosive dont le bruit de friction est allongé (réalisation affriquée).....82
Figure 22: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /p/.....83
Figure 23: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /t/.....84
Figure 24: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /k/.....85
Figure 25: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /p/.....86
Figure 26: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /t/.....87
Figure 27: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /k/.....88

Table des annexes

Annexe 1
Protocole (étude acoustique).....75

Annexe 2
Correspondance cible-production.....77

Annexe 3
Graphes concernant l'intensité du « burst ».....82

Annexe 4
Graphes concernant la corrélation entre l'intensité maximale et la durée du bruit.....83

Annexe 5
Graphes concernant la corrélation entre le CDG et l'intensité maximale du bruit.....86

Annexe 1
Protocole (étude acoustique)

PROTOCOLE – Effort vocal et efficacité dans le Human Beatbox : Apport pour la rééducation pneumo-phono-résonantielle

Le sigle ° représente un équivalent de la voyelle avec un bruit de percussion gorge fermée. S'il y a plusieurs possibilités pour produire le son d'une consonne, les principales possibilités seront explorées.

Première série :

Elle est partie avec ton tonton, ton Taine et ton thon / ° ° P T ° K T T T N T T

Pâtes au pistou / P T Ps T

Petit pot de pesto / P T P D Ps T

Ecartons ton carton / ° Kr T T Kr T

Pourquoi pas / Pr K P

Tant pis pour toi / T P Pr T

N'aie pas peur / N P Pr

Peux-tu patiner? / P T P T N

Apporte un petit pot / ° Pr T ° P T P

Paul ne peut pas passer / P N P P P S

Qu'en penses-tu? / K Ps T

Deuxième série :

Tais toi! / TT

Ou étais-tu? / ° ° T T

On t'attend / ° T T

Tu étais têtu / T ° T T T

Tu as tout ton temps / T ° T T T

Tu es tout honteux / T ° T ° T

As-tu été à Tahiti? / ° T ° T ° T ° T

Ta tortue est toute petite / T T T ° T T P T T

Ton thé t'a t'il oté ta toux / T T T T ° T T T

Troisième série : (avec la snare)

Mon képi est kaki / M K P ° K K

Qui n'a pas compris / K N P K P

Il manque quelques briques / ° ° K K K B K

Les canards font coin coin coin / ° K N F K K K

Qui a commencé / K ° K ° s

Qu'est-ce qu'il y a? / Ks K ° °

Que c'est compliqué / K S K P K

Quatrième série :

Compte les tickets / K T ° T K

Luc est parti en tacot / ° K ° P T ° T K

Ma culotte est en coton / ° K ° T ° ° K T

Combien a couté ton couteau? / K ° ° K T T K T

As-tu un train électrique / ° T ° T ° K T K

Annexe 2
Correspondance cible-production

Beatbox

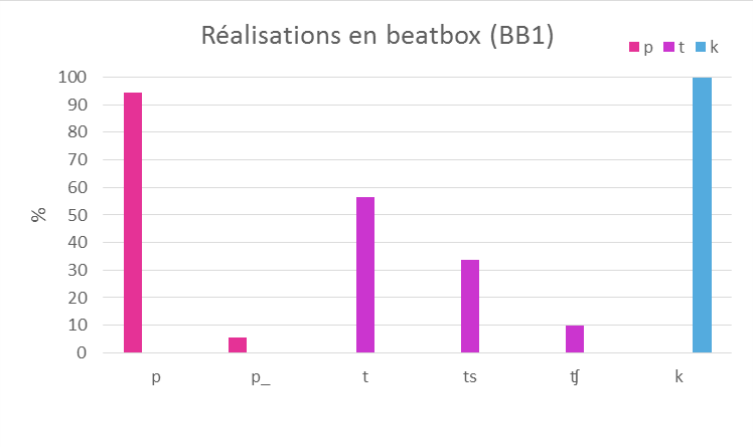

Figure 1: Consonnes réalisées en Beatbox par BB1.

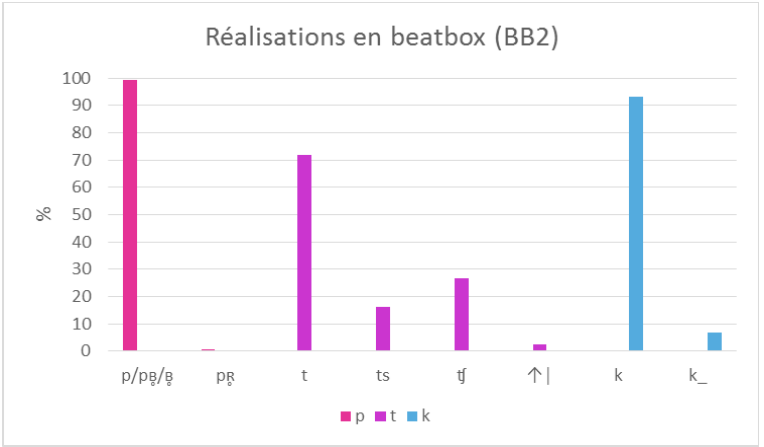

Figure 2: Consonnes réalisées en Beatbox par BB2.

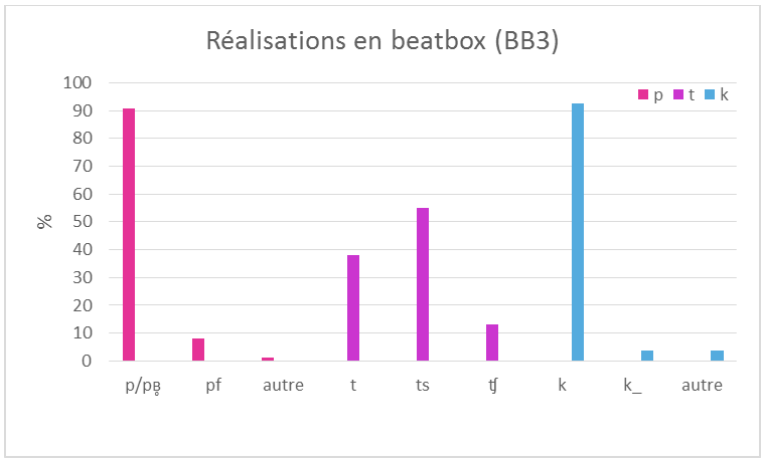

Figure 4: Consonnes réalisées en Beatbox par BB3.

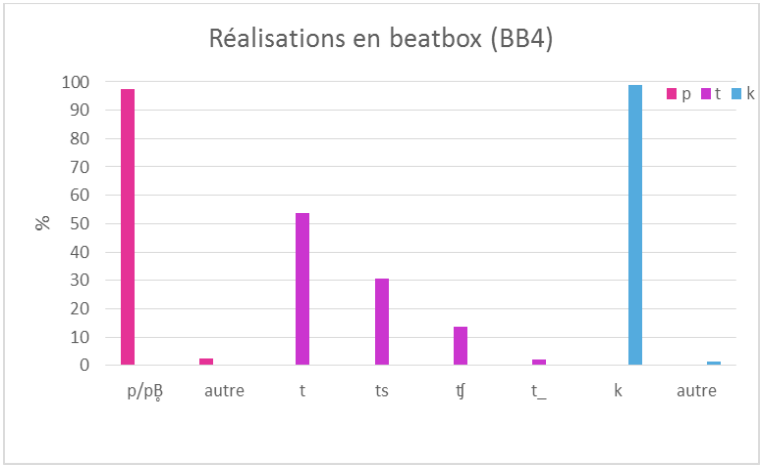

Figure 3: Consonnes réalisées en Beatbox par BB4.

Parole conversationnelle

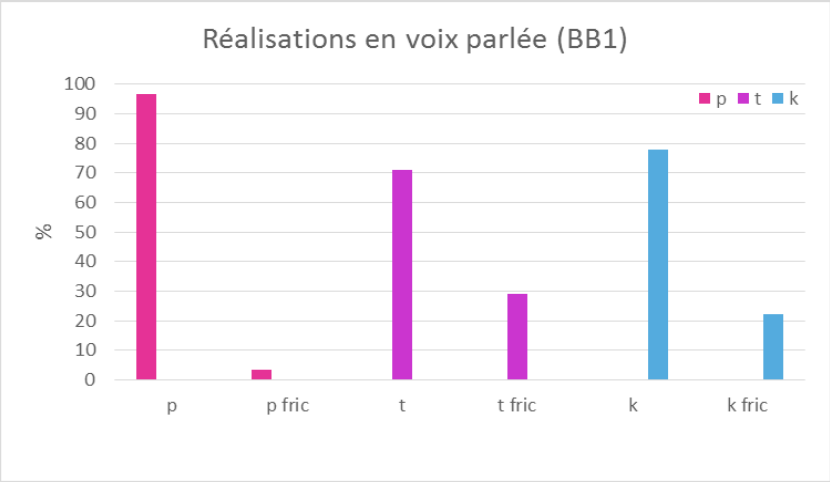

Figure 5: Consonnes réalisées en parole par BB1.

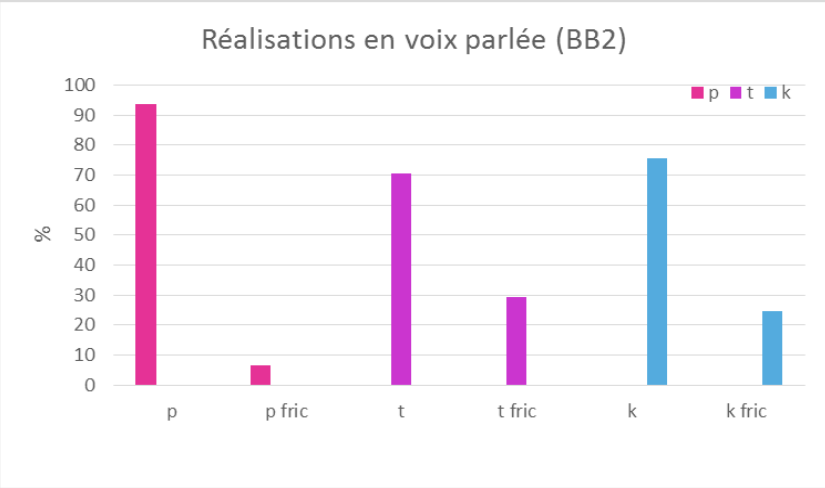

Figure 6: Consonnes réalisées en parole par BB2.

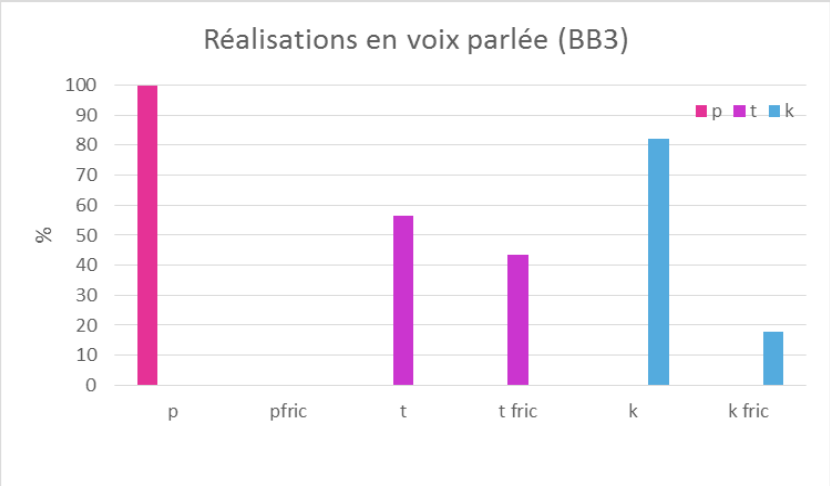

Figure 8: Consonnes réalisées en parole par BB3.

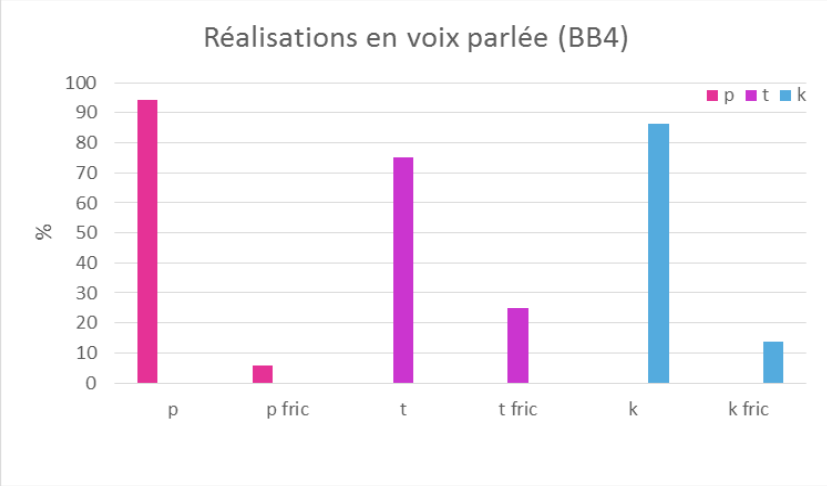

Figure 7: Consonnes réalisées en parole par BB4.

Parole criée

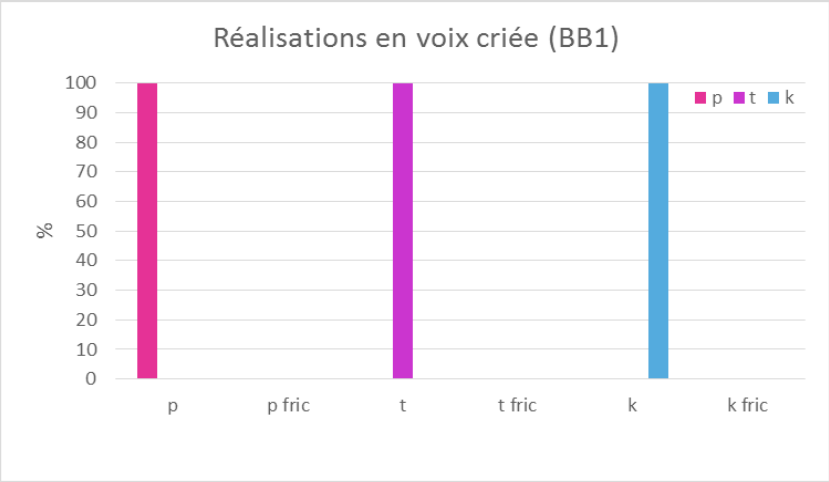

Figure 9: Consonnes réalisées en voix criée par BB1.

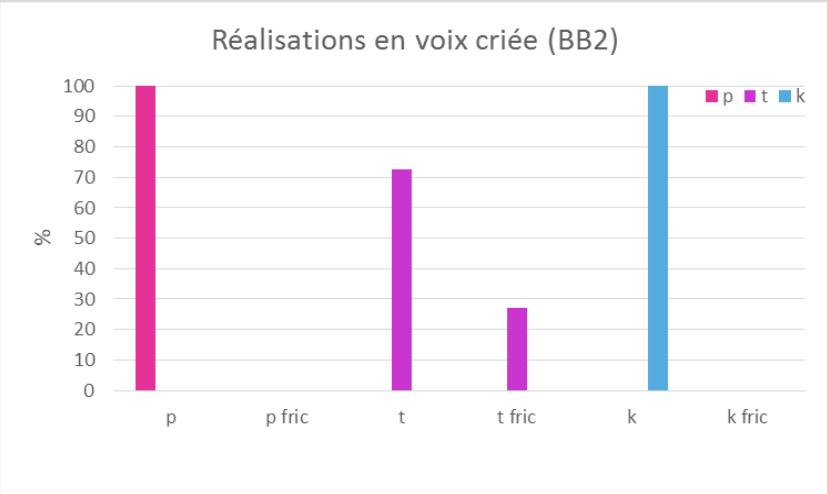

Figure 10: Consonnes réalisées en voix criée par BB2.

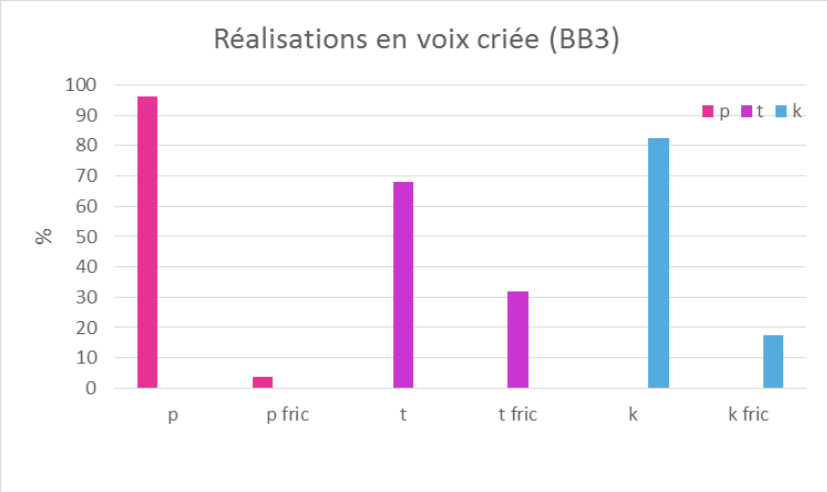

Figure 11: Consonnes réalisées en voix criée par BB3.

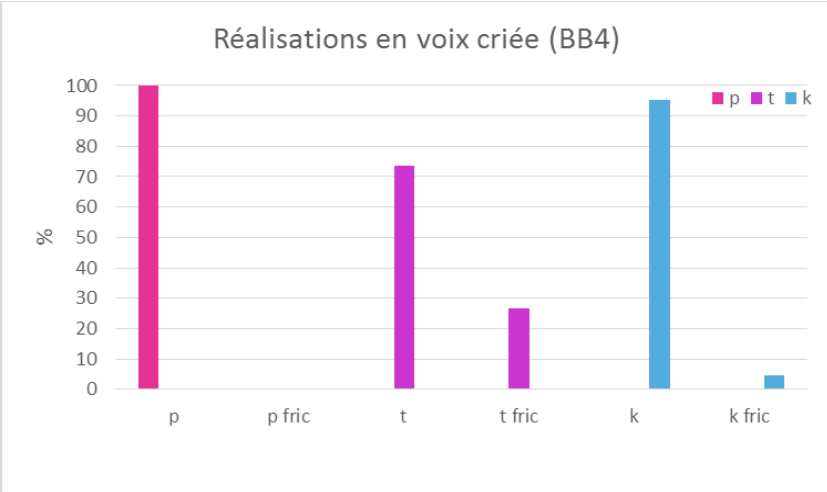

Figure 12: Consonnes réalisées en voix criée par BB4.

Parole projetée

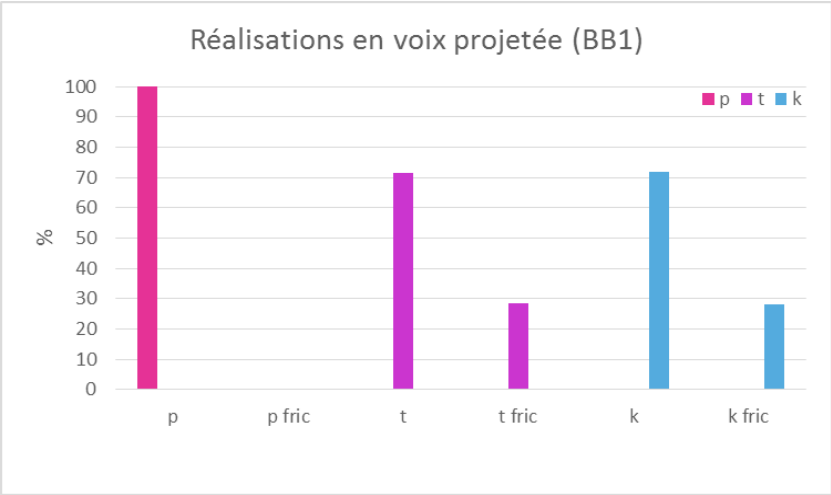

Figure 13: Consonnes réalisées en voix projetée par BB1.

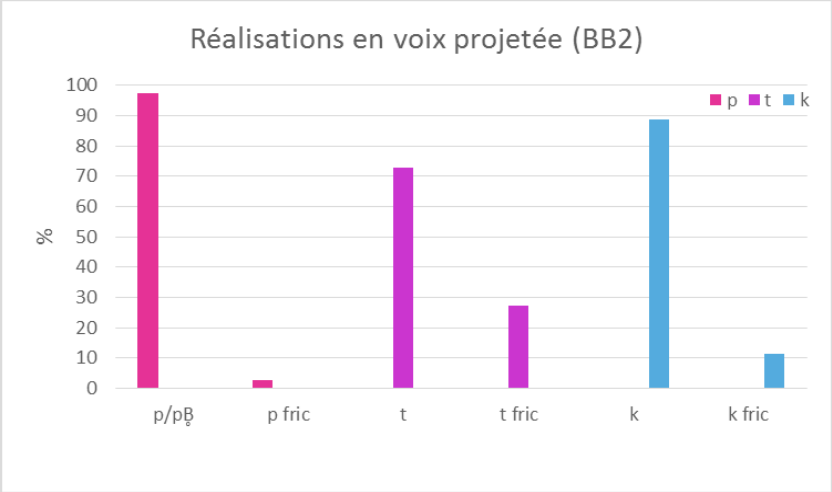

Figure 14: Consonnes réalisées en voix projetée par BB2.

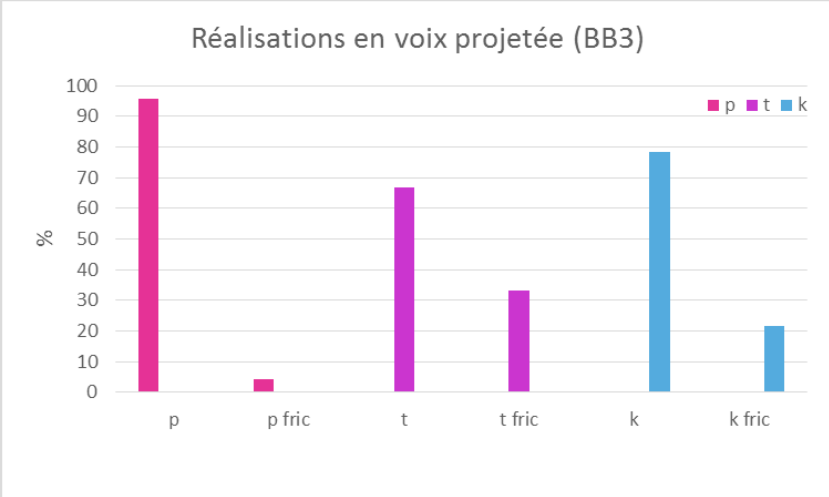

Figure 15: Consonnes réalisées en voix projetée par BB3.

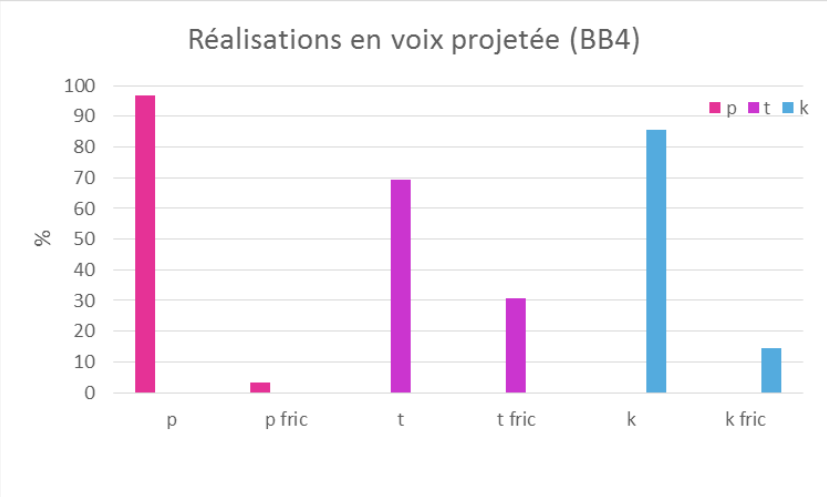

Figure 16: Consonnes réalisées en voix projetée par BB4.

Chant

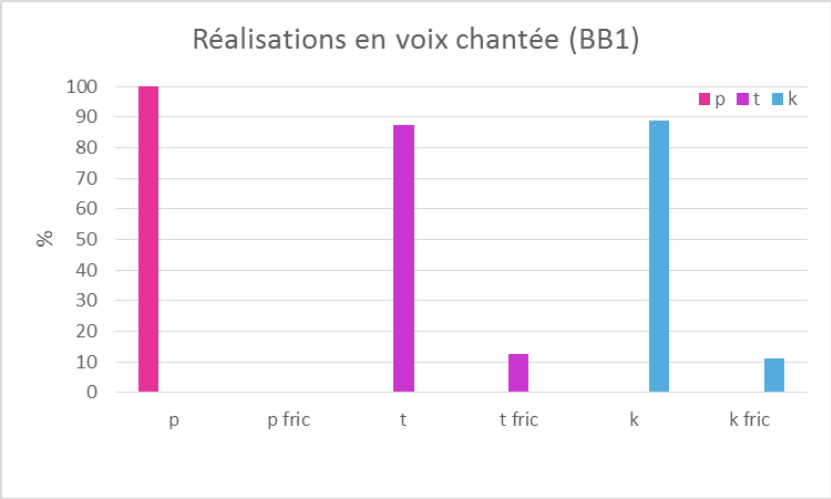

Figure 17: Consonnes réalisées en voix chantée par BB4.

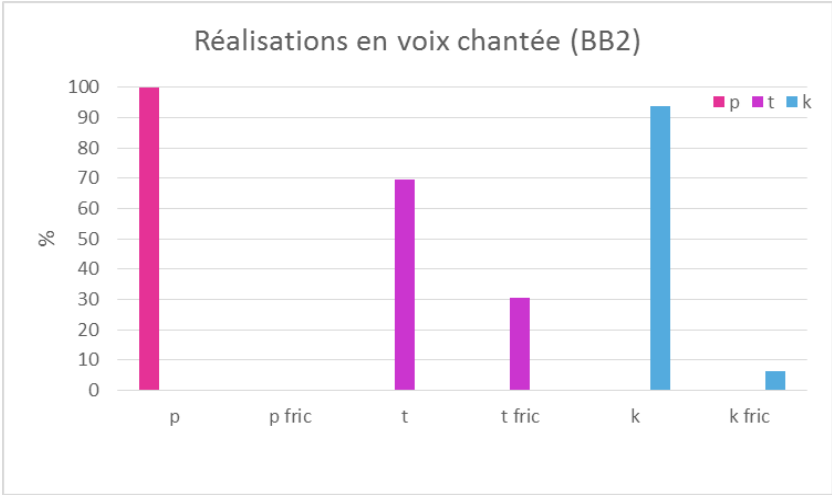

Figure 18: Consonnes réalisées en voix chantée par BB4.

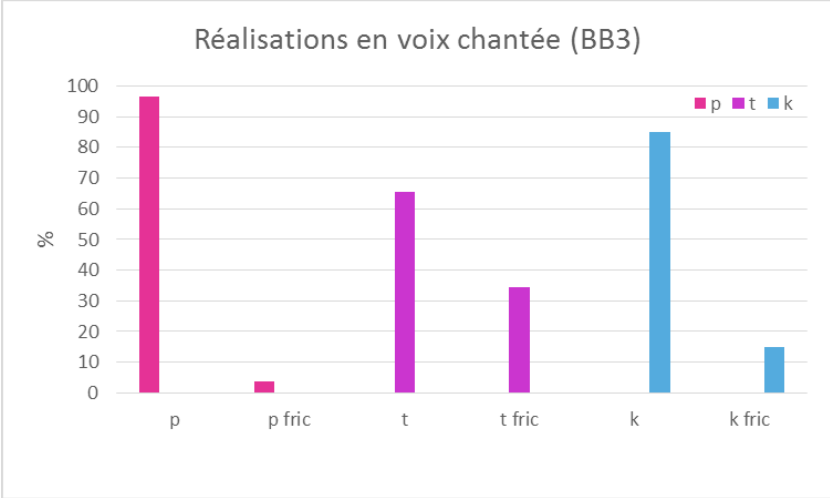

Figure 19: Consonnes réalisées en voix chantée par BB4.

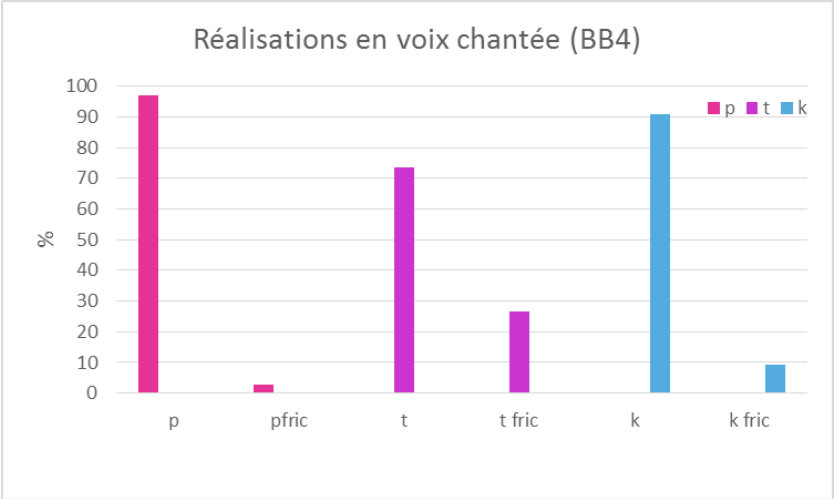

Figure 20: Consonnes réalisées en voix chantée par BB4.

Annexe 3
Graphes concernant l'intensité du « burst »

Valeurs maximales d'intensité pour tout type de réalisation des cibles /p/, /t/, /k/.

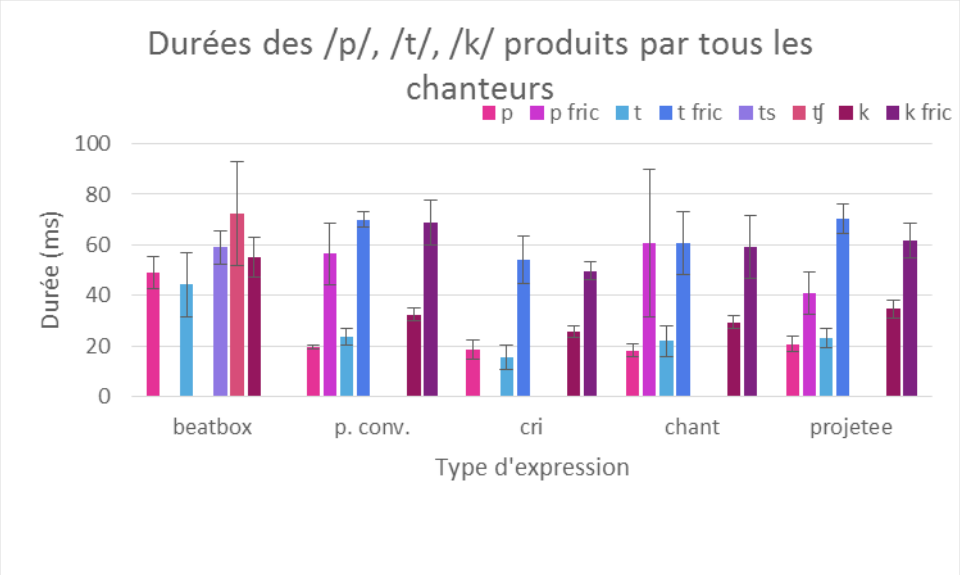

Figure 21: Valeurs maximales d'intensité pour toutes les réalisations des cible /p/, /t/, /k/. « fric » désigne la réalisation d'une plosive dont le bruit de friction est allongé (réalisation affriquée).

Annexe 4
Graphes concernant la corrélation entre l'intensité maximale et la durée du bruit.

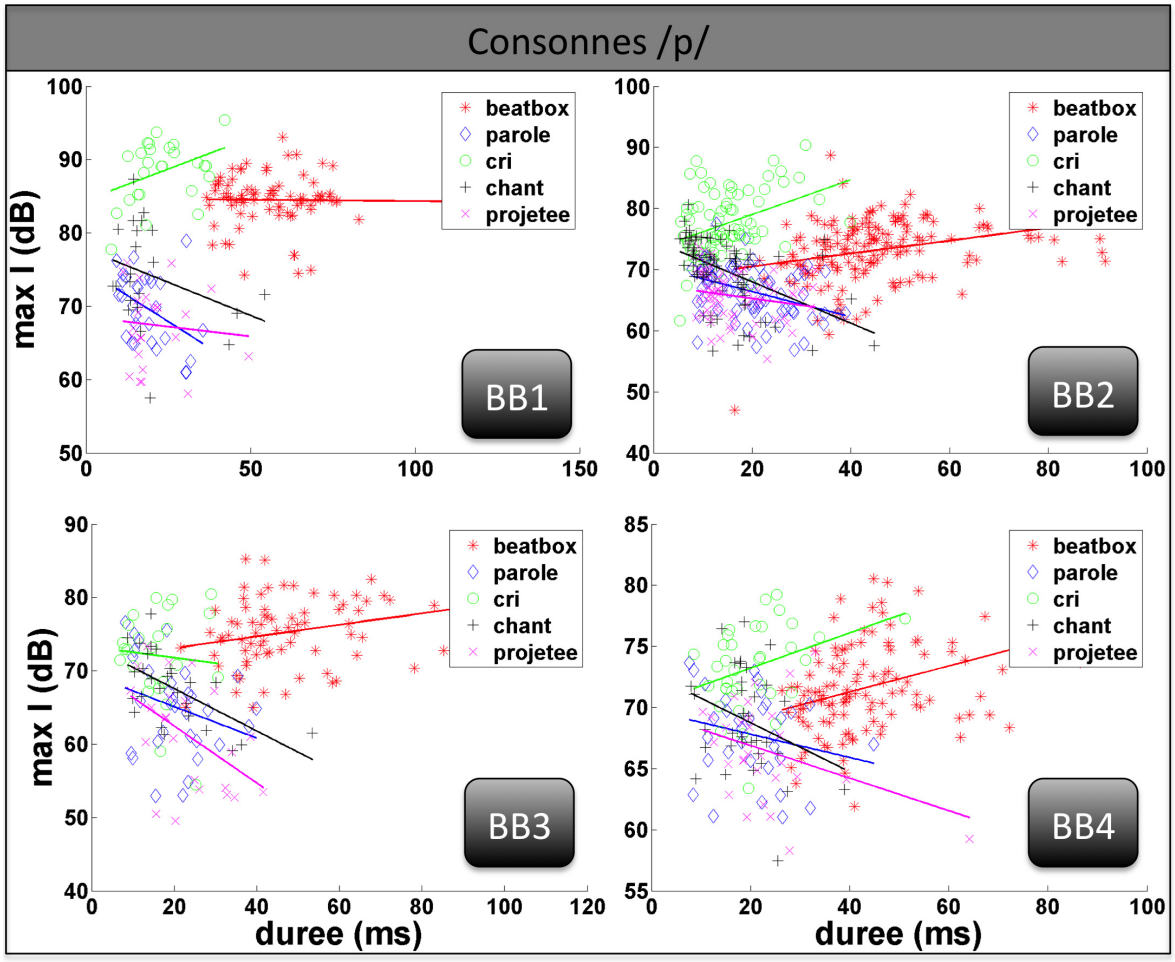

Figure 22: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /p/.

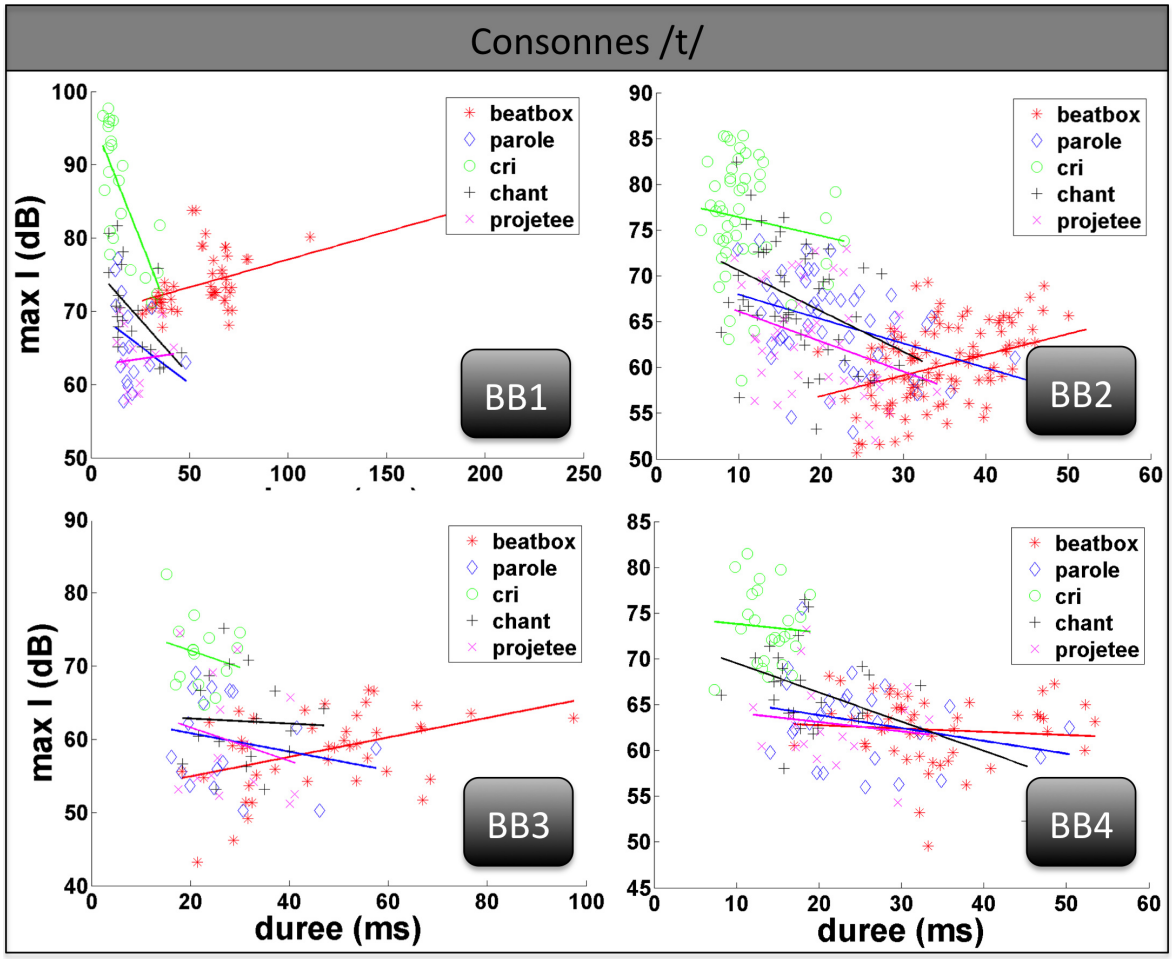

Figure 23: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /t/.

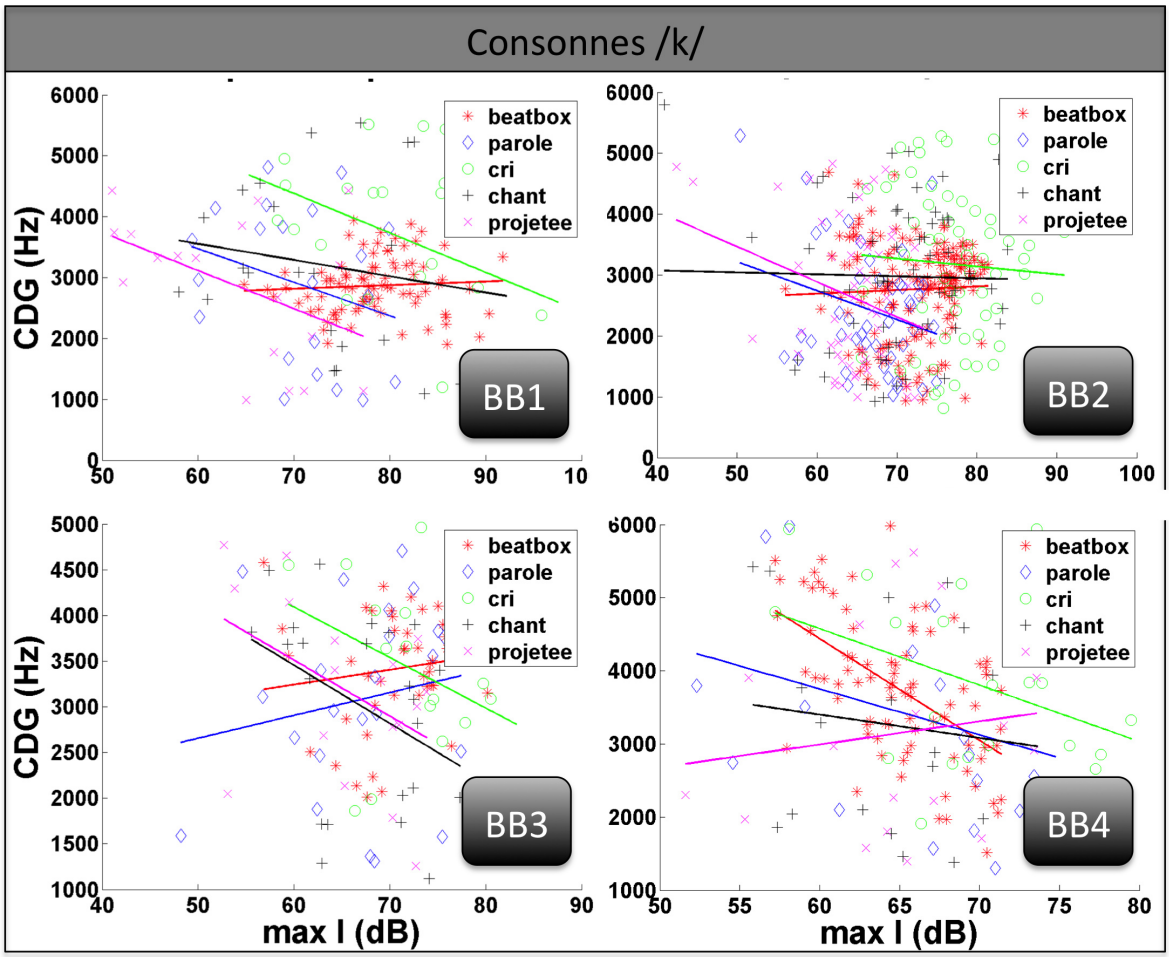

Figure 24: Corrélation entre l'intensité maximale et la durée du bruit pour la cible /k/.

Annexe 5
Graphes concernant la corrélation entre le CDG et l'intensité maximale du bruit.

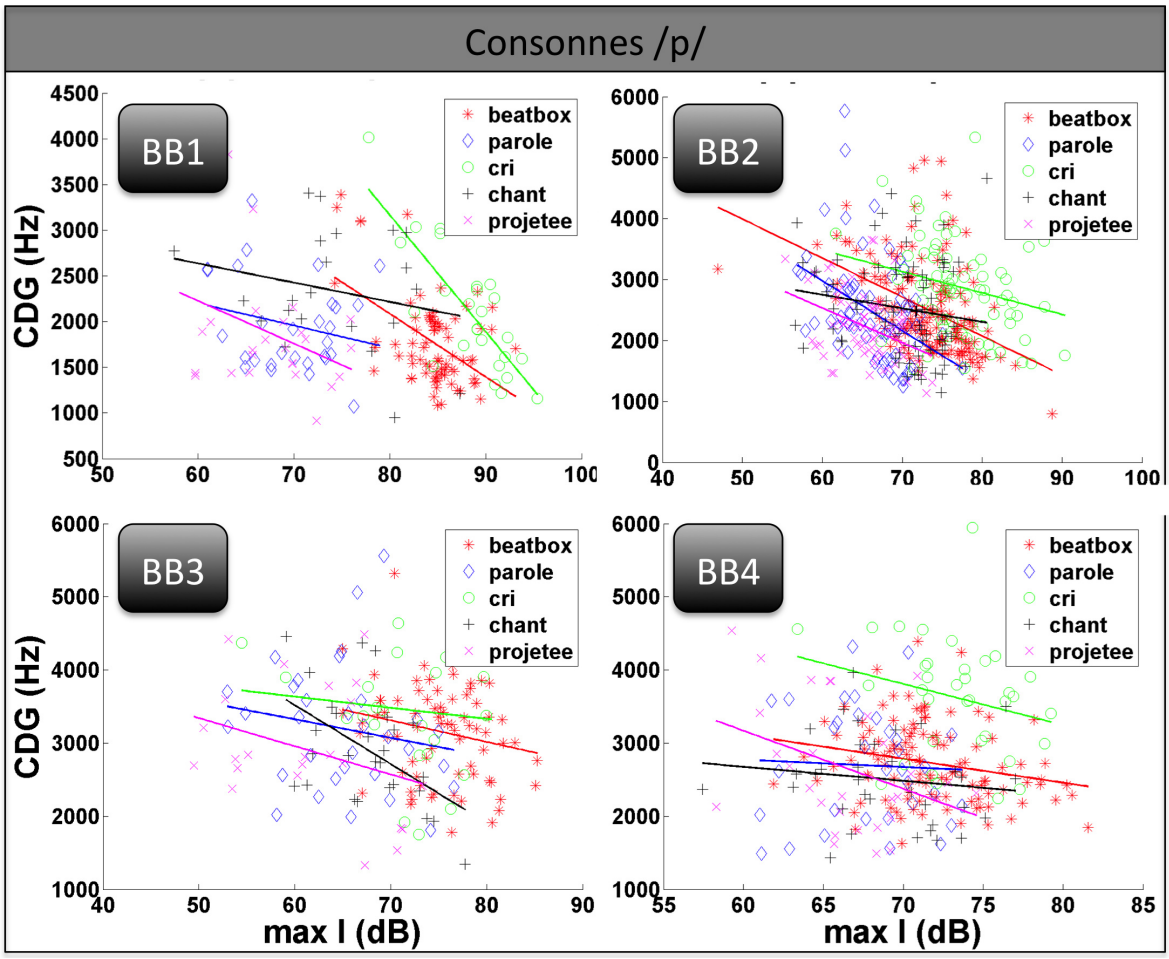

Figure 25: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /p/.

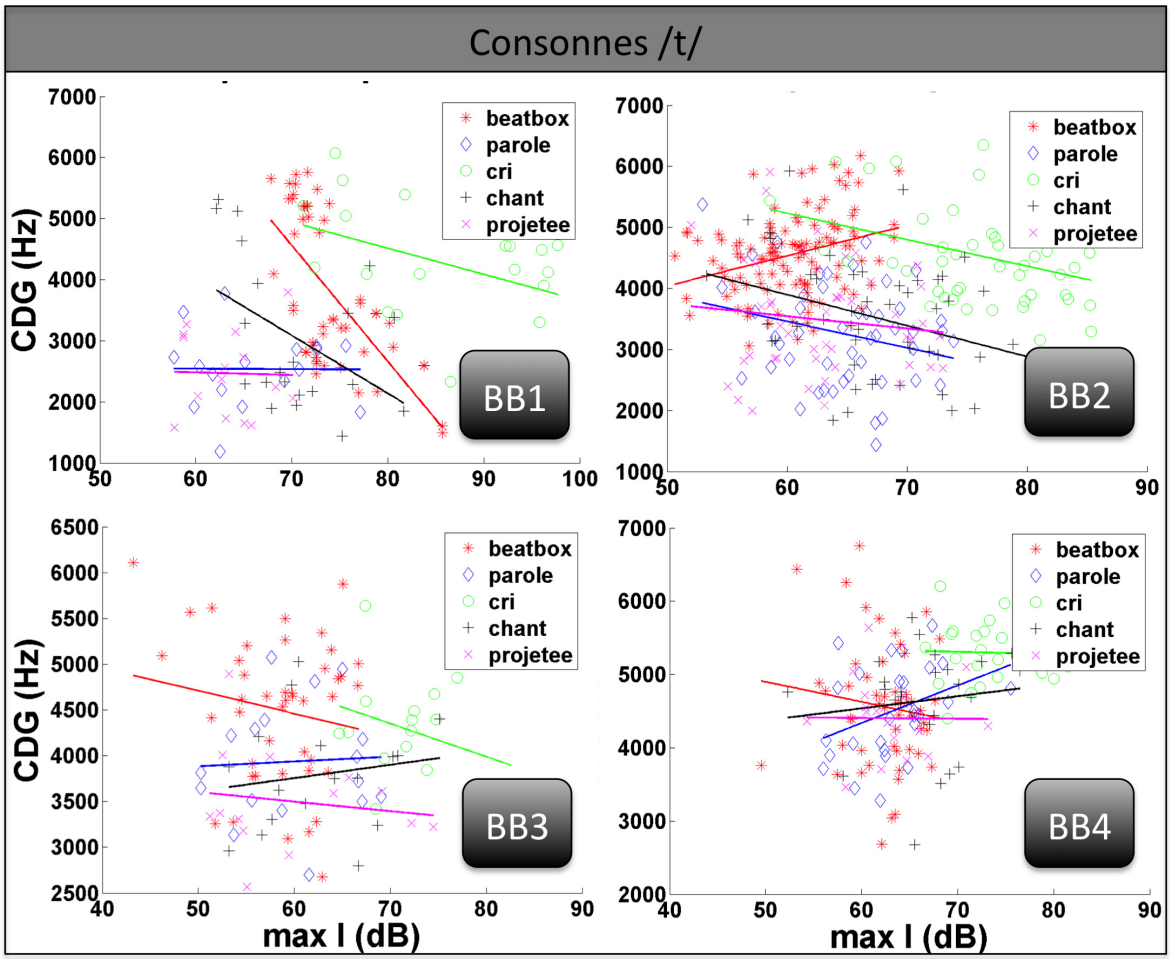

Figure 26: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /t/.

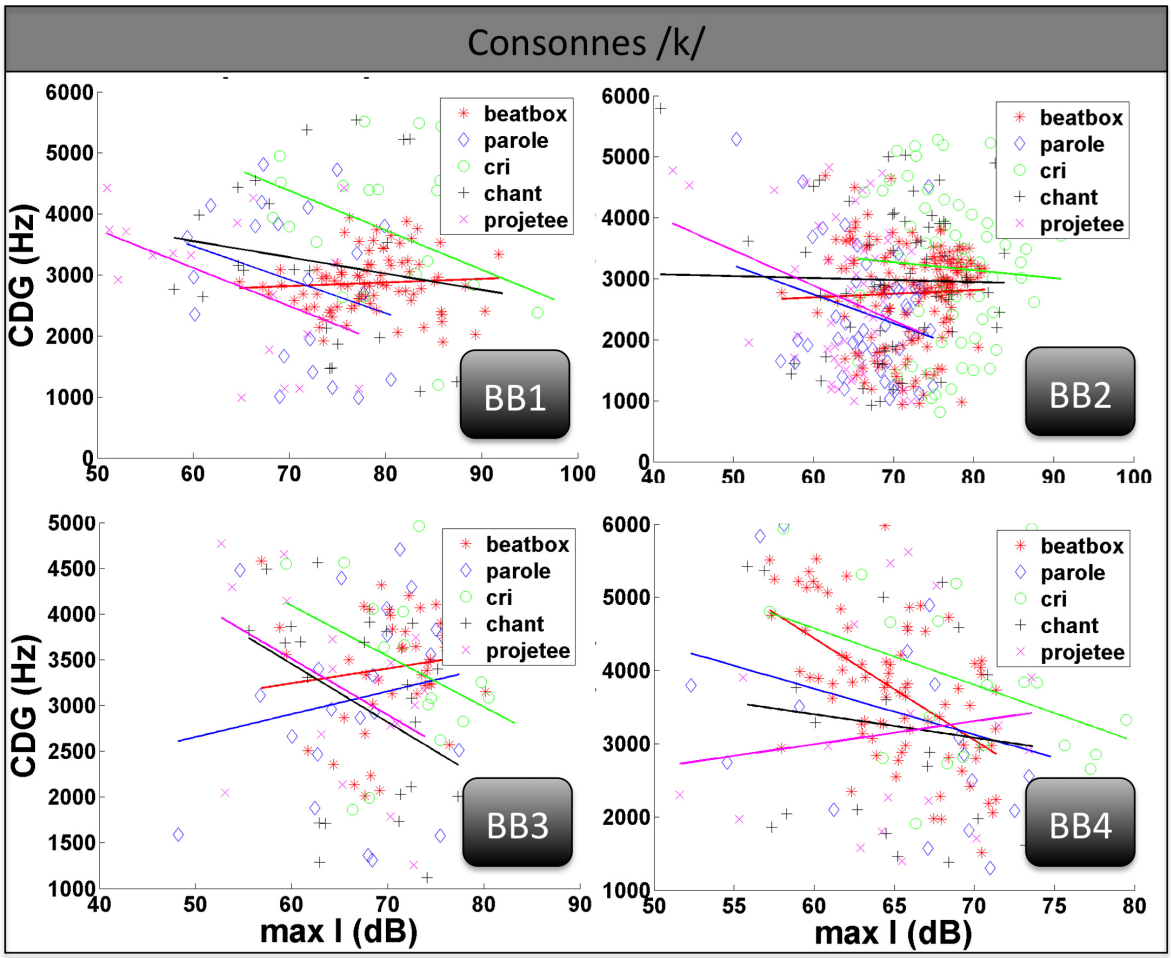

Figure 27: Corrélation entre le CDG et l'intensité maximale du bruit pour la cible /k/.

Table des matières

Remerciements....4
Sommaire....5
Introduction....6
Partie 1 : Partie théorique....7
Chapitre 1. Contexte de recherche....8
1.Le GIPSA-lab....8
2.Le projet StopNCo....9
3.Les notions d'effort vocal et d'efficacité....11
Chapitre 2. Les consonnes....15
1.Les consonnes du français....15
2.Les consonnes plosives orales non voisées....15
3.Indices acoustiques de discrimination des consonnes plosives....17
3.1.Indices de voisement....18
3.2.Indices de lieu d'articulation....18
4.Contrôle des indices acoustiques de discrimination des consonnes plosives....20
4.1.Contrôle du VOT....20
4.2.Contraintes et contrôle de l'énergie du « burst »....21
Chapitre 3. Une production experte : le cas du Human Beatbox....23
1.Intérêt pour l'étude du Human Beatbox....24
2.Cadre théorique....25
3.Le Human Beatbox comme paradigme expérimental....26
4.Problématiques et hypothèses....28
Partie 2 – Partie expérimentale....30
Chapitre 5. Étude acoustique....31
1.Matériel et méthodes....31
1.1.Sujets....31
1.2.Protocole et corpus....32
1.3.Méthodologie et outils d'analyse....33
1.3.1.Statistique descriptive....36
1.3.2.Statistique inférentielle....37
2.Résultats....38
2.1.Correspondance cible-production....38
2.1.1.Structure de la phrase....38
2.1.2.Correspondance consonne cible-consonne produite....38
2.2.Caractéristiques acoustiques des bruits de plosion....40
2.2.1.Durée du bruit....40
2.2.2.Intensité de la consonne....42
2.2.3.CDG....45
2.2.4.Coefficient d'asymétrie....47
2.2.5.Coefficient d'aplatissement....49
2.3.Caractéristiques aérodynamiques des bruits de plosion....50
2.3.1.Pression intra-orale (Pio)....50
2.3.2.Vitesse du débit d'air oral....52
3.Discussion....54
3.1.H1 : différences acoustiques, aérodynamiques et articulatoires....54
3.2.H2 : corrélations concernant l'intensité du « burst »....56
3.3.H3 : corrélations concernant le spectre du « burst »....56
4.Conclusions....57
Partie 3 – Perspectives de recherche....58
Chapitre 6. Étude articulatoire....59
1.Problématiques et hypothèses....59
1.1.Problématiques générales....59

1.2.Problématique spécifique et hypothèses....60
2.Matériel et méthodes....61
2.1.Sujet....61
2.2.Corpus et protocole....62
2.2.1.Corpus....62
2.2.2.Procédure d'enregistrement et dispositifs....63
3.Résultats....65
4.Discussion et conclusions....66
Bibliographie....67
Sigles et abréviations utilisés....71
Table des illustrations....72
Table des annexes....74
Table des matières....89

MOTS-CLÉS : Plosives non voisées, Efficacité vocale, Effort vocal, Human Beatbox.

RÉSUMÉ

La parole est une action complexe, qui nécessite d'une précise coordination entre la respiration, les gestes laryngés et les gestes articulatoires. Dans le cas où cette coordination est déséquilibrée, il peut se produire des tensions et des efforts excessifs et le risque de développement d'une pathologie de la voix augmente. L'étude de la production des consonnes plosives s'avère intéressant pour mieux comprendre le contrôle de la parole, puisqu'elle nécessite la coordination des gestes en amplitude et force, ainsi que dans leur organisation temporelle. A ce jour, de nombreuses questions demeurent encore quant aux gestes permettant de contrôler finement les caractéristiques acoustiques du bruit de plosion. Les chanteurs du Human Beatbox montrent une maîtrise particulière du contrôle et de la variété des sons plosifs, qu'ils produisent avec une efficacité sonore importante. Dans notre travail, nous comparons la production de trois sons plosifs dans le Beatbox, en trois types de parole (conversationnelle, criée et projetée) et en chant, de façon à caractériser les différences acoustiques et aérodynamiques entre ces modes de production. Nos résultats montrent que la variété de sons produits est plus ample en Beatbox que dans les autres modes d'expression, les sons du Human Beatbox sont deux fois plus longs que leur contreparties parlées et chantées et les paramètres acoustiques et aérodynamiques atteignent des valeurs qui se situent généralement entre celles du cri et celles des autres modes d'expression (parole conversationnelle, parole projetée et chant). Cependant, la vitesse du débit d'air oral est plus importante en Beatbox que dans les autres modes d'expression.

KEYWORDS : Voiceless plosives, Vocal efficiency, Vocal effort, Human Beatbox.

ABSTRACT

Speech is a complex action that requires precise coordination among respiration, laryngeal gestures and articulatory gestures. If this coordination is unbalanced, excessive tension and effort can occur, increasing the risk of voice pathology. The study of the production of plosive consonants is interesting to gain a better understanding of speech control, since it requires the coordination of gestures in terms of their amplitude and force, as well as their timing. At present, many questions regarding the gestures that finely control the acoustic characteristics of the burst remain unanswered. Human Beatbox singers show a deep mastery of the control and variety of plosive sounds, which they produce with high acoustic efficacy. In our work, we compare the production of three plosive sounds in Beatbox, in three types of speech (conversational, shouted and projected) and in singing, so as to characterize the acoustic and aerodynamic differences of these production modes. Our results show that the range of sounds produced is larger in Beatbox than in the other modes, that the Beatbox sounds are twice as long as their spoken and sung counterparts, and that the acoustic and aerodynamic parameters reach levels between those of shouted speech and the other modes of expression (conversational and projected speech and singing). Nevertheless, the velocity of the oral air flow is greater in Beatbox than in all the other modes of expression.