



**HAL**  
open science

# Les sons percussifs : des consonnes plosives au Human Beatbox, corrélations acoustiques, aérodynamiques et endoscopiques

Christol Fabre

► **To cite this version:**

Christol Fabre. Les sons percussifs : des consonnes plosives au Human Beatbox, corrélations acoustiques, aérodynamiques et endoscopiques. Sciences de l'Homme et Société. 2018. dumas-01888799

**HAL Id: dumas-01888799**

**<https://dumas.ccsd.cnrs.fr/dumas-01888799>**

Submitted on 5 Oct 2018

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

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



# **Les sons percussifs : des consonnes plosives au Human Beatbox, corrélations acoustiques, aérodynamiques et endoscopiques**

**Christol FABRE**

Sous la direction des Docteurs

**Nathalie HENRICH BERNARDONI<sup>1</sup> et Ihab ATALLAH<sup>2-3</sup>**

<sup>1</sup> Directrice de Recherche CNRS, GIPSA-Lab, Grenoble

<sup>2</sup> Praticien Hospitalier Universitaire Service d'ORL et Chirurgie Cervico-Faciale, CHU de Grenoble

<sup>3</sup> Institut Albert Bonniot, Grenoble

Laboratoire : GIPSA-Lab, Département Parole et Cognition

**UFR LLASIC**

Département Sciences du Langage et Français Langue Etrangère

---

Mémoire de master 2 Recherche

Parcours : Sciences du Langage – 30 crédits ECTS

Année universitaire 2017-2018





# **Les sons percussifs : des consonnes plosives au Human Beatbox, corrélations acoustiques, aérodynamiques et endoscopiques**

**Christol FABRE**

Sous la direction des Docteurs

**Nathalie HENRICH BERNARDONI<sup>1</sup> et Ihab ATALLAH<sup>2-3</sup>**

<sup>1</sup> Directrice de Recherche CNRS, GIPSA-Lab, Grenoble

<sup>2</sup> PHD Service d'ORL et Chirurgie Cervico-Faciale, CHU de Grenoble

<sup>3</sup> Institut Albert Bonniot, Grenoble

Laboratoire : GIPSA-Lab, Département Parole et Cognition

UFR LLASIC

Département Sciences du Langage et Français Langue Etrangère

---

Mémoire de master 2 Recherche

Parcours : Science du Langage – 30 crédits ECTS

Année universitaire 2017-2018

### DÉCLARATION

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 : ...FABRE.....

PRENOM : ...Christol.....

DATE : .....15/05/2018.....

SIGNATURE





# Sommaire

Sommaire .....	1
Introduction .....	2
Partie 1 - Cadre théorique.....	3
<b>CHAPITRE 1. CONTEXTE DE RECHERCHE : .....</b>	<b>4</b>
1. BREF RAPPELS DE PHYSIOLOGIE DE LA PAROLE .....	4
2. LE LARYNX COMME ARTICULATEUR .....	5
3. LES CONSONNES.....	6
4. MESURES EXPERIMENTALES ARTICULATOIRES DES CONSONNES PLOSIVES .....	10
5. LE HUMAN BEATBOX : ETAT DE L'ART .....	20
<b>CHAPITRE 2. QUESTIONS DE RECHERCHE.....</b>	<b>26</b>
Partie 2 - Evaluation expérimentale.....	27
<b>CHAPITRE 4. MATERIEL ET METHODES.....</b>	<b>28</b>
4. DESCRIPTION DES SUJETS .....	28
4. EFFETS ETUDIES ET CONSONNES ASSOCIEES .....	28
5. PROTOCOLE EXPERIMENTAL .....	29
6. METHODES D'ANALYSE.....	33
7. APPROCHE STATISTIQUE .....	39
<b>CHAPITRE 5. RESULTATS.....</b>	<b>41</b>
1. CARACTERISATION ENDOSCOPIQUE DES MOUVEMENTS LARYNGES CONCOMITANTS A LA PRODUCTION DES TROIS EFFETS. ....	41
2. ETUDE AERODYNAMIQUE ET ACOUSTIQUE .....	76
3. CORRELATIONS ENTRE LES DONNEES ENDOSCOPIQUES ET EVA2 .....	89
4. ECHOGRAPHIE.....	96
Partie 3 - Discussion et mises en perspective .....	99
<b>CHAPITRE 6. DISCUSSION .....</b>	<b>100</b>
1. DISCUSSION .....	100
2. LIMITES ET PERSPECTIVES.....	103
Bibliographie.....	107
Table des illustrations.....	117
Table des annexes.....	119
<b>ANNEXE 1 CHAPITRE ANNEXE. ELABORATION DES TECHNIQUES EXPERIMENTALES.....</b>	<b>120</b>
<b>ANNEXE 2 TABLEAU DES MOYENNES DES TEMPS VIDEOENDOSCOPIQUES .....</b>	<b>122</b>
Table des matières .....	130

## **Introduction**

Le Human Beatbox est un art vocal dont le principe initial est l'imitation de percussions, associée ou non à d'autres techniques ou effets vocaux. Apparu récemment, il est de plus en plus pratiqué grâce à l'essor des plateformes multimédias et d'une communauté de pratiquants très active.

Des travaux antérieurs ont montré que les sons de type percussifs produits en Human Beatbox avaient une efficacité sonore proche de la voix criée et une variabilité acoustique faible et terme d'intensité, sans pour autant provoquer de lésion des structures laryngées malgré une pratique intensive. La proximité acoustique de ces sons avec les consonnes plosives leur octroie un certain potentiel dans l'élaboration de nouvelles techniques de rééducation orthophonique permettant l'augmentation de l'intelligibilité des patients.

Ce travail s'intéresse aux gestes phonatoires qui accompagnent la production des consonnes occlusives dans la parole et des sons percussifs dans le Human Beatbox. Il ambitionne de décrire les structures laryngées impliquées, et de mener une comparaison systématique entre production de parole et Human Beatbox.

Après une première partie théorique concernant le contexte de recherche et les techniques d'exploration de production des consonnes plosives, nous étudierons les mouvements laryngés à partir de données endoscopiques, mises en relation avec les données acoustiques, aérodynamiques, électroglottographiques et échographiques mesurées sur deux beatboxeurs. Enfin nous discuterons des perspectives ouvertes par les résultats obtenus.



## **Partie 1**

-

## **Cadre théorique**

## **Chapitre 1. Contexte de recherche :**

### ***1. Bref rappels de physiologie de la parole***

Le son de parole est façonné tout au long du conduit vocal, il est courant de distinguer trois étages dont la coordination permet la création d'un son particulier.

#### ***1.1. La soufflerie***

Elle est composée de l'étage pulmonaire et des muscles de la respiration (diaphragme, intercostaux...). La phase expiratoire, normalement passive par surpression pulmonaire, est contrôlée dans le cas de la parole afin de s'adapter aux besoins de celle-ci. Le cycle respiratoire se modifie et la phase expiratoire ralentit afin d'avoir une pression régulière, pendant que la phase d'inspiration devient plus rapide pour garder un débit de parole relativement continu.

On parle de mécanisme pulmonaire égressif dans le cas où l'air est chassé des poumons pendant la production de sons, et ingressif dans le cas contraire.

#### ***1.2. Le vibrateur***

Le larynx est le carrefour des voies aérodigestives et a une triple fonction : la protection des voies respiratoires lors de la déglutition tout en garantissant la respiration le reste du temps quand il est ouvert et enfin une fonction phonatoire.

Les plis vocaux soumis au flux d'air d'origine pulmonaire se mettent à vibrer (voisement) et à créer un son à la fréquence fondamentale  $F_0$ . Plusieurs théories sur les mécanismes à l'œuvre existent : la théorie myo-élastique traditionnelle principalement basée sur les déplacements d'air formalisée par Ewald en 1898 (Ewald, 1898) complétée par d'autres auteurs, la théorie muco-ondulatoire dans laquelle est intégré un déplacement indépendant de la muqueuse des plis vocaux met en évidence le phénomène de Bernouilli adapté au larynx (Perello, 1973), en intégrant les propriétés de chaque tissu composant les plis vocaux. Complétée par Van den Berg sous la forme de la théorie aérodynamique élastique (Van den Berg, 1958). La théorie myoélastique révisée par Dejonckere propose que la glotte traversée par le flux d'air réagit en fonction des composantes neuromusculaires comme l'amortissement progressif d'un ressort (Dejonckere, 1985).

Différents mécanismes laryngés sont possibles, impliqués dans la parole et dans le chant, et sont caractérisés par la configuration glottique (Laver, 1980). Il existe ainsi le mécanisme M0, avec des plis vocaux très épais et peu tendus (vocal fry). Certains auteurs

ont aussi décrit un vocal fry avec des plis vocaux compressés (Christian DiCanio, 2015). Le mécanisme M1 est courant en parole, surtout chez l'homme, les plis vibrent sur la totalité de leur longueur. Pour le mécanisme M2, les plis sont fins et ne vibrent que sur 2/3 de leur longueur. Le mécanisme M3 se distingue par des plis vocaux très fins et tendus, avec une amplitude de vibration très faible (voix de sifflet). (Roubeau et al., 2009)

### ***1.3. Les articulateurs et résonateurs***

Les articulateurs sont des organes mobiles du conduit vocal coordonnés volontairement, qui modifient le conduit et ses cavités (résonateurs) jouant un rôle d'amplification ou d'amortissement de l'onde sonore, ou de l'arrêt/relâchement du flux d'air.

Dans la théorie source-filtre le courant d'air devenu son au passage des plis vocaux, remonte dans le conduit supraglottique, pharyngé, buccal et/ou nasal et est modifié tout au long du tractus vocal par les résonateurs, l'enrichissant par exemple, en structures formantiques pour les voyelles.

Pour les consonnes occlusives, l'air est initialement bloqué totalement en un point ou plus du conduit par les articulateurs puis relâché. (Ladefoged & Maddieson, 1996).

La constriction en un endroit du conduit par les articulateurs peut aussi provoquer un bruit de friction par accélération du flux et participent aux consonnes fricatives.

## ***2. Le larynx comme articulateur***

Le larynx est traditionnellement considéré comme la source de phonation, et dans la théorie source-filtre ses capacités d'articulateur n'étaient pas reconnues (Fant, 1960).

Le rôle du pharyngo-larynx en tant qu'articulateur a été longtemps sujet à débat. Les difficultés d'expliquer certaines différences de registre pour des voyelles avec la langue en position arrière dans certaines langues sémitiques et d'Asie du sud-est ont fait suspecté à Hockett puis Catford la présence d'un trait distinctif pharyngo-laryngée. (Ladefoged, P. 1979).

L'apparition de la fibroscopie permettant une vision claire de l'espace pharyngo-laryngé a permis de confirmer ces hypothèses en clarifiant les composantes glottiques, supraglottiques et pharyngées.

Ainsi Esling dès les années 90 puis son élève Moisik, suite aux travaux de nombreux auteurs, formalisent la thèse de l'indépendance des structures pharyngo-

laryngées en tant qu'articulateurs propres et créent le modèle du « larynx articulateur ». Dans ce modèle le conduit vocal est divisé en deux parties : laryngée et orale. La partie laryngée est définie par 3 valves (formées par les plis vocaux, les plis vestibulaires et les replis aryépiglottiques), qui sont toutes capables de rentrer en vibration et de modifier les qualités acoustiques du signal (Edmondson & Esling, 2006). Les traits [pharyngé] ou [épiglottal] seraient surtout dus aux mouvements d'élévation du larynx permettant la fermeture du sphincter pharyngé et pas seulement à rétraction de la langue qui serait moins contrôlable que la position laryngée (Esling 1999; Moisik, Czaykowska-Higgins, et Esling 2011). Leur théorie permet ainsi de caractériser le babillage des enfants dont les vocalises se trouvent principalement modifiées pour l'articulation postérieure, par le contrôle des différents sphincters pharyngo-laryngés à un stade précoce de développement précédant le recul de la base de langue (Esling, 2012). Il convient de noter l'adéquation de leur théorie avec les observations du comportement laryngé dans de nombreuses langues du monde, ainsi que dans des productions paralinguistiques. (nous renvoyons ici aux nombreux travaux sur le sujet de Moisik, Esling et autres auteurs)

### **3. Les consonnes**

Nous ne traiterons ici que des consonnes occlusives et affriquées

#### **3.1. Les consonnes occlusives ou plosives**

Les consonnes occlusives diffèrent des autres types de consonnes (fricatives, liquides...) par une phase d'occlusion sur un lieu d'articulation. L'occlusion provoque une augmentation de pression en arrière de l'obstacle, qui, une fois relâché entraîne un bruit de plosion ou burst très court (10ms) caractéristique.

D'un point de vue articulatoire, il existe une phase d'occlusion par un articulateur mobile, suivi d'une tenue puis un relâchement ou l'air est rapidement expulsé.

Le lieu d'occlusion détermine les volumes en avant et en arrière de celui-ci, ce qui entraîne un certain nombre de modifications sur le spectre acoustique du burst, et des transitions formantiques qui sont des indices acoustiques importants au niveau de la reconnaissance de ces consonnes.

Ces consonnes existent dans la totalité des langues du monde (König, 1994).

On peut diviser les consonnes plosives en deux grands types :

\_ les consonnes plosives voisées ou sonores (exemple : /b/,/d/,/g/) en langue française), la vibration des plis vocaux existe pendant le plus souvent pendant la totalité de la production de la consonne (pourcentage de voisement, variable selon les contextes phonétiques), avec un spectre acoustique retrouvant le voisement précédant le burst (Figure 1). La pression intra orale dans ce cas n'atteint pas la pression sous glottique afin de permettre la vibration des plis vocaux. La phase d'occlusion est plus courte pour ces consonnes par rapport aux consonnes sourdes (Abdelli-Beruh, 2004).

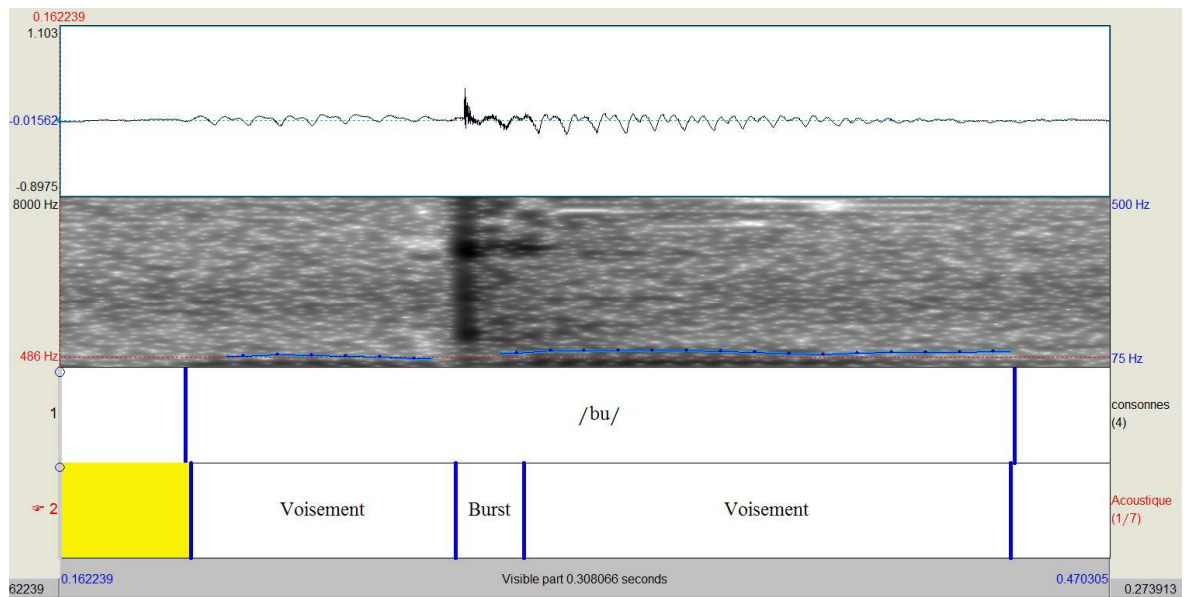


Figure 1 : Signal acoustique de /bu/ et son spectrogramme.

\_ les consonnes sourdes (exemple : /p/, /t/, /k/ en français), avec un silence précédant le burst, un bruit court de plosion correspondant à la phase de relâchement (Figure 2). Au niveau des plis vocaux, il survient un délai pour la reprise du voisement (VOT+) avec une phase de transition formantique (les articulateurs sont en mouvement vers la cible vocalique, les formants ne sont pas stables). La pression intra-orale est ici plus forte que pour les consonnes sonores (Lubker & Parris, 1970). Le volume en arrière de l'occlusion semble plus petit avec un larynx qui est élevé et une langue en position plus reculée que pour les consonnes correspondantes voisées (J. R. Westbury, 1983).

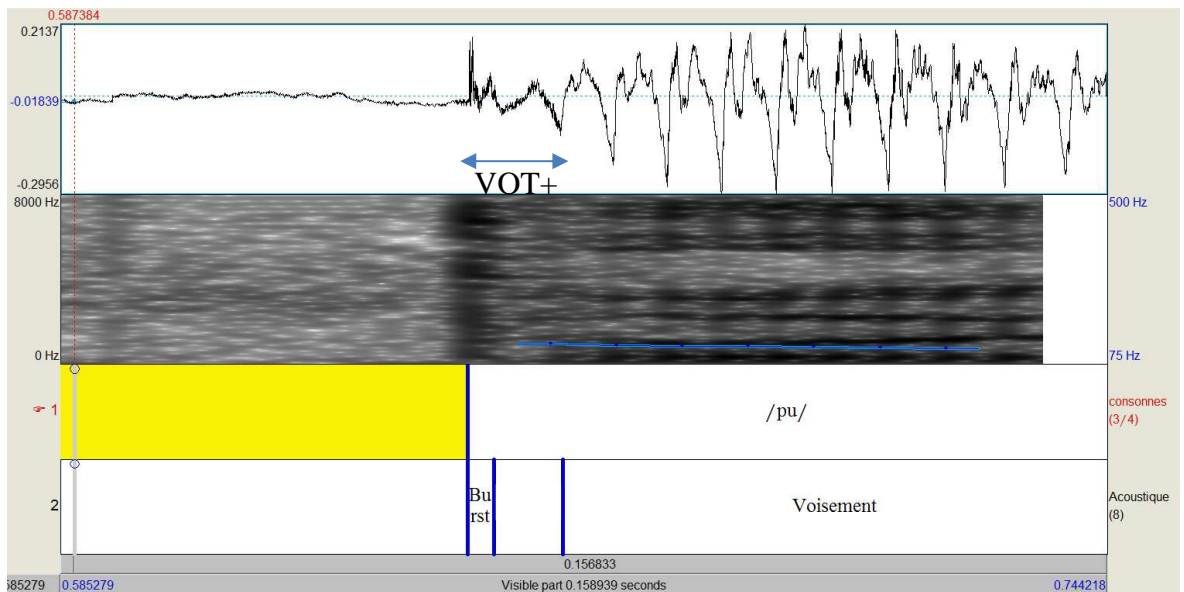


Figure 2 : Signal acoustique de /pu/ et spectrogramme.

Ces consonnes sonores ou sourdes sont dites pulmoniques car le flux d'air est d'origine pulmonaire.

Dans le cas d'une occlusion laryngée concomitante avec coup de glotte on parle de consonnes glottalisées. La montée laryngée simultanée, glotte fermée, provoque un flux non pulmonaire par mouvement de piston (appelé parfois flux d'origine glottique) créant une consonne éjective. A l'inverse, son abaissement, glotte fermée, entraînant un flux d'air entrant, est caractéristique des consonnes injectives.

Le délai de reprise du voisement est déterminé par :

- \_ la durée de tenue de l'occlusion.
- \_ par la vitesse de relâchement de l'occlusion ou vitesse articulaire (Hardcastle, 1973)
- \_ la différence entre la pression intraorale et la pression sous-glottique (Ohala & Riordan, 1979).
- \_ l'importance de la surface de contact articulaire (Stevens, Manuel, & Matthies, 1999)
- \_ ainsi que le lieu d'articulation et la voyelle environnante.

Les différents lieux d'articulation et les consonnes correspondantes sont reportés dans les tableaux ci-dessous.

### the international phonetic alphabet (2005)

consonants (pulmonic)	LABIAL		CORONAL				DORSAL				RADICAL		LARYNGEAL
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Alveolo-palatal	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal
Nasal	m	ɱ	n			ɳ	ɲ		ŋ	ɴ			
Plosive	p b		t d			ʈ ɖ	c ɟ		k g	q ɢ			
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ɕ ɟ̟	x ɣ	χ ʁ	ħ ʕ	ħ̥ ʕ̥	h ɦ	
Approximant		ʋ	ɹ			ɻ	j		ɰ				
Tap, flap		ⱱ	ɾ			ɽ							
Trill	ʙ		r							ʀ		ʀ̤	
Lateral fricative			ɬ ɮ			ɮ̠	ɬ̺		ɮ̺				
Lateral approximant			l			ɭ	ɭ̺		L				
Lateral flap			ɭ			ɭ̠							

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured *h*. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

### CONSONANTS (NON-PULMONIC)

Anterior click releases (require posterior stops)	Voiced implosives	Ejectives
⊙ Bilabial fricated	ɓ Bilabial	' <i>Examples:</i>
Laminal alveolar fricated ("dental")	ɗ Dental or alveolar	p' Bilabial
! Apical (post)alveolar abrupt ("retroflex")	ɟ Palatal	t' Dental or alveolar
‡ Laminal postalveolar abrupt ("palatal")	ɡ Velar	k' Velar
Lateral alveolar fricated ("lateral")	ɠ Uvular	s' Alveolar fricative

Figure 3 : lieux d'articulation des consonnes occlusives et éjectives des langues du monde. (IPA)

### 3.2. Consonnes affriquées

Les consonnes affriquées sont constituées d'une phase occlusive comme les plosives suivies d'une phase fricative, où l'air est relâché par une ouverture étroite provoquant un bruit de friction caractéristique. Il y a un seul lieu articulaire pour les deux

phases à la différence d'une double articulation qui possède deux lieux d'articulation différents (cf figure 4).

AFFRICATE LIGATURES	
dz = dz̥	ts = ts̥
dʒ = dʒ̥	tʃ = tʃ̥
dʒ̥ = dʒ̥	tʃ̥ = tʃ̥

Figure 4 : Consonnes affriquées. (IPA)

#### 4. Mesures expérimentales articulatoires des consonnes plosives

Nous recensons ici les méthodes expérimentales utilisées pour l'exploration articulatoires des consonnes.

##### 4.1. Les signaux acoustiques (Calliope 1989)

L'analyse à partir des signaux acoustiques des consonnes plosives permet de déterminer les lieux d'articulation. De multiples indices acoustiques ont été retrouvés permettant la distinction des consonnes et de leur articulation. Les consonnes dont les plosives sont en général considérées comme moins sujettes à la variation phonétique comparées aux voyelles en réalisation standard.

L'intensité du burst, c'est-à-dire du bruit de plosion initial, permet une bonne discrimination des consonnes plosives de l'anglais : si l'on en modifie l'amplitude énergétique spectrale, cela entraîne des difficultés de reconnaissance (Kapoor & Allen, 2012)



Nous avons relevé dans la partie précédente le Voice Onset Time (VOT ou délai de voisement) décrit par Lisker (Lisker & Abramson, 1964), qui correspond au délai entre le relâchement de l'occlusion et la reprise de vibration des plis vocaux, toujours positif pour les consonnes plosives sourdes selon cette définition. Une autre définition (Klatt, 1975) qui est le délai entre le relâchement et le début de la structure formantique, rend possible d'identifier un VOT pour les consonnes plosives et fricatives.

En situation normale, ces deux indices semblent être les plus importants pour la discrimination des consonnes plosives.(Francis, Kaganovich, & Driscoll-Huber, 2008)

Les moments spectraux, c'est-à-dire les moments mathématiques calculés sur le spectre de puissance du burst (traité comme une loi de probabilité), liés aux différences de volume créées par l'occlusion ont été retrouvés comme indices acoustiques des lieux d'articulation, parmi eux nous pouvons citer le centre de gravité spectrale (équivalent de la moyenne, il serait corrélé négativement avec la longueur de la cavité de résonance en amont du lieu d'occlusion/constriction), le coefficient d'aplatissement (kurtosis : le fait que l'énergie soit concentrée sur une bande de fréquence plus ou moins large), et les coefficients d'asymétrie (skewness : répartition de l'énergie du spectre par rapport à la moyenne). (Chodroff & Wilson, 2014; Forrest, Weismer, Milenkovic, & Dougall, 1988)

Une consonne C produite entre deux voyelles V (VCV) implique des transitions formantiques visibles sur les phases d'attaque ou de terminaison des voyelles et sont liées à la mise en mouvement des articulateurs. Ces transitions, très étudiées, sont de bons moyens de reconnaissance des lieux d'articulation.(Sussman, Fruchter, & Cable, 1995; Sussman, Hoemeke, & Ahmed, 1993)

Les signaux acoustiques sont très souvent mis en relation avec les moyens d'exploration traités ci-dessous.

## ***4.2. Evaluation de l'articulation bucco-linguale***

### **4.2.1. La palatographie**

La statopalatographie est une méthode ancienne d'étude de la parole, réalisée à partir d'encre ou équivalent, permettant de définir le lieu d'articulation par la mesure du contact de la langue avec le palais. Cette méthode ne permet pas d'appréhender la dynamique du geste articulatoire.

Elle fut ensuite modernisée par l'électropalatographie (EPG) qui consiste à mesurer les contacts entre la langue et le palais au moyen d'un faux palais muni d'électrodes. Cette méthode permet une analyse dynamique des mouvements linguaux (cf figure 5). Elle a comme inconvénient principal de gêner le locuteur par l'épaisseur du faux palais et doit être réalisé sur mesure avec un coût non négligeable.(Brunner, Fuchs, & Perrier, 2005)

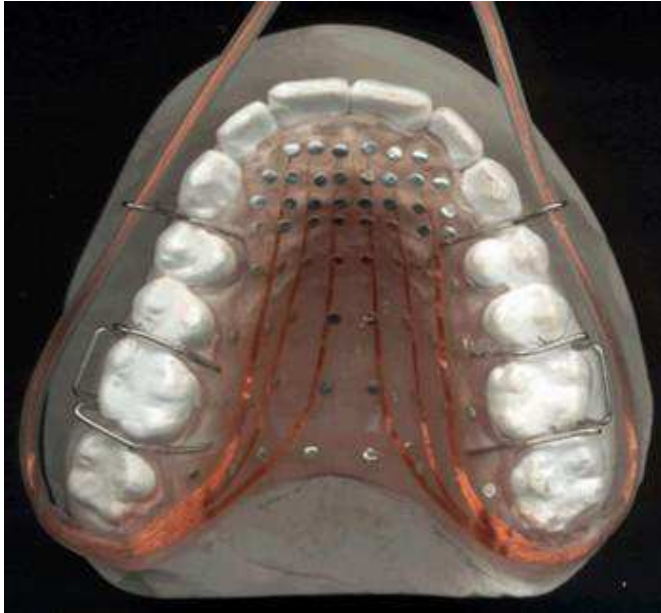


Figure 5 : faux palais d'électropalatographie. (image trouvée sur le site du Laboratoire de Phonétique et Phonologie (CNRS Sorbonne Nouvelle)

#### **4.2.2. Les techniques radiologiques**

Les techniques de radiographies ou vidéofluoroscopie ont été utilisées dès les années 1920 pour l'étude dynamique de la parole (Russel, 1928). Elles permettent de repérer les différentes structures dures (cartilages laryngés, os hyoïde, crâne) ainsi que les ombres projetées par les structures molles (langue, replis, pharynx...) sur la globalité du tractus vocal. Elle n'est actuellement plus usitée pour l'étude phonétique de la parole en raison de l'irradiation induite sur le sujet (figure 6).

Un autre inconvénient est l'impossibilité de déterminer précisément les mouvements des structures dans le plan frontal à partir de mesures prises dans le plan sagittal. (Branderud, Lundberg, Lander, & Djamshidpey, 1998).

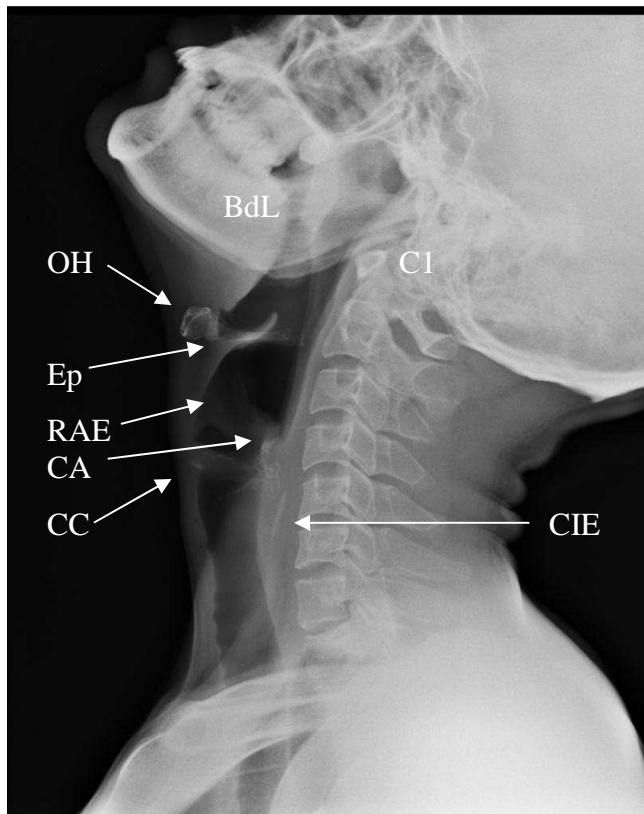


Figure 6 : Exemple de radiographie cervicale dans le plan sagittal prise dans un contexte médical (CHU de Grenoble). Les structures annotées sont : base de langue (BdL), os hyoïde (OH), épiglotte (Ep), cartilage cricoïde (CC), Repli aryépiglottique (RAE), cartilages aryénoïdes (CA) Vertèbre cervicale C1, corps étranger intra oesophagien (CIE, arrête de poisson).

Le X Ray Microbeam (micro-faisceau de rayons X) permet l'étude des mouvements de la langue, des joues et des lèvres en suivant les trajectoires de quelques pastilles d'or fixées, mais l'inconvénient de l'irradiation persiste bien que les doses irradiantes soient moins importantes. (Westbury, Milenkovic, Weismer, & Kent, 1990)

#### 4.2.3. Imagerie par Résonance Magnétique (IRM)

L'Imagerie par Résonance Magnétique n'utilise pas de rayons ionisants mais un champ magnétique puissant permettant de mesurer le signal émis par les atomes d'hydrogène lors de leur excitation par une onde radio.

C'est le seul examen qui permet d'étudier en trois dimensions et dans sa totalité le conduit vocal sans effet secondaire. Il est de plus en plus utilisé mais reste un examen difficile d'accès et onéreux. De plus sa vitesse d'acquisition est encore faible (20 images par seconde en IRM « classique »), bien que l'apparition d'IRM plus puissantes et de

nouveaux types de séquences d'acquisition ouvrent un champ de possible (Patil, Greer, Blaylock, & Narayanan, 2017a).

Un autre problème est la nécessité pour le sujet d'être positionné en décubitus dorsal, ce qui provoque un recul des structures vocales comme la base de langue.

Enfin la machine d'IRM étant extrêmement bruyante, il est difficile de réaliser simultanément des enregistrements acoustiques exploitables. A noter que certaines équipes développent des dispositifs de réduction du bruit.

#### **4.2.4. Echographie**

L'échographie utilise le principe de réflexion des ondes ultrasonores lorsque le milieu change d'impédance pour donner des images en deux dimensions des structures étudiées. Nous ne parlerons ici que de l'échographie en mode B, c'est-à-dire de la reconstruction d'une image bidimensionnelle échographique.

L'application de cette technique dans le domaine de l'étude de la parole a commencé dans les années 1980 (Stone, 1983) (STONE, 1983). L'absence de référence spatiale absolue et le risque de perturbation des données par les mouvements de translation ou de rotation de la sonde échographiques obligent le sujet à une stricte immobilité dans un casque rigide. L'échographie est très utilisée pour l'analyse des mouvements linguaux. Le développement d'algorithmes de suivi et de reconnaissance des contours linguaux (EdgeTrak par exemple) ont permis l'analyse semi-automatique des données. En dehors de l'immobilité du sujet qui peut être gênante dans l'étude des mouvements linguaux en condition réelle (exemple du Beatbox : le chanteur accompagne le plus souvent ses productions orales par des mouvements des mains mais aussi de la tête et de la mâchoire pour la production de certains effets), l'échographie donne des images dans le plan de la sonde, et ne permet donc pas l'évaluation des plans orthogonaux ou adjacents. De plus il n'est pas possible de suivre un point de chair précisément.

Le système HOCUS (pour Haskins Optically Corrected Ultrasound System) (Whalen et al., 2005) permet de s'affranchir de l'immobilité entraînée par le casque rigide en corrigeant les données par la position relative de la sonde par rapport à la tête par un suivi optique. Le système Optotrack effectue une correction de la position de la sonde par rapport aux structures externes céphaliques.

L'échographie peut être couplée avec l'Articulographie ElectroMagnétique (Aron, Kerrien, & Laprie, 2007).

Enfin l'échographie est aussi utilisée dans l'évaluation des mouvements laryngés par certains auteurs (Esling & Moisik, 2011). Il reste cependant le problème de la position mobile de la sonde car tenue par l'opérateur, il n'existe pas actuellement de dispositif garantissant son immobilité. De plus les images échographiques laryngées sont d'une qualité relativement faible en raison de la présence de l'air dans le conduit vocal et dans l'espace ventriculaire.

#### **4.2.5. Articulographie Electro Magnétique (EMA)**

Le principe de l'EMA est simple : des bobines sont placées sur différents points de la langue et des structures étudiées (joues, lèvres...) et le courant induit selon leur place dans le champ électromagnétique environnant permet de déterminer leur position en trois dimensions (Schönle et al., 1987). Bien que les premiers systèmes soient apparus dans les années 1990 (Perkell et al., 1992), la commercialisation et l'utilisation de l'EMA date de la fin des années 2000.

L'avantage de cette méthode exploratoire est la mesure des mouvements dans les trois plans de l'espace, ainsi que sa rapidité d'acquisition permettant un suivi temporel très fin et la possibilité d'acquérir des données sur plusieurs articulateurs simultanément.

L'EMA reste cependant sensible aux mouvements parasites du sujet et la présence de fils et de bobines intraorales peuvent gêner la production de certains sons.

#### **4.3. L'exploration laryngée**

Nous avons déjà recensé ci-dessus l'IRM, les méthodes radiologiques et l'échographie qui peuvent être utilisées dans ce cadre.

##### **4.3.1. L'endoscopie**

L'étude pharyngo-laryngée peut se faire par deux types d'endoscopie dans le cadre de l'étude de la parole :

\_ l'endoscopie rigide, nécessitant l'ouverture buccale, et une certaine position avancée de la langue pour une bonne évaluation laryngée. Les endoscopes rigides permettent une visualisation de qualité de par leur taille de prisme et de par leur éclairage

des structures. Ainsi on peut utiliser les caméras ultra-rapides pour la visualisation des mouvements glottiques par exemple.

Cependant la position est relativement inconfortable, invasive et gêne la production normale de la parole puisque la fermeture buccale ou un recul de la base de langue empêche la visualisation des structures en amont (figure 7).

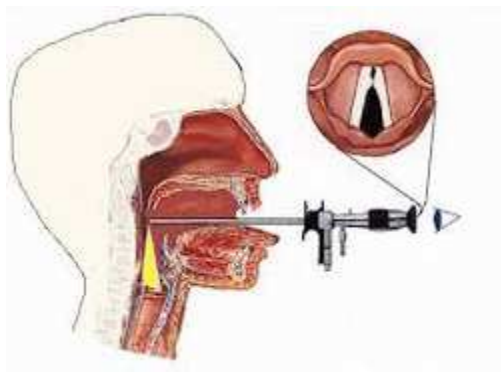


Figure 7 : principe de la pharyngoscopie rigide. (traité d'ORL)

\_ la naso-fibroscopie avec un endoscope souple (figure 8) : passé par les fosses nasales au-dessus du velum, il permet une vue en plongée sur le pharyngolarynx. Cette technique permet la production de parole sans gêner les articulateurs.

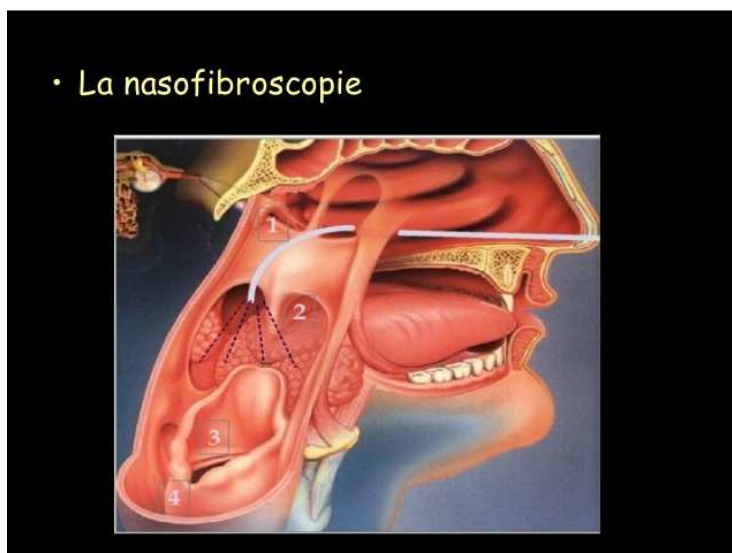


Figure 8 : Principe de la nasofibroscopie souple. (traité d'ORL)

En revanche, les systèmes actuels utilisés sur le marché sont limités par la taille des cellules optiques ne permettant pas un éclairage optimal, ce qui peut gêner pour l'acquisition de séquences d'images à haute vitesse. De plus le fait que l'endoscope souple repose sur le velum, les mouvements de celui-ci rendent difficile l'analyse des

mouvements dans l'axe cranio-caudal et donne une certaine instabilité de position aux images.

La stroboscopie peut être réalisée avec l'un ou l'autre des endoscopes et consiste en l'enregistrement d'images à une fréquence très légèrement différente de la fréquence de phonation permettant une reconstruction artificielle des cycles vocaux. En effet l'enregistrement des images sur des temps décalés d'un peu moins (ou plus) d'une période sur des cycles vocaux consécutifs permet une visualisation artificiellement « ralentie » de l'ondulation muqueuse. Cependant elle nécessite une vibration périodique et ne peut être utilisée dans le cadre d'étude des consonnes.

La cinématographie ultrarapide est coûteuse en terme de matériel (caméras) (Granqvist & Lindestad, 2001), et de stockage. Son indépendance par rapport au signal acoustique par rapport à la stroboscopie en fait une méthode de plus en plus utilisée et vouée à se développer permettant la visualisation de certains mouvements de structures laryngés trop rapides pour l'enregistrement en vidéo-endoscopie classique (Deliyski, Hillman, & Mehta, 2015; Hassan, Esling, Moisik, & Crevier-Buchman, 2011). Nous avons déjà relevé les difficultés pour l'utiliser en nasofibroscopie souple, mais l'utilisation de caméras ultra-rapides très sensibles permet de contourner cet inconvénient (Kaneko, Watanabe, Kihara, Sakaguchi, & Takahashi, 2017). La vitesse d'acquisition d'au moins 4000 images par seconde est recommandée par Deliyski (Deliyski, Powell, Zacharias, Gerlach, & de Alarcon, 2015) ce qui nécessite des algorithmes d'analyse automatique des images.

La vidéokymographie développée par Svec et Schutte (Švec & Schutte, 1996), ne sélectionne qu'une bande visuelle de l'image complète et crée un empilement afin de créer un kymogramme représentant l'évolution au cours du temps de la coupe sélectionnée. Elle permet de visualiser facilement les aperiodicités, mais est limitée par la perte de vue de la globalité des structures. Elle est le plus souvent réalisée en cinématographie ultra-rapide pour une meilleure précision. Le vidéokymogramme peut aussi être reconstruit à partir des images stroboscopiques.

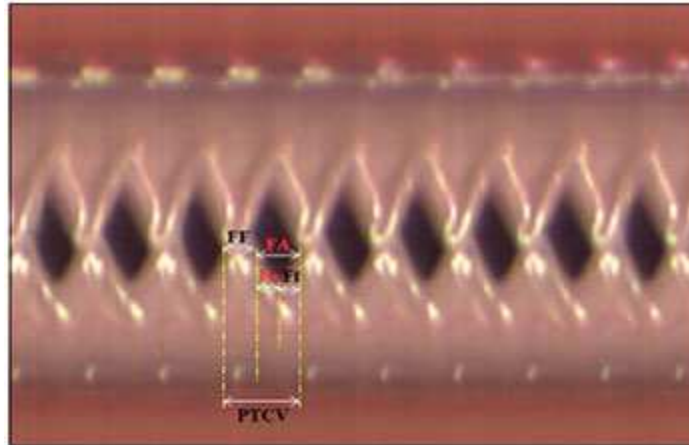


Figure 9 : Vidéokymogramme d'oscillation glottique.

#### 4.3.2. Autres systèmes optiques

L'illumination de l'étage sous glottique et l'enregistrement de la luminosité résiduelle sus glottique en fonction des états laryngés s'appelle la photoglottographie. La variation de luminosité est proportionnelle à l'aire glottique et permet de donner indirectement l'état d'ouverture glottale (Habermann, Jiang, Lin, & Hanson, 2000). Initialement l'illumination se faisait par voie transcutanée directement dans la trachée ou par une source descendue dans le pharynx puis sous l'étage glottique (Löfqvist & Yoshioka, 1980). Il existe de nouveaux dispositifs par voie externe pure pour la transillumination et la récolte de la luminosité. (Ridouane, Audibert, & Minh Nguyen, 2012; Suthau, Birkholz, Mainka, & Simpson, 2016)

#### 4.3.3. Mesures électromyographiques

L'électromyographie laryngée nécessite l'implantation d'aiguilles électrodes par voie transcutanée dans les muscles à étudier. En général sont étudiés les muscles intrinsèques laryngés à travers la membrane crico-thyroïdienne mais le même type de dispositif peut être utilisé pour l'étude des muscles extrinsèques du larynx. Sont enregistrés les potentiels d'action musculaires et ainsi le recrutement des muscles et leur relâchement. En raison de l'implantation des électrodes, ce dispositif est assez invasif dans le cadre de l'étude de la parole.

L'électromyographie de surface nécessite la mise en place d'électrodes cutanées en regard du muscle à étudier pour l'enregistrement indirect de sa contraction. Le nombre de muscles ainsi que leur disposition au niveau cervical ne permettent pas de rendre compte



précisément de l'activité volontaire unique d'un seul muscle contrairement à l'implantation d'électrodes. L'interprétation des enregistrements est donc délicate et très approximative.

#### **4.3.4. L'Electroglottographie**

L'électroglottographie (EGG) mise au point par Philippe Fabre (Fabre, 1957) est un moyen non invasif d'exploration glottique : des électrodes cutanées disposées en regard de la glotte permettent de mesurer les variations de potentiel électrique consécutives aux mouvements glottiques, permettant de déterminer la mise en contact des plis vocaux (Marasek, 1996; Orlikoff, 1991). Cependant elle est limitée par l'impossibilité de décomposer l'évolution du contact des plis dans le plan horizontal (axe antéro-postérieur) et vertical (axe inférieur-supérieur). Si les fréquences rapides de variation du potentiel sont très étudiées et correspondent à la vibration (contact et relâchement des plis vocaux), concernant les variations amples et lentes recueillies, on ne peut que poser l'hypothèse de changements d'état dans les autres plans (notamment vertical) ou la contraction d'autres groupes musculaires du voisinage (extrinsèques ou intrinsèques laryngés) faisant varier les potentiels. Des dispositifs avec un certain nombre d'électrodes empilées ou disposées verticalement sont supposées permettre le suivi dans le plan vertical du larynx (Kob & Frauenrath, 2009).

#### **4.3.5. Mesures aérodynamiques indirectes**

Les mesures de débit d'air et de pression atmosphérique/intraorale par différents systèmes (système d'Evaluation Vocale Assistée (EVA) par exemple )sont simples à mettre en œuvre (Mamadou M. et al., 2011). Certains systèmes renseignent de manière indirecte sur le comportement laryngé en estimant l'Onde de Débit Glottique dans le cadre de l'exploration de la vibration des plis vocaux.

Les débits d'air oral et nasal ont été étudiés pour la production des consonnes plosives et nasales montrant une certaine variation en fonction de l'âge, de lieu d'articulation et de la présence ou non de voisement (Demolin, 2004; Trullinger & Emanuel, 1983).

La pression intra-orale (PIO) enregistrée au moyen d'une sonde placée dans le pharynx permet de donner une bonne évaluation de l'effort vocal nécessaire à la production des consonnes plosives.(Prosek, 1975) De plus elle est une très bonne approximation de la pression sous-glottique enregistrée de manière invasive par un cathéter

en position sous glottique pour les consonnes plosives non voisées (Hertegård, Gauffin, & Lindestad, 1995; Löfqvist, Carlborg, & Kitzing, 1982).

Il est considéré qu'un différentiel de pression transglottique (sous-glottique vs intraorale) doit être d'au moins 2 hPa pour permettre un voisement.

L'avantage du système EVA2 cité au-dessus, est la possibilité d'enregistrement simultané de la PIO, des débits d'air nasal et oral, de l'électroglottographie et du signal acoustique, de façon simple et rapide.

#### ***4.4. Conclusion sur les méthodes exploratoires***

Le recensement de ces différentes méthodes d'exploration de la production de la parole et la mise en exergue de leurs avantages et limites met en évidence la complémentarité de ces examens dans l'étude de la parole. La production vocale résultant d'une interaction complexe entre gestes respiratoires, phonatoires et articulatoires, il y a un très grand intérêt à combiner ces nombreuses méthodes pour une meilleure compréhension d'une production vocale donnée.

### ***5. Le Human Beatbox : état de l'art***

#### ***5.1. Un art vocal récent***

Le Human Beatbox est défini comme l'art de production des sons rythmés avec la bouche (Martino, 2009).

L'imitation des percussions par la voix humaine est ancienne et existe dans de nombreuses cultures : les premières formes de musique jouées avec la bouche sont anciennes en Occident (il existe des témoignages du XVe siècle). Certaines musiques rituelles africaines ou indiennes (*bol*) utilisent aussi ce genre de production rythmique mais sans essayer de cacher leur origine vocale humaine (Atherton, 2018). A partir des « années folles » avec l'apparition du be-bop et du *scat*, les chanteurs de jazz ont essayé d'imiter les mélodies instrumentales uniquement par des onomatopées. Toutefois il n'est pas relevé d'autre courant musical utilisant de façon simultanée des sons percussifs et d'autres produits par les plis vocaux (Lebrun, ethnomusicologue, Communication sur ArteRadio 2010).

Initialement né aux Etats-Unis au sein du mouvement hip-hop afin de remplacer les boîtes à rythme électroniques onéreuses, il se développe rapidement dans le pays en tant que technique à part entière, porté notamment par le groupe The Fat Boys dans les années

80. Les premiers sons simples de type percussifs s'enrichissent petit à petit d'imitations de sons électroniques, de scratch, et de sons inventés selon les cultures et styles musicaux. Le premier album composé quasi intégralement de Beatbox de Razhel en 1999 marque l'essor international de la technique et inspire toute une génération telle que Killa Kela au Royaume-Uni ou les Saïan Supa Crew en France. En plus d'un répertoire de sons reproduits extrêmement large, il est le premier à utiliser la technique du « humming », c'est-à-dire d'arriver à produire le rythme (« beat ») tout en chantant simultanément.

L'apparition d'Internet au début des années 2000 dans les foyers, marque un tournant pour la discipline en démocratisant le beatbox auprès de milliers de personnes, qui, fascinées par les performances des beatboxeurs s'essayent à ces techniques. En effet la communauté des beatboxeurs est extrêmement présente et active sur le nouveau média, partageant par exemple des vidéos pédagogiques sur leurs différentes techniques (Martino, 2009). De même des artistes internationaux (comme Bjork pour l'ouverture des Jeux Olympiques d'Athènes en 2004) invitent des beatboxeurs sur scène et participent à la popularisation de cet art. Enfin un championnat mondial est organisé tous les trois ans depuis 2005, ainsi que différents championnats nationaux annuels, libérant la discipline de la vision initiale d'accompagnement musical.

Le Human Beatbox démocratisé s'affranchit peu à peu de son origine hip-hop pour devenir un art vocal à part entière, qui s'adapte à l'envie de l'interprète, de son style musical et de la recherche quasi quotidienne de nouveaux sons par ses pratiquants. C'est pourquoi on peut noter un certain nombre de courants correspondant aux générations d'interprètes :

\_ de style « classique », les pères du Beatbox, qui ont participé à son élaboration, parmi lesquels nous pouvons citer un certain nombre de chanteurs français (Tiko, le groupe UnderKontrol, Ezra, L.O.S,...), musicalement ancrés dans la culture hip-hop, avec des sons de type percussifs, instrumentaux et de scratches.

\_ de style « moderne », plus jeunes, qui adaptent leurs performances aux goûts musicaux d'aujourd'hui dans un univers plus tourné vers la musique techno et des sons électroniques.

De même, au sein de chaque courant, peuvent se retrouver des artistes plutôt portés sur la technique et la recherche de nouveaux sons ou d'autres plus portés sur le côté mélodique et musical.

Le Beatbox est ainsi devenu en quelques années un véritable art populaire rassemblant des milliers d'interprètes amateurs ou professionnels, qui sont de véritables passionnés de la discipline et la pratiquent au quotidien, de par sa facilité (un seul instrument : soi-même) et de par l'étendue de ses possibilités...

## ***5.2. Etude scientifique du Beatbox***

Les performances vocales multiples des beatboxeurs en font une source importante d'information pour la recherche dans de nombreux domaines.

### **5.2.1. Aspect paralinguistique**

Le beatboxeur cherche à imiter le plus fidèlement possible un son en partant quasi systématiquement d'un son du langage, puis en supprimant les indices acoustiques linguistiques qui pourraient donner à l'auditeur, l'impression que ce son est produit par une voix humaine (Stowell & Plumbley, 2008). Le Beatbox serait donc une utilisation paralinguistique de patterns phonétiques associés aux langues humaines, mais modifiés afin de créer un véritable langage musical à part entière et cohérent. C'est pourquoi il est suspecté une certaine difficulté pour le beatboxeur à se détacher des patterns articulatoires de sa langue native pour l'imitation de sons proches des phonèmes de sa langue (Lederer, 2005). Garrigues a confirmé la présence de sons n'appartenant pas au système phonologique natif des interprètes (Garrigues, 2015).

Ainsi, différents auteurs proposent une formalisation de notation des effets du Beatbox par une notation basée sur l'Alphabet Phonétique International(API) de par leur proximité avec les sons des langues du monde (Sinyor, Rebecca, Mcennis, & Fujinaga, 2005; Stowell & Plumbley, 2008). D'autres systèmes de notation sont utilisés dont la notation standardisée du Beatbox, plus simple que la formalisation par l'API, elle aussi basée sur une certaine ressemblance avec les consonnes de l'anglais (TyTe & Splinter, 2002).

Récemment, un système de notation a été imaginé par un graphiste, Adrien Contesse. Dénommé « VocalGrammatics », il s'inspire de la phonétique articulatoire pour présenter visuellement les effets du Beatbox par des pictogrammes, permettant une meilleure compréhension par l'apprenant (Contesse, 2015).

Les discussions avec nos différents sujets pratiquant le Beatbox nous apprennent qu'il existe une certaine variation pour les mêmes effets selon les pays d'origine du

beatboxeur : par exemple, l'effet grosse caisse (kick) serait voisé pour les pratiquants anglais à la différence des beatboxeurs d'autres origines. Cependant il n'a pas été retrouvé d'étude sur la variation de ce véritable langage musical.

### **5.2.1. Aspect acoustique et phonétique**

Les beatboxeurs ont une efficacité vocale remarquable, avec une gestion des paramètres acoustiques et aérodynamiques comparables voire supérieures à la voix criée. L'énergie du signal acoustique se concentre au moment du relâchement de la consonne avec une absence de voisement et une répartition spectrale sur toutes les fréquences et non en décalé comme dans les autres modalités (parole, voix chantée, criée, projetée) (Bourdin & Navion, 2013). Il n'a en revanche pas été montré de franche démarcation du Beatbox par rapport aux autres modalités sur différents moments spectraux : centre de gravité, coefficients d'asymétrie et d'aplatissement. La pression intra-orale (PIO) n'est pas non plus corrélée à l'intensité des plosions, à la différence de la dérivée du débit d'air oral qui est majorée dans cette modalité (Paroni, 2016).

Les techniques vocales employées pour en arriver à ces résultats sont extrêmement nombreuses et nécessitent l'emploi de la totalité des structures du tractus vocal (vibrateurs, résonateurs, articulateurs) et des organes associés (respirateurs) avec une coordination forte entre ceux-ci tout en ayant une indépendance des structures, inhabituelle dans leur usage quotidien à vocation de communication.

Il a ainsi été montré que le beatboxeur catégorise les sons qu'il produit en référence aux sons de la parole mais la conformation de ses structures vocales est dans un certain sens « extraordinaire » : par exemple l'étude des mouvements du pharyngo-larynx par nasofibroscopie retrouve une tubulisation de l'épiglotte pour les sons instrumentaux imitant la trompette ou la guitare (de Torcy et al., 2014a). Ces nombreuses conformations laryngées permettant la production d'effets instrumentaux divers, montre l'importance des structures du pharyngo-larynx dans le Human Beatbox à la manière des combinaisons articulaires laryngées décrites dans le modèle de l'articulateur laryngé de Esling et Moisik.

Chez le chanteur professionnel, les modifications même mineures du conduit vocal, entraînées par les mouvements des structures pharyngo-laryngées, influencent de façon importante sur le résultat acoustique produit (Echternach, Sundberg, Markl, & Richter, 2010; Kitamura, Honda, & Takemoto, 2005). Certaines similarités de changement de

forme du conduit pharyngo-laryngé, par l'utilisation notamment des muscles constricteurs du pharynx, ont été relevées et semblent expliquer le rendement vocal important des sons percussifs du Beatbox (Sapthavee, Yi, & Sims, 2014).

Il convient de citer ici l'étude de Proctor et al. qui a étudié les mouvements des articulateurs pour certains effets du Beatbox, par Imagerie par Résonance Magnétique (IRM), dans un plan médio-sagittal qui montre des mouvements de piston du larynx fermé, caractéristiques des consonnes éjectives (ces mouvements verticaux avaient été relevés en nasofibroscopie par de Torcy et Clouet (2010)). Toutefois la vitesse d'acquisition des images de l'IRM reste faible et cette étude n'a été réalisée que sur un seul sujet (Proctor, Bresch, Byrd, Nayak, & Narayanan, 2013). Une étude récente par IRM à séquences d'acquisition rapides, a comparé la production d'effets du Beatbox entre des sujets experts et des amateurs retrouvant les mécanismes éjectifs déjà montrés par Proctor et al., avec un certain nombre de variations articulatoires pour un effet similaire. Il semble aussi que les experts ont un meilleur contrôle dans la production de ces éjectives, notamment lié à l'effet stylistique recherché que pour les amateurs (Patil et al., 2017a).

L'exploration échographique de l'articulation linguale des effets du Beatbox a permis de démontrer par une analyse kymographique des images la précision et le contrôle extrêmement fin des gestes linguaux (Paroni, 2014).

Les mesures aérodynamiques accompagnées des analyses acoustiques et de l'électroglottographie (EGG) retrouvent une gestion très particulière de la respiration du beatboxeur lors de son interprétation en Beatbox avec des sons non pulmonaires avec mécanismes égressifs et ingressifs mais aussi de la phonation inversée leur permettant de garder le rythme dans leur « beat » (ou phrase rythmique) tout en reprenant leur respiration (Garrigues, 2015).

Il a été relevé dans nos recherches bibliographiques la création d'une interface d'apprentissage automatique appelée Beatbox permettant à l'utilisateur de classifier ses propres productions (Hipke, Toomim, Fiebrink, & Fogarty, 2014), ainsi que la création d'algorithmes de reconnaissance automatique (Picart, Brognaux, & Dupont, 2015).

### ***5.3. Son potentiel thérapeutique***

Comme nous l'avons souligné dans le paragraphe précédent, les sons percussifs produits dans le Beatbox ont un rendement vocal très important malgré un comportement laryngé à risque : certaines conformations du larynx, ainsi que les mécanismes de

phonation inversée associés pourraient rapidement provoquer un malmenage vocal jusqu'à des traumatismes et lésions des plis vocaux. De plus les pratiquants de cet art sont souvent d'infatigables travailleurs qui peuvent répéter pendant des heures quotidiennement. Hormis quelques désagréments subjectifs comme une fatigue vocale ou quelques légers troubles gastro-oesophagiens (reflux ressentis), et malgré des conduites à risque pour la voix (tabagisme, consommation régulière d'alcool et de cannabis), il n'a pas été relevé de pathologie vocale handicapante. Saphavee pose l'hypothèse que la présence et l'association de la contraction pharyngée aux mouvements du larynx, l'utilisation de la totalité des structures du conduit vocal et la présence de voisement en adduction des cordes sont des facteurs protecteurs concernant la santé vocale (Saphavee et al., 2014).

Il semble ainsi important de comprendre d'un point de vue fondamental les raisons et le rationnel de cette absence de malmenage vocal. De plus le côté ludique du Beatbox et les possibilités de travail sur la totalité du conduit vocal en feraient un outil de choix pour l'élaboration de techniques de rééducation orthophonique, dont certains exercices répétitifs et rébarbatifs sont difficilement suivis et réalisés par les enfants par exemple. Enfin l'utilisation des consonnes éjectives plus sonores avec des mouvements ascensionnels laryngés, ainsi que les conformations pharyngées aidant potentiellement à la production de certains sons pourraient augmenter l'intelligibilité de patients ayant subi une opération chirurgicale laryngée de type laryngectomie partielle par exemple et qui ont une voix souvent soufflée, peu audible, en raison de l'absence de certains muscles laryngés, mais pour lesquels le reste des structures est globalement conservé (muscles suprahyoïdiens, constricteurs pharyngés...).

Un travail récent montre que les orthophonistes sont relativement conscients et plutôt intéressés par les possibilités qu'offrent le Beatbox en pratique clinique de rééducation de la voix (Poupineau, 2018).

## Chapitre 2. Questions de recherche

Le Human Beatbox de par la richesse des gestes vocaux laryngés et oraux impliqués, de par la variété des effets sonores, et de par la puissance vocale produite avec un minimum d'effort, nous amène à différentes questions :

**1. *Est-il possible de corréler les gestes laryngés produits pendant le Human Beatbox à des mesures aérodynamiques et acoustiques ?***

Nous posons l'hypothèse que les gestes articulatoires laryngés du Human Beatbox de par leur nature précise sous contrôle volontaire et leur dynamique permettent d'influer sur la Pression Intra Orale et le débit d'air oral, et par conséquent sur la production sonore et l'efficacité acoustique.

**2. *Quelles sont les différences entre les consonnes plosives de la parole normale et leurs équivalents en Human Beatbox tant au point de vue des gestes laryngés qu'au niveau de l'aérodynamique et de la production sonore ?***

Nous pensons que l'importance de la composante éjective/injective des effets du Human Beatbox permet principalement de distinguer les deux modalités.

**3. *Un même effet peut-il être produit par des stratégies acoustiques différentes ?***

Pour répondre à ces questions nous avons évalué les gestes laryngés à partir d'enregistrements endoscopiques, acoustiques et aérodynamiques, effectué sur un beatboxeur professionnel et un beatboxeur amateur à partir d'un corpus contenant des consonnes plosives dans la parole normale et leurs équivalents en Beatbox.



## **Partie 2**

-

## **Evaluation expérimentale**

## Chapitre 4. Matériel et méthodes

La mise en place des techniques d'exploration que nous avons utilisées est décrite dans le chapitre en annexe.

Nous présentons ici notre protocole réalisé dans le cadre de l'étude du Beatbox.

### 4. *Description des sujets*

Notre premier sujet est un beatboxeur professionnel, locuteur natif du français, plusieurs fois champion du monde avec le groupe UnderKontrol. Il a découvert la discipline dans les années 2000 au moment de son expansion internationale. Très impliqué sur la scène internationale, il organise aussi des cours de Beatbox auprès du public. Il pratique le Human Beatbox quotidiennement à raison de plusieurs heures par jour. Il se considère comme de « l'école classique », c'est-à-dire plutôt porté sur l'imitation instrumentale, des percussions et de scratches. Il n'a jamais eu de pathologie vocale, mais décrit une période avec des symptômes d'inconfort intestinal à type de reflux gastro-oesophagiens qu'il lie à son hygiène de vie de mauvaise qualité (alimentation déséquilibrée, manque de sommeil, consommation régulière d'alcool, tabagisme et abus de café).

Le deuxième sujet est un beatboxeur amateur, locuteur français natif, actuellement étudiant, qui a commencé le Beatbox il y a 5 ans. Il se produit de temps en temps dans des concerts locaux. De style « nouvelle école », il est plutôt tourné sur la production de sons de type électroniques, tant au point de vue mélodique que pour sa rythmique. Il n'a pas de comportements à risque. Il signale une certaine fatigue vocale après la pratique intensive de certains sons de type « harsh voice » et « creaky voice », et a développé une attitude de rotation de la tête pour palier cette gêne et pour produire ces sons particuliers de façon efficace.

Les sujets ont signé un consentement d'accord pour l'enregistrement et l'utilisation des données récoltées dans notre travail.

### 4. *Effets étudiés et consonnes associées*

Afin de simplifier la lecture et de pouvoir catégoriser les effets en grandes familles, nous utiliserons notre propre notation pour celles-ci, sauf mention pour les variantes qui seront en notation standard du Beatbox (SBN). Les notations en SBN sont entre accolades.

L'étude portait sur les consonnes plosives /p/, /b/, /t/, /k/ du français et les sons percussifs du Beatbox s'en rapprochant.

Nous avons étudié les effets percussifs suivants ainsi que leurs variations :

\_ grosse caisse (ou kick, {B}), noté P dans la suite. Ils sont l'équivalent en Beatbox de la consonne /p'/ dans sa version non voisée et /b/ dans sa version voisée.

\_ cymbale de type charleston (hi-hat, {t}), noté T qui correspond à un /t/ plus ou moins suivi d'une fricative comme dans l'affriquée /ts/.

\_ clave (K-snare ou Rimshot {^K}) que nous noterons K, ressemblant au /k/.

\_ caisse-claire (Psh snare {^Psh}) que nous noterons PS, qui est une combinaison d'un /P/ suivi d'une fricative /f/ /s/ ou /ʃ/ selon les productions, correspondant à la version non voisée de « p-V-s » ou de « p-V-ch » ou V est une voyelle.

## **5. Protocole expérimental**

### **5.1. Corpus**

Un corpus a été constitué à partir de phrases utilisées par les beatboxeurs pour l'apprentissage du Beatbox, leur permettant un enchaînement rythmique cohérent et facile à mémoriser. Ces phrases contiennent les consonnes plosives /p/, /b/, /t/, /k/. Les consonnes sont, en parole normale, situées dans des contextes vocaliques différents selon les phrases. Les phrases étudiées ci-dessous sont mises en correspondance avec leur équivalent en Beatbox :

\_ « Boots and cats » / P – T – K – E-H

\_ « Pâtes au pesto » / P – T – PS – E-H

\_ « Pose ta capuche, ta qu'à poser ta capuche » / PS – T – K – PS , T – K – PS – T – K – PS

Chaque phrase était répétée entre 5 et 10 fois en parole normale, puis il était demandé au sujet d'effectuer une transition la plus régulière possible vers le Beatbox (nous appellerons cette phase « continuum »), puis entre 5 et 10 fois la version Beatboxée.

La tâche de parole ainsi que la tâche en Beatbox ont été répétées 5 à 10 fois à des vitesses d'environ 80 syllabes par minute, puis 120 syllabes par minute, et enfin en accélérant de façon continue.

Chaque famille d'effet a été enregistrée en séries séparées de 5 à 10 répétitions de l'effet puis de quelques-unes de ses variantes au choix du sujet.

Enfin des enregistrements de production de Beatbox en séquences rythmiques libres utilisant ces effets, des mêmes séquences avec ajout de mélodies produites par voisement (appelé « humming » en Beatbox), d'instruments ou d'effets libres, et de séquences totalement libres ont été réalisés.

## ***5.2. Méthodes de mesure***

### **5.2.1. Enregistrements endoscopiques**

L'enregistrement endoscopique du corpus par nasofibroscopie souple a été effectué au CHU Albert Michalon par le Dr Ihab ATALLAH, ORL phoniatre dans le service d'ORL et Chirurgie cervico-faciale.

Pour ce faire, il a été utilisé un vidéo-endoscope de la marque Xion, branché sur une colonne d'enregistrement EndoSTROB de la même marque. Le signal acoustique était récupéré par le microphone du vidéo-endoscope. Après un bref examen médical de routine afin de vérifier l'absence de pathologie ORL, et une anesthésie locale à la Xylocaine 5% pulvérisée par voie nasale pour le confort du sujet, ce qui n'a pas provoqué de gêne pour la production au dire des sujets, l'enregistrement du corpus a été réalisé en vue « classique » c'est-à-dire en vue plongeante au-dessus du larynx. Des enregistrements des mêmes séquences ont été faits en tournant le nasofibroscope afin d'avoir une vue postérieure du larynx, nous appellerons cette vue « postérosupérieure ». Elle permet une vision plus précise du pied de l'épiglotte et de la commissure antérieure, mais présente le défaut de perdre les structures épilaryngées (base de langue, oropharynx et épiglotte).

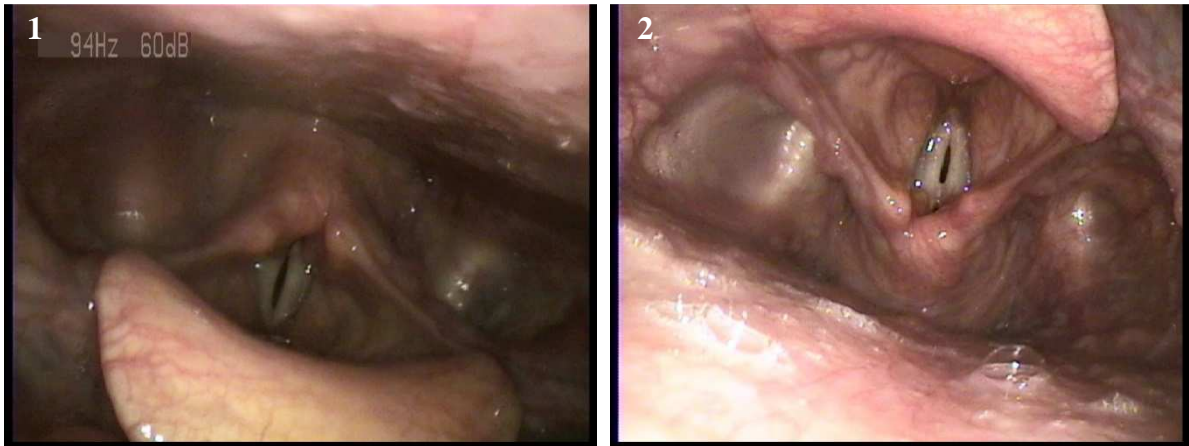


Figure 10 : Vues prises en fibroscopie souple : 1) vue classique supérieure, 2) vue postérosupérieure haute.

### ***5.3. Mesures aérodynamiques et EGG***

L'enregistrement du corpus sur le sujet 1 a eu lieu dans la chambre anéchoïque du GIPSA-Lab. Pour des raisons de fermeture administrative de bâtiment et d'indisponibilité de matériel, le sujet 2 n'a pas encore pu bénéficier de ces mesures.

Nous avons utilisé le système EVA2 avec l'environnement Phonedit pour un enregistrement simultané des :

- \_ Débit d'Air Oral (DAO),
- \_ Débit d'Air Nasal (DAN),
- \_ Pression Intra Orale (PIO),

La fréquence d'échantillonnage de ces mesures était de 6250Hz.

\_ Electroglottogramme (EGG) au moyen de l'ElectroGlottograph (model 7950A, Ltd, Londres, Royaume-Uni), branché sur la sortie EGG de l'EVA2, à la fréquence d'échantillonnage de 25000Hz.

\_ Signal acoustique par le micro de la pièce à main. Sa fréquence d'échantillonnage était de 25000Hz.

En raison d'un microphone défectueux sur la pièce à main, le signal acoustique a aussi été recueilli par un micro-cravate OMNI ATR3350 (Audio-Technica, Tokyo, Japon) échantillonné à 44100 Hz, disposé à 30cm environ du sujet, branché sur une carte son externe (AudioBox 22VSL , PreSonus, Baton rouge, USA).

Il n'a pas pu être installé de dispositif de synchronisation manuel (« trigger ») entre la carte son et le système EVA2, les séquences ont donc été enregistrées séparément puis

synchronisées à partir d'une fonction de corrélation croisée que nous détaillerons dans les méthodes d'analyse.

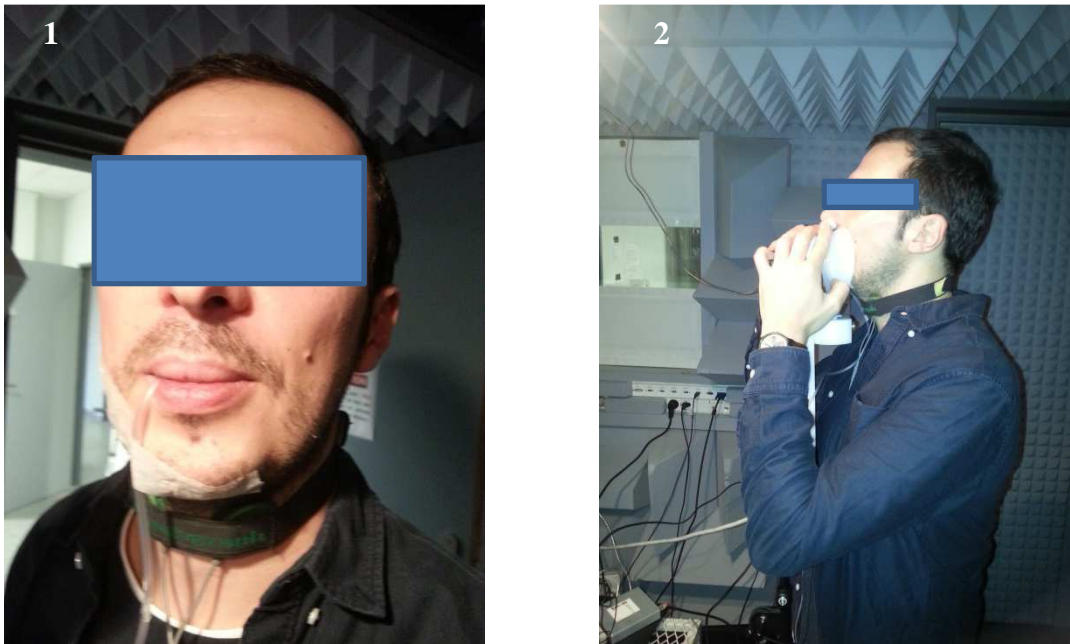


Figure 11 : 1) sujet avec le tube de mesure de Pression Intra Orale en position, il est glissé par le sujet au niveau du sillon amygdalogue. 2) Recueil de mesure avec le système EVA2.

#### ***5.4. Echographie laryngée***

La présence des électrodes nécessaires à l'EGG étant incompatible avec l'examen échographique laryngé, le corpus a été répété une troisième fois avec un enregistrement synchronisé du signal acoustique (fréquence échantillonnage : 44100Hz, micro-cravate), et des images échographiques recueillies au moyen du dispositif Telemed MicrUS EXT-1H (Telemed, UAB, Lituanie) et d'une sonde échographique du même fabricant.

Une reconnaissance des structures laryngées du sujet a d'abord été effectuée, puis la position de la sonde a été ajustée en fonction de la morphologie du sujet, restant relativement similaire à la position décrite au chapitre précédent.

Le logiciel d'acquisition UltraSpeech a permis la synchronisation des données acoustiques et échographiques. Nous avons cependant eu un problème de décalage initial des données de 790ms sur la première séquence et de 129ms pour la seconde.

Après discussion avec l'auteur du logiciel, ce décalage apparaît épisodiquement, et est dû à la présence d'une mémoire interne circulaire dans les systèmes d'enregistrement échographique initialement destinés à l'usage médical et permettant la visualisation des images précédant une acquisition. Ce décalage ne permet pas une bonne analyse

temporelle des données obtenues. Elles seront donc présentées de façon à illustrer nos propos avec une certaine réserve.

De même que pour les mesures aérodynamiques, le sujet 2 n'a pu être évalué.

## **6. Méthodes d'analyse**

### **6.1. Endoscopie**

Les images endoscopiques recueillies au CHU ont été exportées à partir du logiciel en format .mp4. Le signal audio a été extrait et analysé sous Praat afin de déterminer les séquences au long de l'examen. Nous avons tenté d'extraire les séquences d'images correspondantes avec le logiciel Elan après changement du script d'installation (cf notice installation Elan), cependant, le script est instable et n'a pas permis de créer ces séquences.

Nous avons donc effectué cette division des fichiers films initiaux à la main avec le logiciel VLC aux temps correspondant aux temps des tâches sur les TextGrid de Praat. Le signal audio de ces séquences étant plus léger en termes de taille et plus facilement exploitable, a bénéficié d'une annotation manuelle sous Praat au niveau du type de séquence (tier *séquence*), au niveau du type de production (tier *type* : parole, continuum, Beatbox), de la syllabe (tier *syllabe*), des consonnes (tier *consonne*), de l'effet (tier *effet*), et des variantes (tier *variante*).

Le début de la consonne ou de l'effet de Beatbox était l'apparition du burst, la fin était la reprise du voisement dans la parole normale ou la fin de l'effet au niveau perceptif.

Nous avons ensuite implémenté un script sous Matlab permettant d'extraire les images correspondant au temps précédent l'effet ou la consonne recherchée (pour pouvoir voir la coarticulation potentielle), jusqu'au début du temps de l'effet suivant pour le Human Beatbox et incluant la voyelle suivante pour la parole (en effet, les consonnes plosives étant plus courtes, nous ne pouvions pas éliminer la voyelle). Pour chaque effet/consonne retrouvé à partir du TextGrid Praat, nous avons donc des images comprises entre : fin (effet - 1) et début (effet+1). Le script extrait en même temps les données de la tier séquence et des effets/syllabes précédents et suivants sous les termes « séquence », « précédent » et « suivant » respectivement. Pour une facilité de lecture, et une analyse sous logiciel approprié, le script reconstruisait donc une séquence d'images avec le framerate correspondant et le signal acoustique en rapport avec les temps de début et fin de l'image sur le signal acoustique. Chaque petite séquence d'effet/consonne ainsi enregistrée

était séparée de la suivante par 5 images noires avec un signal acoustique nul pour une meilleure différenciation des effets. Etant donné que la durée du signal extrait n'est pas exactement le même que celui des séquences d'images (différence d'échantillonnage), il a été décidé de modifier le signal audio correspondant à chacune des séquences pour éviter un décalage, en ajoutant des zéros ou en soustrayant le délai correspondant entre le nombre d'image de la « miniséquence » et le signal audio extrait.

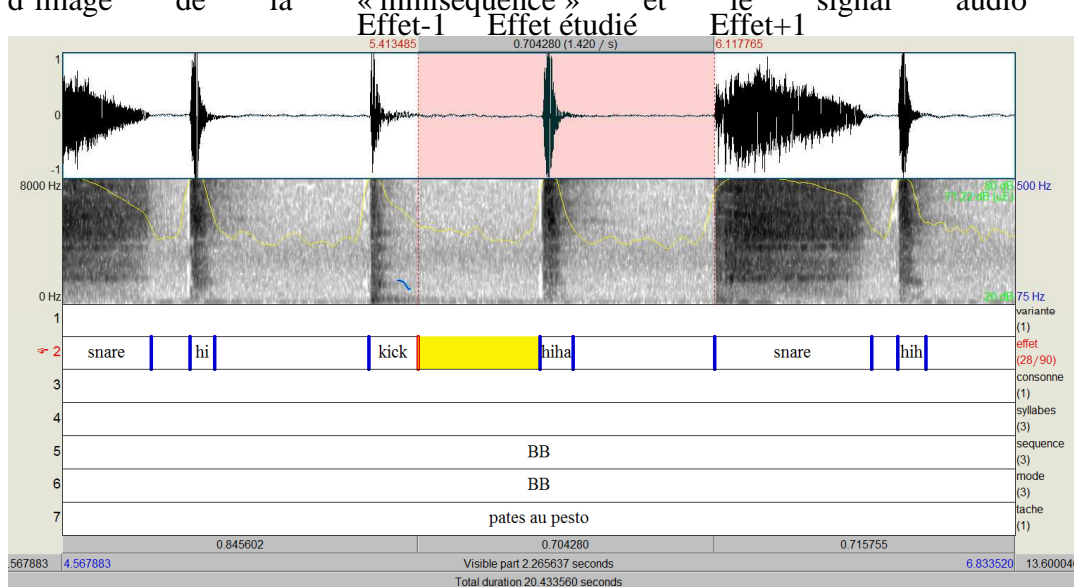


Figure 12 : exemple d'extraction de la durée totale d'un effet, ici hihat dans une production de Beatbox dans la tâche « pates au pesto », s'étendant de la fin du kick au début du snare.

Le temps d'ouverture de l'objectif n'est pas fixé en raison de l'adaptation du logiciel Xion à la luminosité et peut prendre toute valeur comprise entre un temps minimum d'ouverture de 5ms et le temps correspondant au framerate. Il n'est pas encodé sur les fichiers exportés. C'est pourquoi nous avons considéré que le temps de shutter (ouverture de l'objectif) est égal à 5ms.

Les séquences jugées inexploitables (luminosité trop faible, salive sur la caméra etc...) ont été éjectées de la base de données.

L'enregistrement du premier sujet a été fait avec le mode stroboscopie de la caméra (ce qui était passé inaperçu lors de notre expérimentation). Nous avons retrouvé un framerate variant selon les séquences sur les enregistrements du sujet 1 entre 47 et 52 images par seconde après découpage. En fait la temporalité de chaque image n'est pas réellement encodée dans le fichier vidéo mais il y aurait un « time-stamp » épisodique qui permettrait une pseudo-synchronisation avec le fichier audio à des temps particuliers, le framerate étant donc moyenné entre ces « time-stamps ». Le framerate change donc lorsque



le logiciel commence à détecter une fréquence fondamentale et augmente ou diminue selon celle-ci. Cependant le logiciel étant à destination médicale, il n'encode pas la temporalité de façon précise. Ainsi des fonctions matlab permettant de retrouver une image à partir de la temporalité d'un fichier vidéo n'ont pas pu être utilisées, nous avons donc dû décomposer chacun des fichiers « séquence » afin d'extraire les images correspondantes aux temps du TextGrid pour les réencoder par la suite.

L'enregistrement du deuxième sujet s'est fait en mode « classique » à 25 images par secondes, à taux fixe.

Le temps correspondant de chaque image par rapport au burst (début de la consonne ou de l'effet) a aussi été encodé dans un tableur (négatif si l'image est avant le burst, positif si l'image est prise après le burst). Les images avec un délai positif qui sont comprises dans le temps de la consonne/effet sont annotées « in », les images hors de la consonne/effet sont annotées « out ».

Les images d'un type d'effet ou de consonne, ainsi extraites avec leurs temps, labels, effets/voyelles précédents et suivants, type de variante (si il y a une variante), nom de séquence, type de production et nom du fichier initial d'extraction, ainsi que le signal audio correspondant ont été encodées dans des structures Matlab avec le type d'effet comme nom. Ces structures ont de plus été exportées sous format vidéo, audio et tableau Excel pour analyse sous logiciels dédiés.

Les limites du logiciel Elan liées à ces problèmes de temporalité et d'entrelacement d'image dues à l'exportation initiales des données ont donc été contournées.

Nous avons ainsi pu annoter un certain nombre de répétition d'effets/consonnes avec le logiciel Elan par une analyse perceptive visuelle des gestes des structures laryngées en nous inspirant de l'échelle OCM (de Torcy et al., 2014) permettant d'évaluer le type de mouvement de la structure, et l'état d'ouverture du « sphincter ».

Les structures étudiées de cette façon sont les plis vocaux, les cartilages aryénoïdes, les plis vestibulaires, l'épiglotte associée à la base de la langue, les parois pharyngées latérales et postérieures notées sous le dénominateur « pharynx haut », l'hypopharynx divisé en sinus piriforme (dénominateur « hypopharynx ») et espace rétrocricoïdien, et l'entonnoir laryngé dans sa globalité (« larynx »).

Les mouvements des plis vocaux ont été annotés en terme de mouvement d'ouverture (OM) et de fermeture (FM). Les degrés d'ouverture (O1, 2 et 3) et de

fermeture (FC pour fermeture complète, FI fermeture incomplète, considérée comme plus proche de la fermeture de l'étage glottique que le degré d'ouverture O1) sont associés avec le trait voisé (V) ou non voisé (NV). Les mouvements globaux laryngés d'ascension (A) étant difficilement évaluable de façon certaine en raison des mouvements des autres structures et de la caméra ont souvent été notés avec « ? » en cas d'interrogation sur la réalité du geste. Les mouvements des cartilages aryténoïdes sont nommés (A) lorsque l'on a une bascule antérieure, pour les bandes ventriculaires nous avons relevé des états d'hypercontraction (C) sans fermeture du sphincter, et d'états d'ouverture importante avec visualisation du ventricule avec une forme arquée des bandes (noté A). L'épiglotte présente des mouvements de bascule avant ou arrière (A ou R), l'hypopharynx (sinus piriformes) et le pharynx haut ont des mouvements de contraction (C) ou de dilatation (D). Pour l'hypopharynx nous avons annoté, pour un certain nombre d'effets, « D » la phase de dilatation de son début jusqu'au retour à son état normal car il est difficile d'évaluer à l'œil l'apogée de la phase sinon en séquence « D » et « C » comme pour le pharynx haut. La durée de visualisation de l'espace rétrocricoïde a été annotée « D ».

Les données temporelles ont été exportées en référence à l'instant du burst correspondant. Les gestes qui ont débuté 500ms avant le burst acoustique ont été exclus de l'analyse car rattachés aux effets précédents.

En dehors de la temporalité gestuelle, nous n'avons pas encore pu faire d'analyses quantitatives et automatisées des mouvements de chaque structure.

Nous avons effectué des analyses des moments spectraux principaux des signaux acoustiques grâce à un script Matlab et avons donc récupéré pour chaque effet ou consonne (sans la voyelle) :

\_ l'intensité moyenne et maximale (Imean et Imax en dB) calculées de façon relative à partir de fenêtres temporelles glissantes de 2ms, ainsi que le temps associé à l'intensité maximale.

\_ le centre de gravité spectral (CDG en Hz).

\_ le coefficient d'asymétrie ou Skewness deviation.

\_ le coefficient de kurtosis.

## **6.2. Mesures aérodynamiques**

Le fichier audio du système EVA2 était trop bruité pour l'analyse en raison d'un problème matériel. Il a d'abord été filtré puis sur-échantillonné à 44100 Hz afin d'appliquer un algorithme réalisé sur Matlab, de corrélation croisée avec le signal audio enregistré en parallèle avec le micro-cravate. Cela nous a permis de synchroniser les deux signaux et d'utiliser donc le signal du micro-cravate pour nos analyses aérodynamiques.

Ce signal audio a été annoté manuellement sous Praat de la même manière que pour les vidéos : c'est-à-dire avec annotation temporelle des consonnes/effets, types de production, séquences, variantes si il y a lieu.

Un script d'extraction des effets/consonnes à partir du TextGrid a permis de mettre en relation temporelle leurs données acoustiques, aérodynamiques (PIO, DAO, DAN) et EGG ainsi que leur visualisation.

Le signal acoustique a été traité de la même manière que pour l'analyse des données endoscopiques (cf ci-dessus).

On peut distinguer pour les consonnes plosives plusieurs phases sur le débit d'air oral et la PIO : une phase de fermeture (« closing ») pendant l'intervalle  $T_c$ , une phase fermée (« occlusion »)  $T_o$ , une phase de relâchement (« release »)  $T_r$ . Nous avons calculé la durée  $dT$  correspondant à la différence entre le début du relâchement et le maximum de PIO (Muller & Brown, 1979). La différence de pression correspondant à  $dT$  est appelée  $dP$ . On peut ainsi distinguer les pressions correspondantes à ces temps ( $P_o$ ,  $P_c$ ,  $P_r$ ) et la pression moyenne d'occlusion ( $P_a$ ) calculée comme la moyenne de l'aire sous la courbe de PIO pendant la phase d'occlusion. Les pentes  $a$  (équivalent de la dérivée moyenne entre le temps de début de la phase fermée et le temps de la pression moyenne d'occlusion), et  $b$  (dérivée moyenne entre le temps de la pression moyenne d'occlusion et le début du relâchement) permettent de caractériser la forme de la courbe et les vitesses correspondantes d'évolution de PIO ainsi que le coefficient  $d$  défini comme ( $d = a-b$ ) qui est positif ou négatif selon la forme de croissance de la PIO en phase fermée. Nous avons choisi de retrouver ces paramètres car nous pensons pouvoir les mettre en relation avec nos données endoscopiques.

Sur chaque signal de pression et de débit nous avons appliqué un algorithme permettant de retrouver automatiquement ces temps. Les phases  $T_o$   $T_c$  et  $T_r$  étaient détectées à partir des maxima du signal de pression ou de débit et des points d'inflexion

des courbes. Les données ont ensuite été vérifiées manuellement et corrigées le cas échéant. Si la forme du débit d'air oral ne permettait pas la détermination précise des instants de fermeture, nous utilisons en approximation la durée ( $T_c+T_o$ ) pour le calcul de la Pa. (cf figure 13)

Nous avons aussi créé une fonction callback sous Matlab afin de déterminer par morcellement le point d'inflexion de la courbe de PIO ( $T_{pw}$ ) en phase fermée, et une fonction permettant de retrouver le point de changement de courbure ( $T_{cp}$ ) sur la phase d'augmentation du signal de PIO (closing + occlusion). En effet la visualisation initiale des données retrouvait des profils différents selon les effets et nous voulions savoir si les temps associés à ces paramètres étaient statistiquement corrélés à certaines conformations laryngées (ces données ne seront pas analysées dans ce travail, sauf mention contraire).

Les temps correspondant aux paramètres décrits ont comme référence le délai avec le burst de leur effet correspondant.

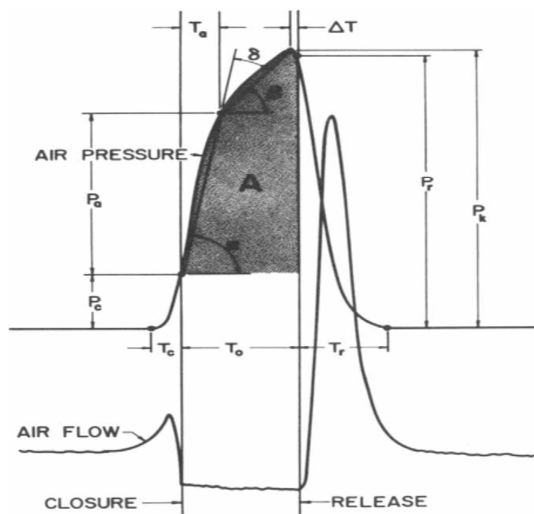


Figure 13 : Mise en relation de la PIO avec le DAO (tirée de Muller et Brown, 1979), la courbe de PIO est située en haut, la courbe de débit d'air oral en bas de la figure.

$T_c$  : phase de fermeture,  $T_o$  : temps d'occlusion,  $T_r$  : temps de relâchement

$P_c$  : PIO au début du temps d'occlusion,  $P_a$  : pression moyenne d'occlusion, a b pentes de pression avant et après  $P_a$ ,  $P_r$  : pression lors du relâchement,  $dT$  : différence entre les temps de maximum de la PIO et le début du relâchement.

Pour l'EKG nous avons retrouvé des patterns répétés et relativement robustes mais nous n'en donnerons qu'une analyse qualitative.

Nous résumons ici les variables sur lesquelles nous nous concentrerons :

\_ variables de production : type de consonne/effet , séquence / type de production, effet/voyelle précédente, effet/voyelle suivante

\_ données temporelles : Tc, To, Tr, Tom (temps correspondant à pression moyenne d'occlusion), dT, TmaxI (temps correspondant au maximum d'intensité acoustique), durée de la consonne/effet

\_ formes de pente : a, b, d, a\_oaf (moyenne de la dérivée de la phase croissante du DAO)

\_ données aérodynamiques : Pc, Pa, Pr, dP (différence de pression entre les temps Tr et temps du maximum de la PIO).

Etant donné la position très arrière du /k/ et du Rimshot, l'immense majorité des données de PIO pour ce couple n'est pas analysable, nous les avons donc exclus de notre analyse. En effet, la sonde de PIO était bloquée par le sujet au niveau du sillon amygdalogue, et était donc obstruée par la langue dans les consonnes vélaires.

### ***6.3. Mesures échographiques***

Un script Matlab permet la reconstruction en film de la séquence d'images enregistrée synchronisée par rapport au son. Le décalage initial relativement important ne permet pas une synchronisation réelle des données comme nous l'avons mentionné plus haut, nous présenterons donc notre analyse qualitative de ces séquences. Nous avons modifié le script UltraMat de T. Hueber dans le but de visualiser l'image échographique synchronisée avec le signal audio correspondant.

## ***7. Approche statistique***

Les données quantitatives EVA2, acoustiques et endoscopiques correspondant aux variables que nous avons présentées ci-dessus ont été analysées sous R, avec l'aide de l'ingénieur statisticien du laboratoire (Sylvain Gerber).

Pour les données temporelles endoscopiques, afin de déterminer l'intervalle de confiance des délais moyens et étant donné le faible nombre d'effets annotés, nous avons créé une fonction Bootstrap à l'aide du package Boot, pour rééchantillonner les données sur 5000 échantillons afin d'en augmenter la significativité.

Il a été mis en place un modèle de type gls (méthode des carrés partiels généralisée) à partir du package lme permettant la comparaison des effets entre eux pour

chacune des variables et selon leur coarticulation, ce modèle sera utilisé pour les mesures aérodynamiques.

La comparaison des données temporelles endoscopiques et EVA a été réalisée avec des tests non paramétriques.

Le risque de première erreur a été choisi à 0.05.

## Chapitre 5. Résultats

### 1. *Caractérisation endoscopique des mouvements laryngés concomitants à la production des trois effets.*

Nous allons décrire à la suite les gestes laryngés relevés pour chacune des consonnes plosives et des effets du Beatbox correspondants, ainsi que leurs principales variantes. Les mouvements seront décrits qualitativement avec les temps approximatifs sur un effet « type » pour faciliter la lecture, puis nous montrerons la séquence temporelle type avec le signal acoustique correspondant. Les données temporelles avec les intervalles de confiance calculées au moyen du Bootstrap évoqué dans le paragraphe méthode sont exposées dans les tableaux en annexe 2 pour les mouvements précédant le burst. Néanmoins le fait qu'une structure ne peut pas être dans deux états donnés (mouvement, état stable, ouverture et fermeture...) n'a pas été appliqué au modèle expliquant le chevauchement important des intervalles de confiance.

Pour notre sujet 2, la durée de prise d'image étant très importante (40ms) fait que nous avons à peine une ou 2 images contemporaines d'une consonne et rend une analyse temporelle difficile. Nous présenterons donc des images de facons qualitative. Les délais dans les figures sont en millisecondes.

La classification perceptive de certains effets du Beatbox a été faite par l'auteur, avec confirmation d'une collègue qui travaille sur le Beatbox depuis plusieurs années. Les noms d'effets variant souvent entre les beatboxeurs et les styles, nous avons tenté de donner les noms de variante les plus proches acoustiquement.

#### 1.1. *kick - /p/-/b/*

##### 1.1.1. **Sujet 1 : beatboxeur professionnel**

**En parole**, la consonne /b/ est retrouvée sur 12 occurrences dans notre base de données. Etant donnée la constitution du corpus, nous la retrouvons uniquement dans le contexte : /ts/ ou inspiration - /b/ - /u/, correspondant à « boots and cats ».

En termes de gestuelle laryngée, les plis vocaux ont un mouvement d'adduction (suivant la production d'une consonne sourde /ts/ ou d'une ouverture inspiratoire) débutant environ 300ms avant le burst et se finissant vers -180ms, avec apparition d'un voisement après établissement du contact (environ 150ms avant le burst, correspondant au voisement du signal audio), puis une abduction 220ms après le burst correspondant à la fin de la

voyelle. On note un léger déplacement en avant de la base de langue entrainant l'épiglotte de -20ms à +140ms induisant un étirement des replis aryépiglottiques probablement en raison de la voyelle /u/. L'hypopharynx se dilate légèrement entre -72ms et +166ms.

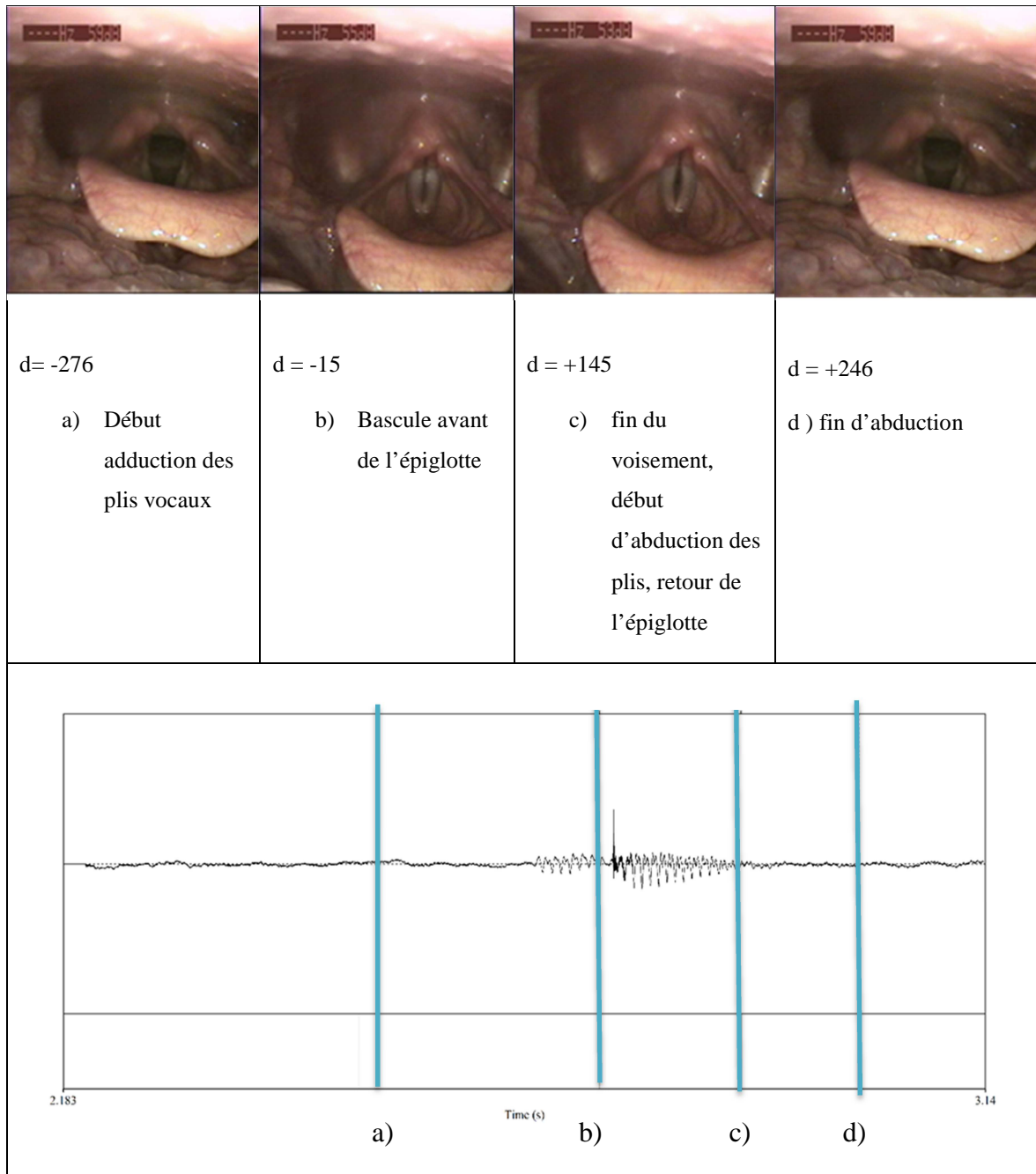
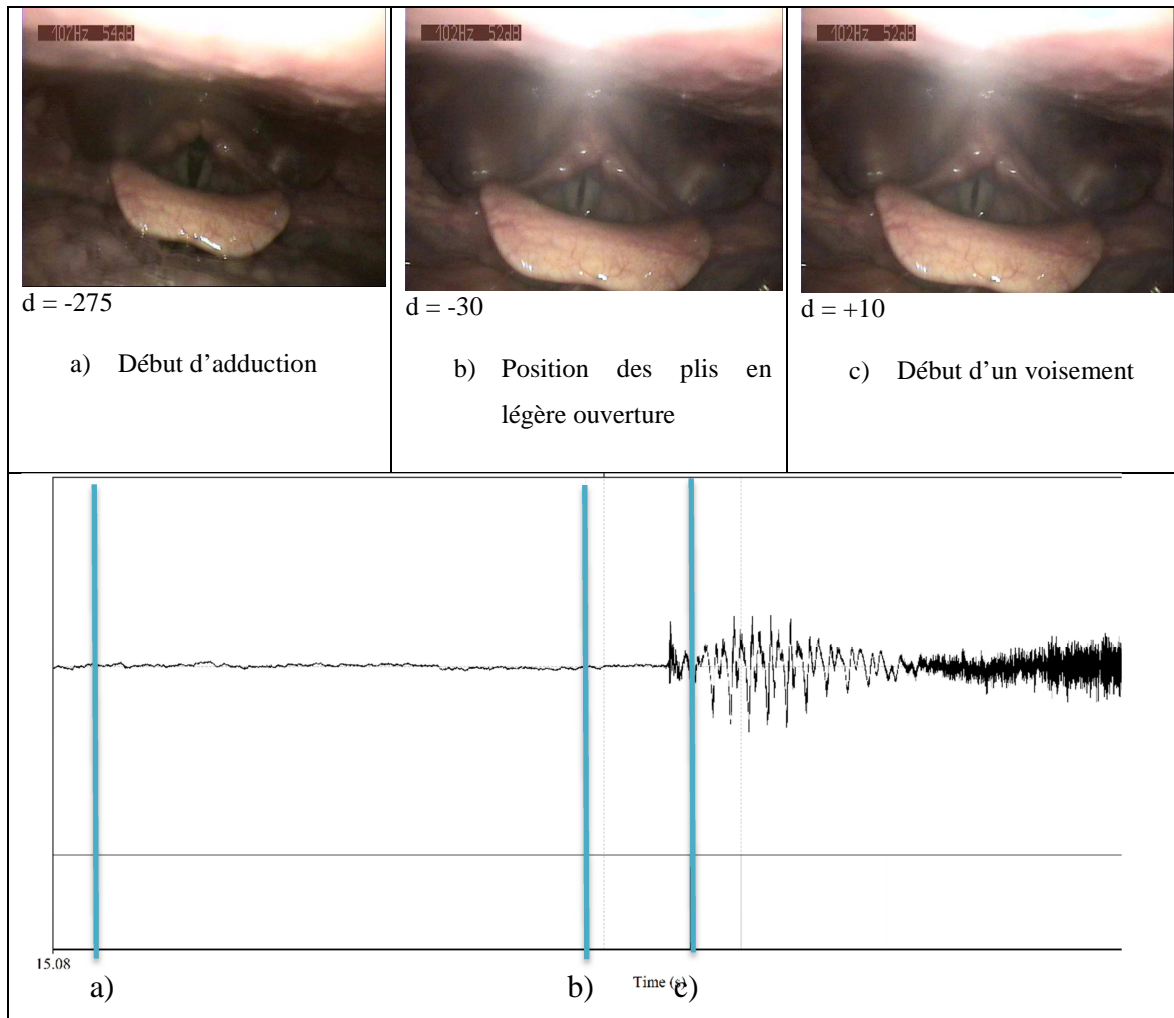


Figure 14 : Illustration d'une séquence laryngée en parole pour l'enchaînement de /bu/ (phrase « boots and cats »). d (ms) se réfère au burst.

Pour la consonne /p/ (60 occurrences), le mouvement d'adduction des plis vocaux est plus tardif (débute vers -220ms). Il ne semble pas y avoir de contact des plis vocaux pendant la production consonantique mais la mise en vibration des plis se fait assez rapidement à la suite vers 14ms environ, ce qui semble en rapport avec l'apparition d'une



structure périodique sur le signal audio, VOT en moyenne 17.5ms. Le test de wilcoxon sur les moyennes de VOT et le temps de vision d'un voisement n'est pas significatif :  $pvalue=0.45$ . La base de langue entraîne l'épiglotte pour les /p/ suivis d'un /u/ ou d'un /y/ et dépend donc du contexte vocalique. L'hypopharynx se dilate légèrement à partir de -180ms.



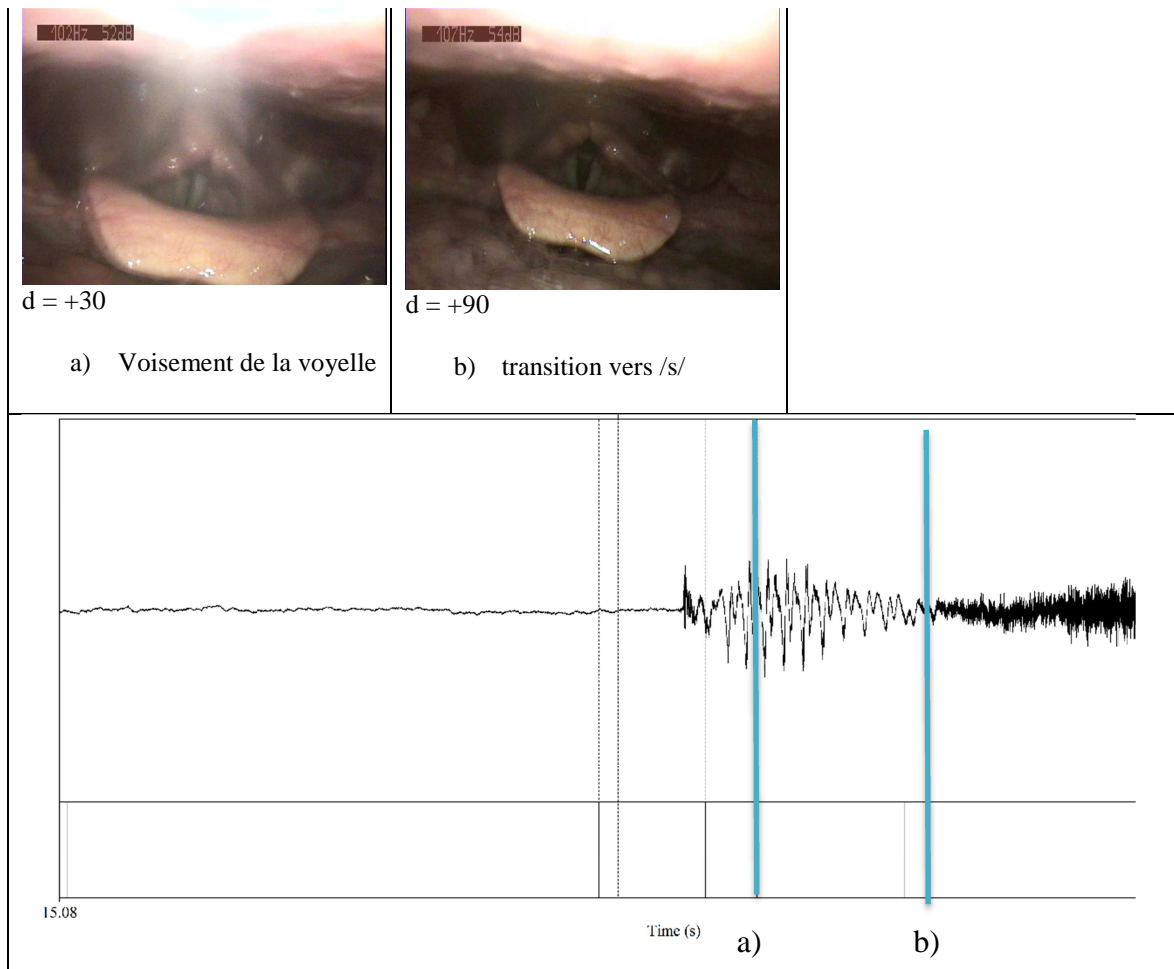


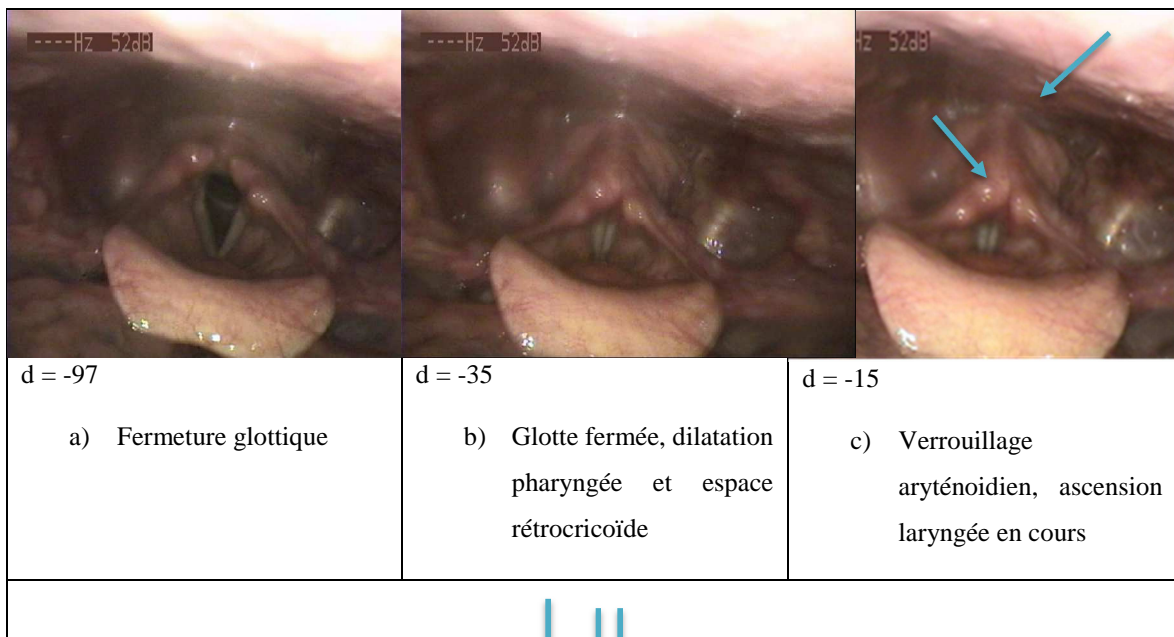
Figure 15 : Illustration d'une séquence laryngée pour l'enchaînement de /pes/ (phrase pâtes au pesto »). d (ms) se réfère au burst

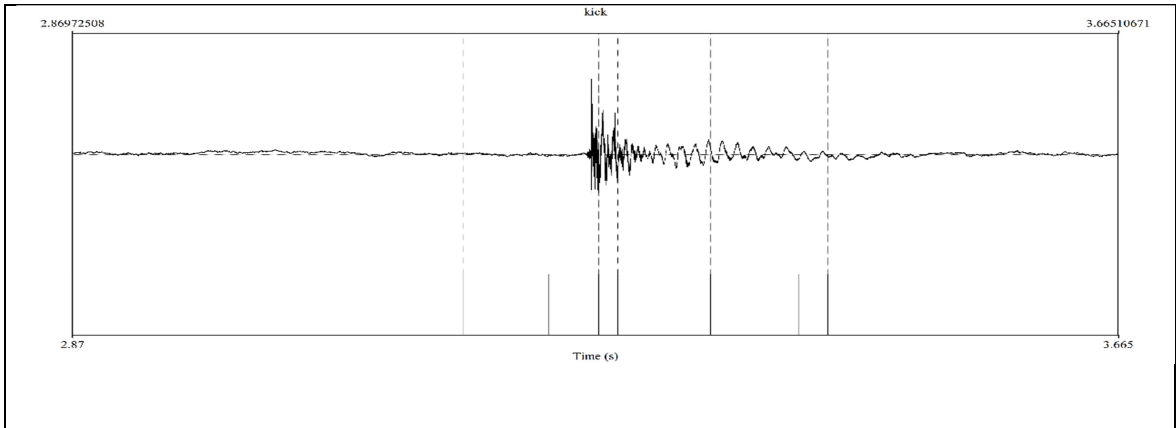
**En Beatbox**, l'effet kick classique non voisé a un pattern gestuel laryngé complètement différent. La fermeture complète laryngée est plus rapide, débutant à -180ms et se terminant vers -70ms. La durée de fermeture n'est pas influencée par la tâche de production (continuum, effet kick simple, rapidité de l'enchaînement wilcoxon test :  $p=0.52$ , non significatif). La fermeture glottique est totale, avec hypercontraction légère des plis vestibulaires, et avec occlusion complète entre les deux apophyses vocales. Puis selon les tâches de production il y a un mouvement d'ouverture vers un état glottique de O2 (position intermédiaire assez fermée) à O3 (position ouverte), avec parfois une mise en vibration des plis vocaux (sans contact) très rapidement après le burst (+10ms).

Pour l'effet kick voisé équivalent à la version anglophone du kick classique (évolution du « /bu/ de « boots and cats »), la glotte est dans une position plus ouverte par rapport à la production d'un /b/ en parole, le relâchement par rapport au kick classique décrit ci-dessus se fait vers une position O1/O2, avec des plis vocaux épais et flasques entraînant une vibration sans contact des plis vocaux. Nous pouvons suspecter que cette

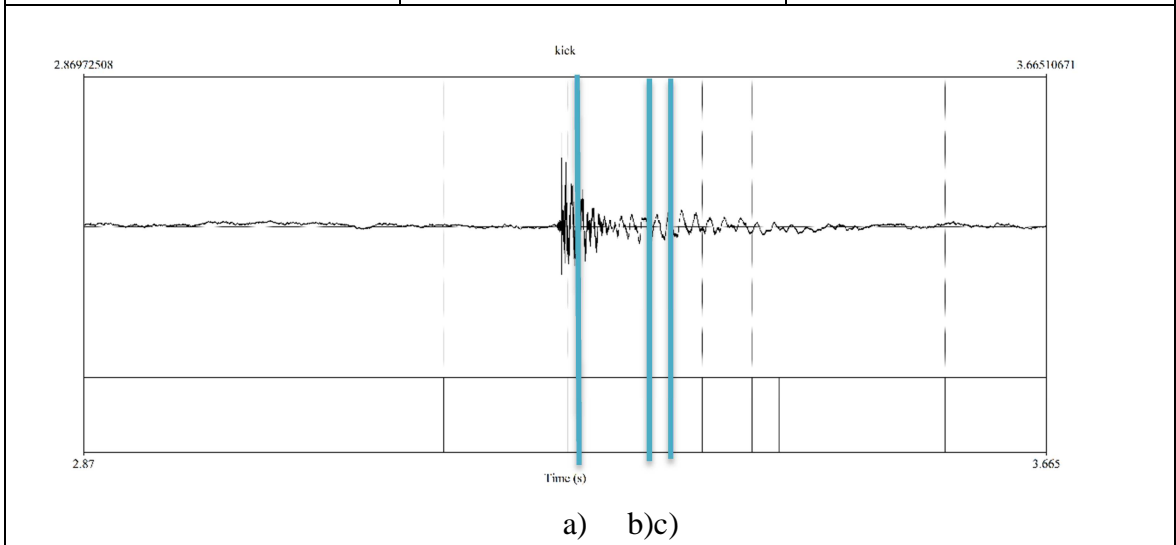
vibration se fait grâce au relâchement du muscle thyro-aryténoïdien. La fréquence fondamentale mesurée sur cette phase de voisement dans ce cas-là est plus basse que pour la parole (moyenne  $f0_{kickV} = 84$  Hz vs  $f0_{ou/parole} = 112$  Hz, test wilcoxon : p-value  $<0,01$ ).

On note un mouvement d'ouverture ou de dilatation de l'espace rétrocricoïdien à la suite de la fermeture glottique (vers -60ms) (figure 16b), qui paraît lié à un mouvement d'ascension laryngée et de bascule en avant du larynx. Le maximum de dilatation précède de peu le burst (-5ms environ) et se finit au moment du déclenchement du burst en un mouvement très rapide. Il faut aussi noter la contraction pharyngée en « tube » qui débute environ 10ms avant le burst et se termine 10ms après. Ces mouvements ne sont jamais observés dans la production de /b/ ou /p/ en parole et sont systématiques sur les kicks en Beatbox. La dilatation de l'hypopharynx est de visu très importante aux environ de -76ms avant le burst, précédée d'une hypercontraction du sphincter glottique vers -85ms (les plis vocaux sont en position fermée avec un « verrouillage des aryténoïdes +/- une hypercontraction des plis vestibulaires).





<p>d = +25</p> <p>a)</p>	<p>d = +107</p> <p>b) Ouverture O2 avec voisement et ascension laryngée</p>	<p>d = +127</p> <p>c) Voisement sans contact</p>
--------------------------	---	--



a) b)c)

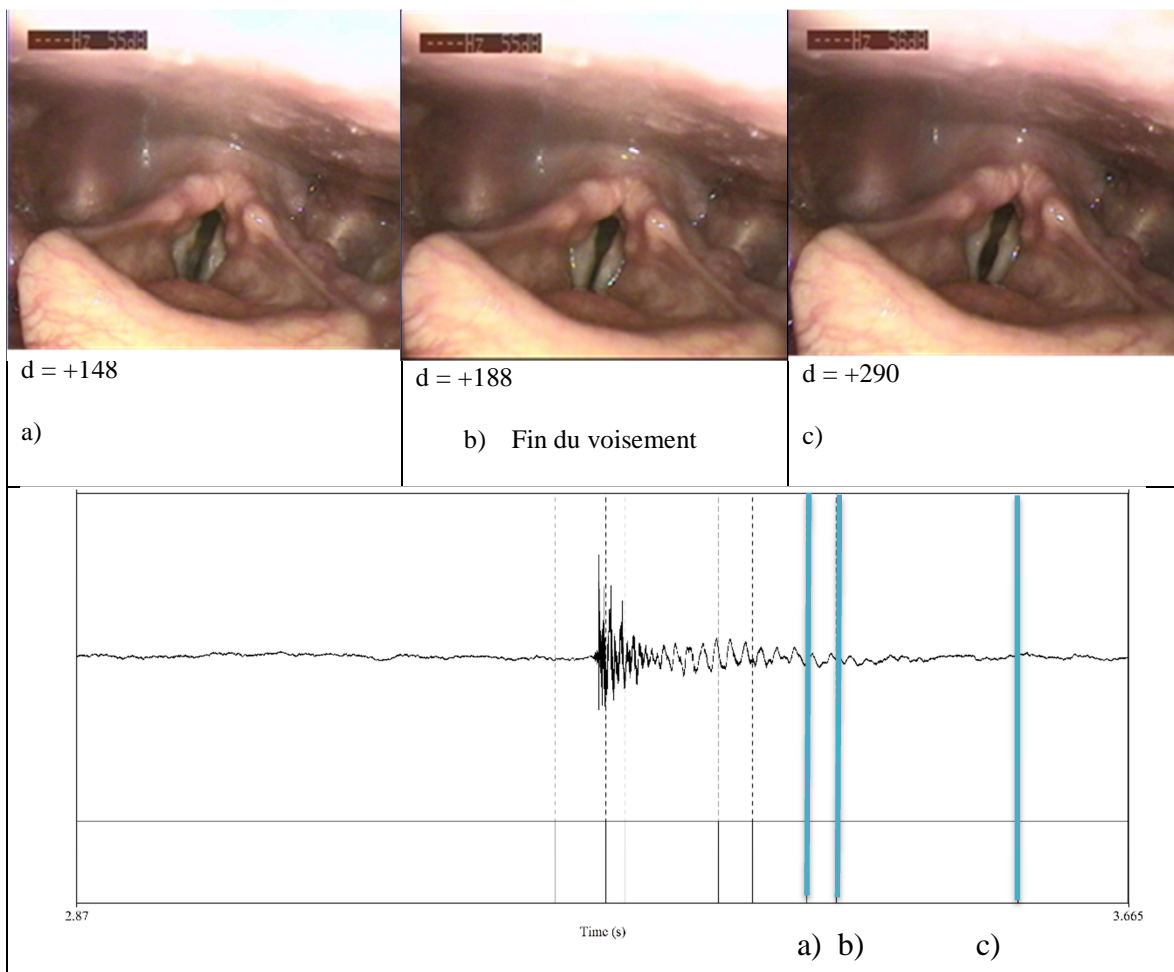


Figure 16 : séquence temporelle d'un kick voisé, les flèches montrent l'espace rétrocricoïdien ouvert, et l'hypercontraction des plis vestibulaires avec le verrouillage aryténoïdien.

Les mouvements sont très reproductibles (>80% des effets étudiés,  $p$ value<0.001) pour l'ensemble des kicks, mise à part l'observation de vibration des plis vocaux dans le kick classique.

Il nous semble que le degré de dilatation hypopharyngée, et de l'espace rétrocricoïdien ainsi que l'amplitude de l'ascension laryngée soient reliées à l'intensité de l'effet, ce qui reste à confirmer sur des mesures quantitatives.

Sur les séquences en continuum nous retrouvons des catégories gestuelles en parfaite adéquation avec les catégories perceptives acoustiques : c'est-à-dire que nous observons une discontinuité entre la consonne et le son percussif correspondant tant d'un point de vue acoustique que de la gestuelle laryngée, malgré une intensité acoustique qui n'est pas statistiquement différente de la production de /b/ et /p/ en voix forte (Imoyennekick = 68dB, Imoyenne/b//p/voix forte = 64dB, wilcoxon test : pvalue = 0.17).

Une analyse rapide sans annotation des séquences d'« accélération » et des plosives libres retrouve un kick légèrement moins intense, non influencé par les effets précédents ou suivant, avec le pattern typique d'un kick.

### **1.1.2. Sujet 2 : beatboxeur amateur**

La production de /b/ est quasi identique au sujet 1 en dehors d'un enchaînement plus fréquent de [en-b-u] pour « boots-and cats-and », ce qui ne modifie pas nos constatations sur une fermeture quasi complète glottique avec vibration précédant le burst (12 occurrences au total).

De même les /p/ ne montrent pas de différence avec ceux du sujet 1 c'est-à-dire une fermeture glottique incomplète (11 occurrences réelles). En revanche notre sujet a du mal à se détacher du Beatbox et a tendance à faire des kicks en lieu et place de /p/ bien qu'il s'en rende compte lui-même.

Pour l'effet kick, on observe également une certaine similarité avec le sujet 1 mais le sujet 2 réalise une constriction complète épilaryngée avec une légère composante de piston laryngé. La présence de voisement n'est pas visible sur les images. Les kicks produits ne sont ni acoustiquement ni perceptivement identiques au sujet 2. On pourrait les catégoriser comme « Punchy kick » (de timbre plus grave et plus sombre). Il n'y a pas une dilatation hypopharyngée aussi importante que pour le sujet 1. L'espace rétrocricoïde n'est pas visible. Cette hypercontraction et mobilisation du larynx en bloc se retrouve sur les autres effets produits dans le Beatbox, pouvant nous interroger sur un éventuel apprentissage de la dissociation pharyngolaryngée, pour un moindre effort vocal, encore non acquis.

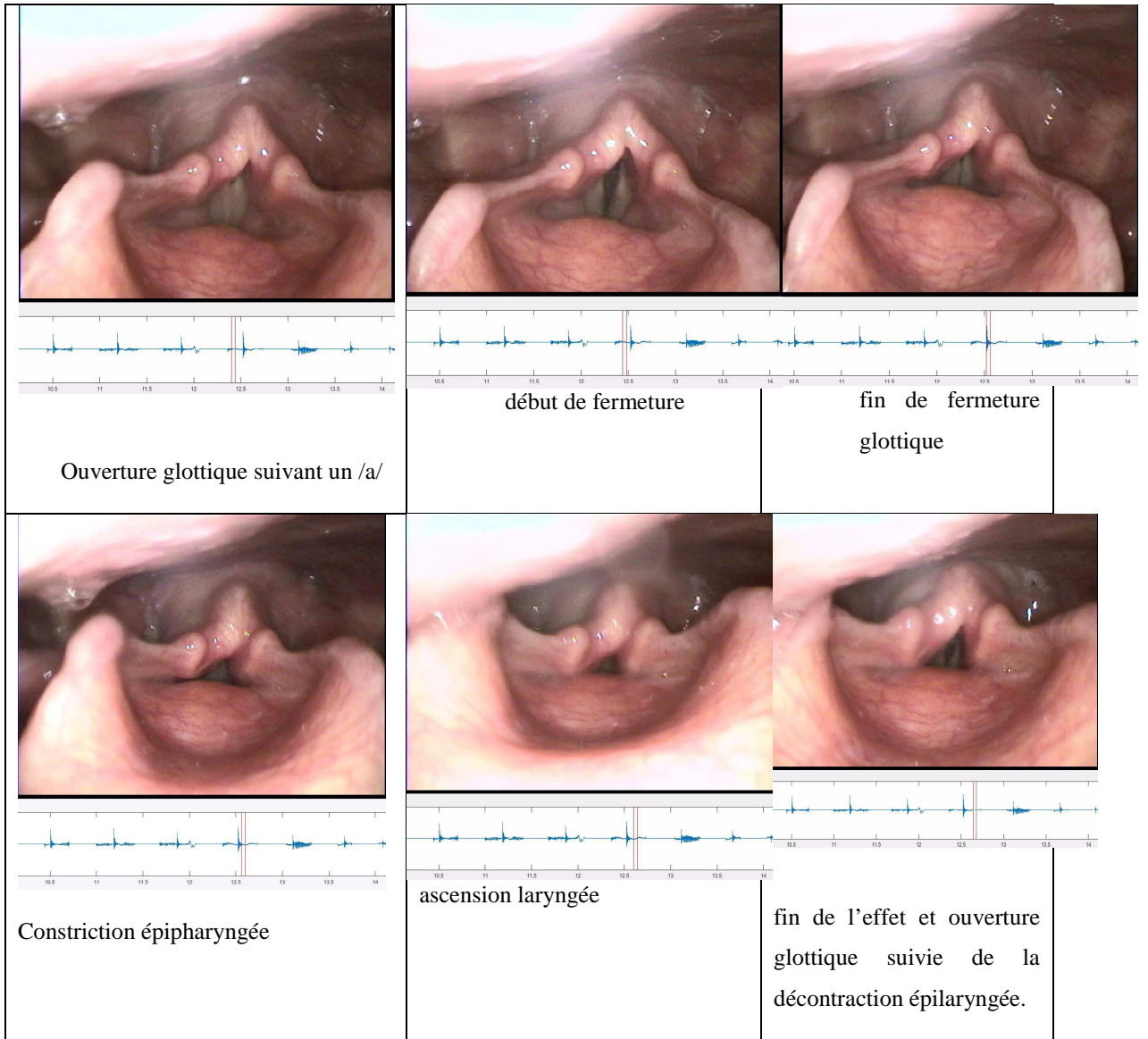


Figure 17 : illustration de la production d'un kick par le sujet 2.

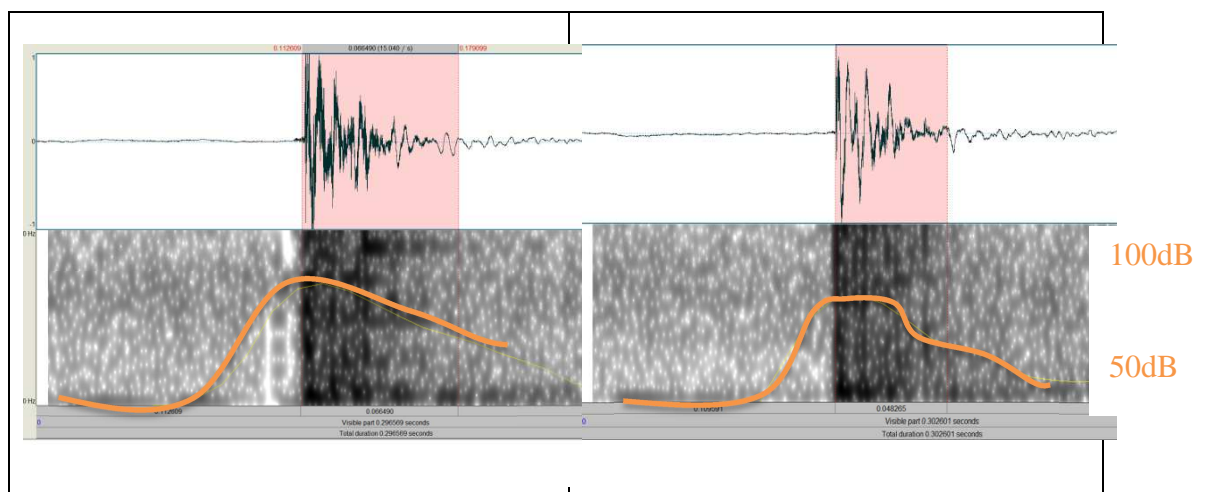


Figure 18 : comparaison des signaux acoustiques du kick du sujet 1 et 2

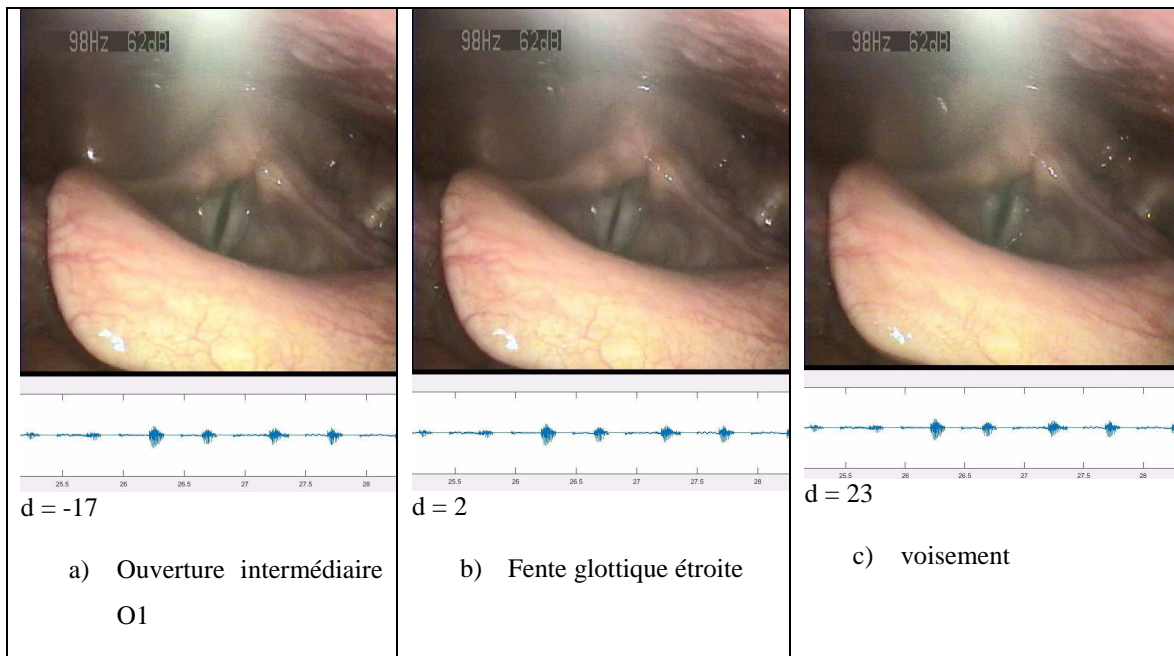
## 1.2. *hihat* - /t/

### 1.2.1. Sujet 1 : beatboxeur professionnel

**En parole**, nous avons 80 occurrences de /t/ dans différents contextes (environ 20 par contexte) correspondants aux tâches décrites dans notre protocole.

Au niveau laryngé comme pour les autres consonnes, nous avons un mouvement initial de rapprochement des plis vocaux (-90 ms), vers un état de fente glottique étroite mais sans contact. Après le burst un très léger rapprochement aryténoïdien permet la mise en vibration des plis si la consonne est suivie d'une voyelle, ou d'une abduction des plis en cas de fricative suivant la plosive (+10 ms).

Un léger mouvement de base de la langue parfois accompagné par l'épiglotte se retrouve pendant la production de la voyelle/consonne suivante mais semble plutôt relié à la coarticulation, c'est-à-dire que ce mouvement se retrouve particulièrement dans /pou/ à l'inverse du /pa/. Pendant les séquences de transition de la parole au Beatbox on peut observer une stratégie d'évolution de la production du /t/ avec contact des plis, apparition d'une dilatation hypopharyngée, et mouvements d'ascension du larynx.





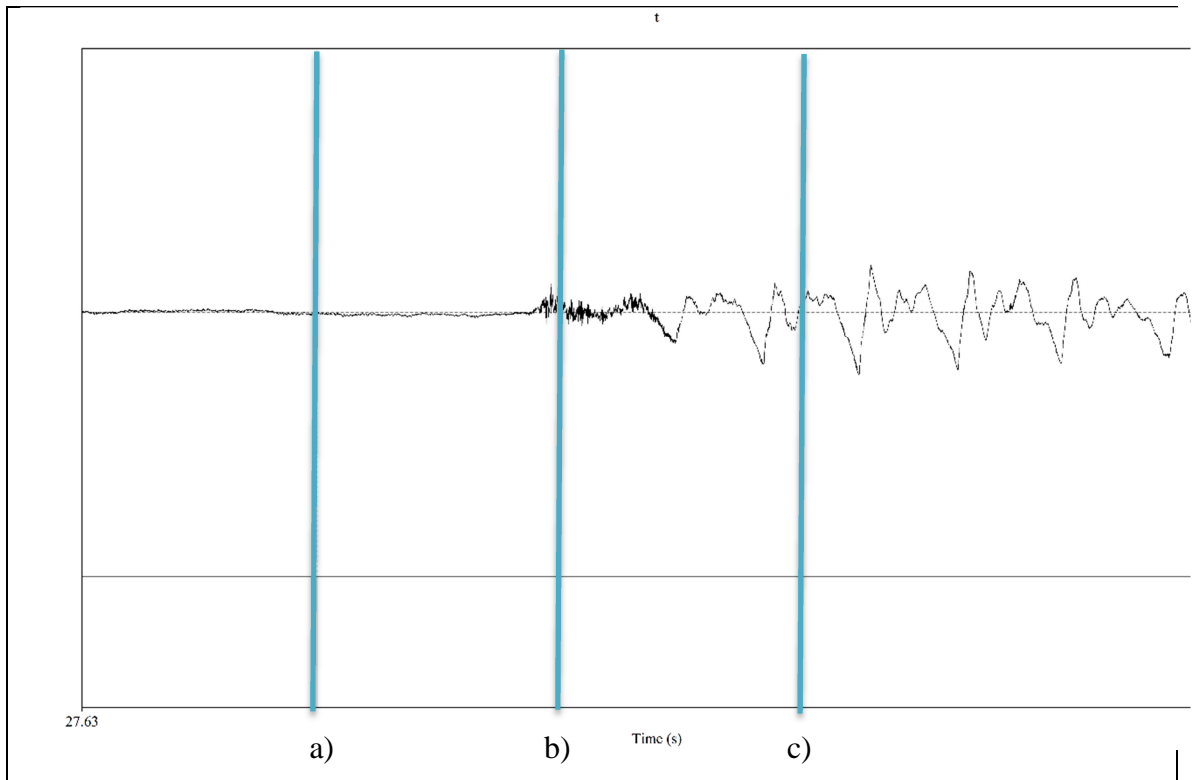
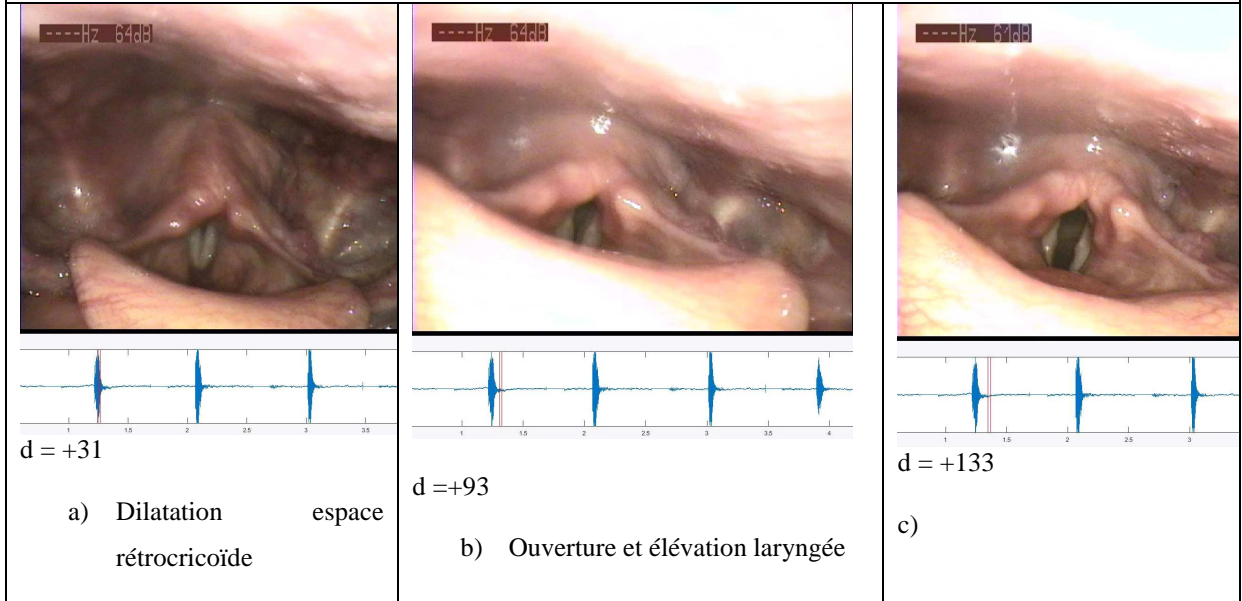
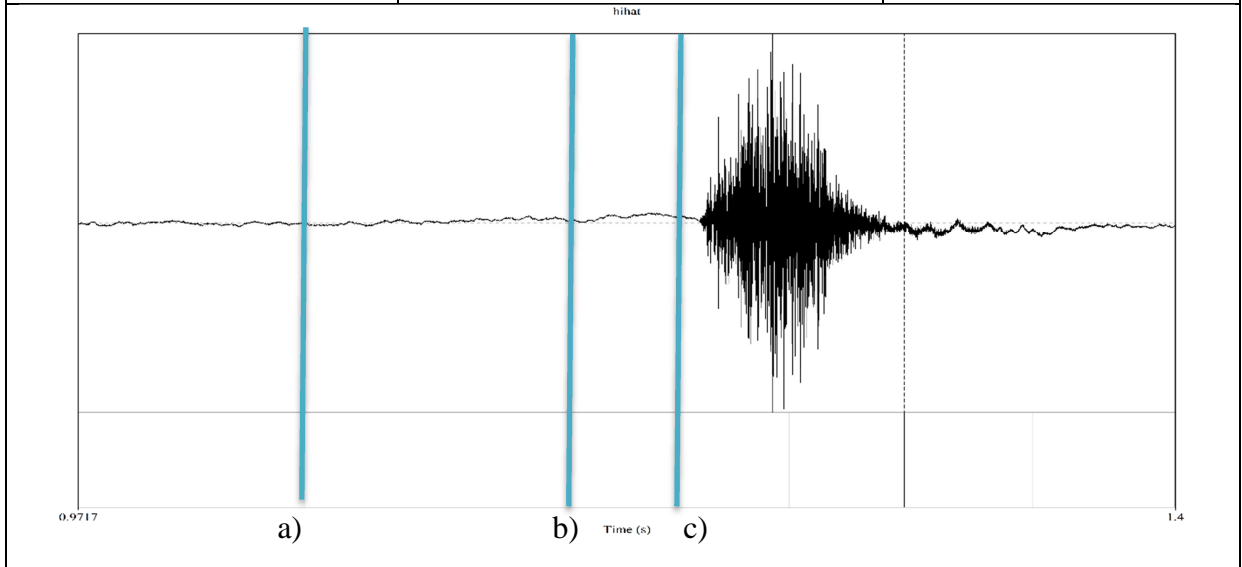
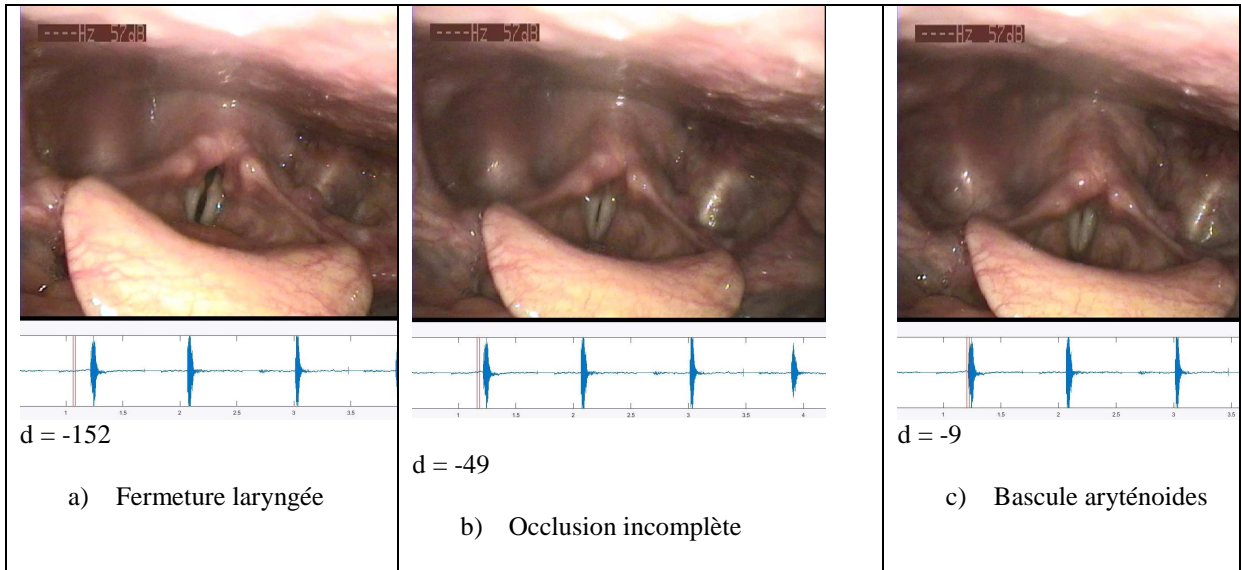


Figure 19 : illustration de la production d'un /t/ du sujet 1 dans /to/ de « pâtes au pesto ».

**Dans le cadre du Beatbox**, pour l'effet hihat typique, la fermeture glottique débute vers -180ms, suivie d'une bascule avant de l'épiglotte contemporaine de la dilatation de l'hypopharynx (-100ms environ). L'espace rétrocricoïdien s'ouvre à partir de -50ms, en synchronisation avec un verrouillage des cartilages aryénoïdes juste après la fin de la fermeture glottique, puis survient un mouvement d'ascension laryngée contemporain d'une contraction pharyngée. La fermeture glottique nous paraît incomplète dans 30% des cas avec la présence d'une très petite ouverture de la commissure postérieure comblée par la bascule et le verrouillage aryénoïdien. Dans 20% des cas un doute subsiste sur la fermeture, les autres images sont trop éloignées temporellement pour nous donner cette information.



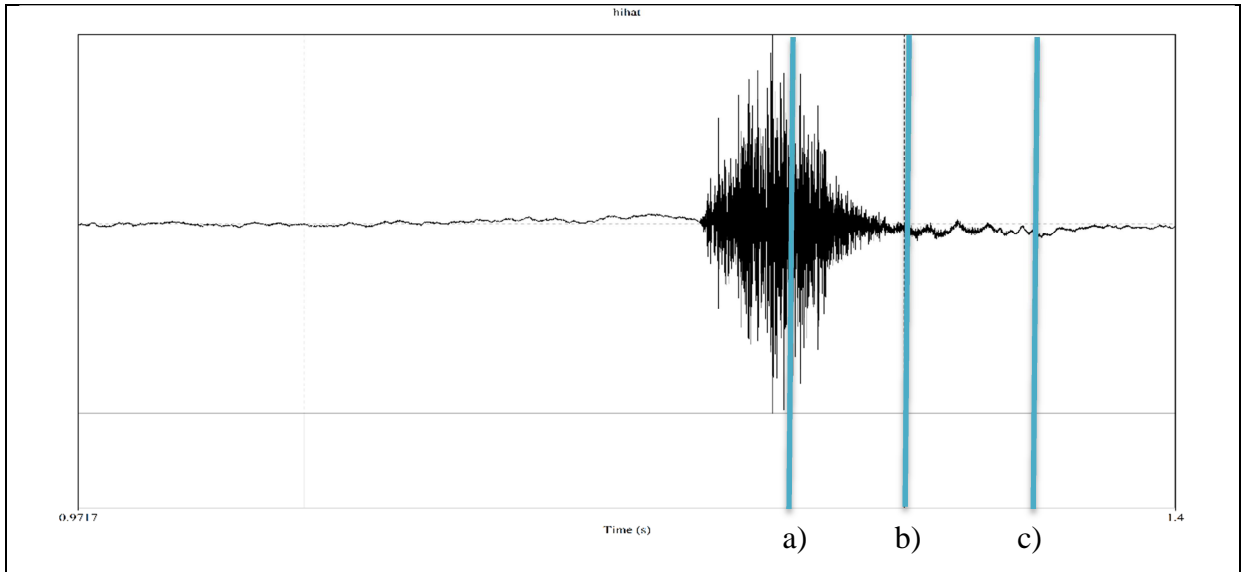
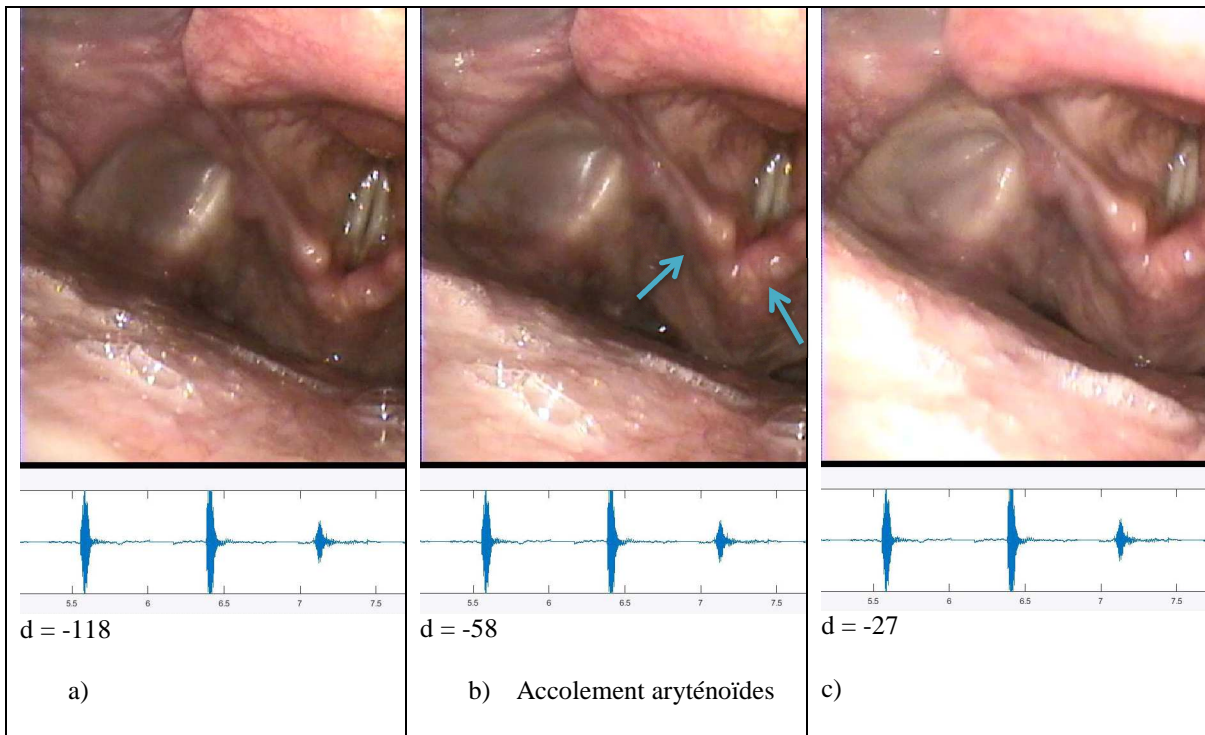
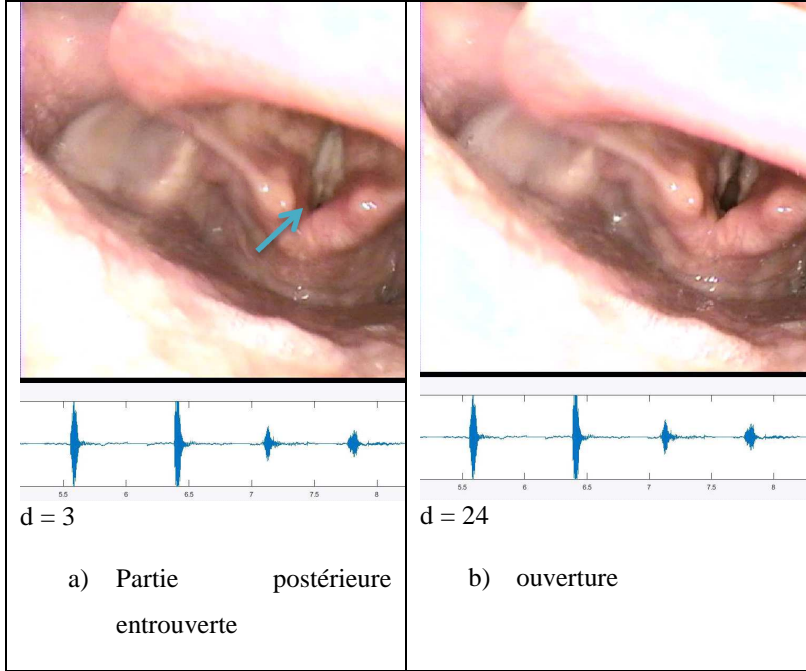
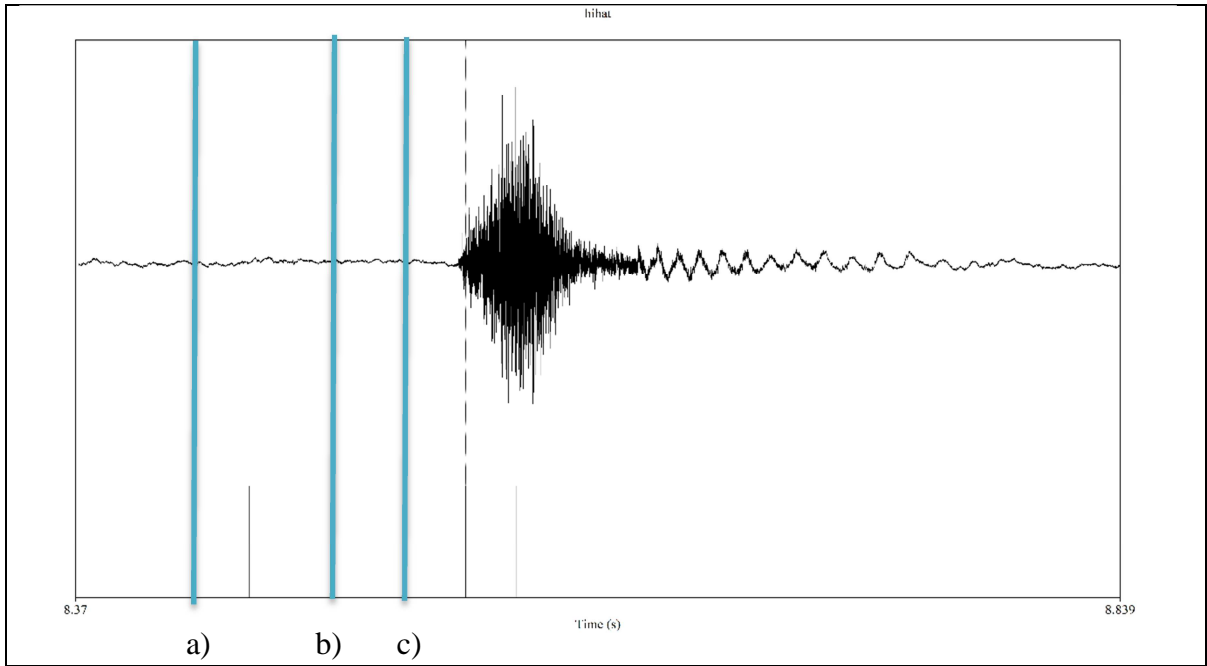


Figure 20 : Séquence temporelle d'un hihat.





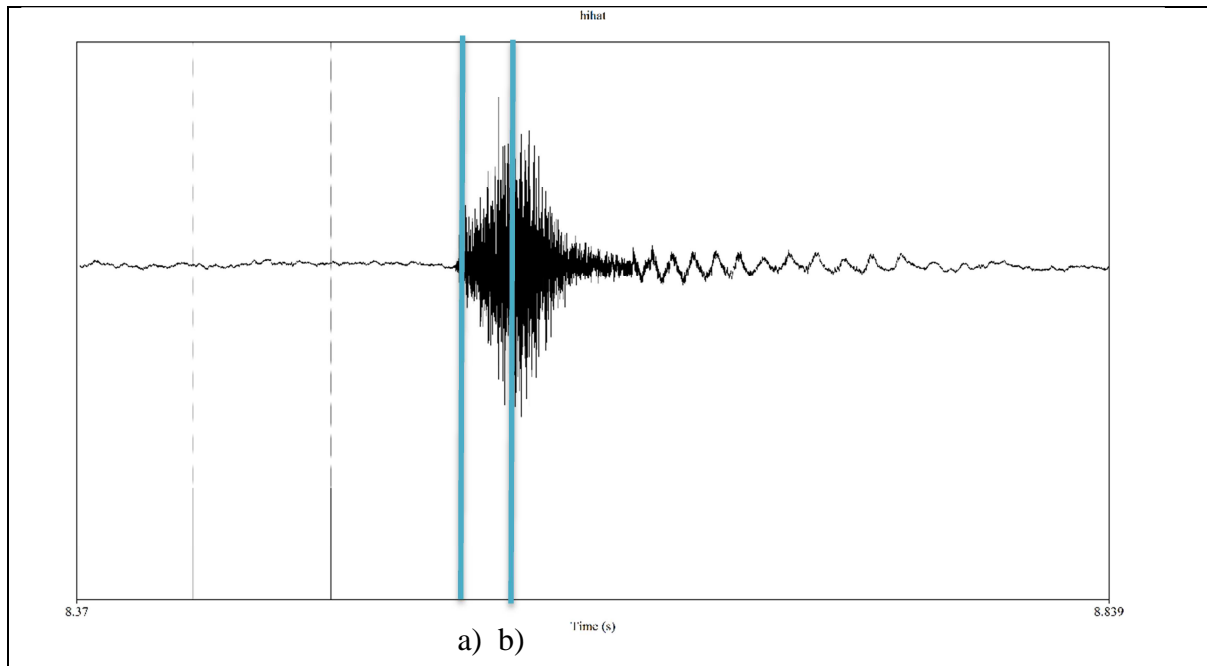


Figure 21 : Séquence de hihat en vue postéro supérieure, on voit bien le verrouillage aryténoïdien en figure 2 et 3. On peut noter le petit triangle ouvert postérieur mentionné ci-dessus.

Nous avons deux variantes du hihat retrouvées sur le corpus consistant en :

\_ hihat rapide appelé fast hihat, {tk} en SBN, d'intensité moindre (intensité maximale de 48dB vs 72dB pour un hihat classique, intensité moyenne de 35dB vs 64dB,  $pvalue < 0.01$ ). Ils sont produits en enchaînement rapide à la suite de hihat classiques, la glotte est restée en état stable de fermeture incomplète telle que décrite précédemment, avec une contraction pharyngée légère dont la temporalité encadre le burst (-2ms à 7ms en moyenne).

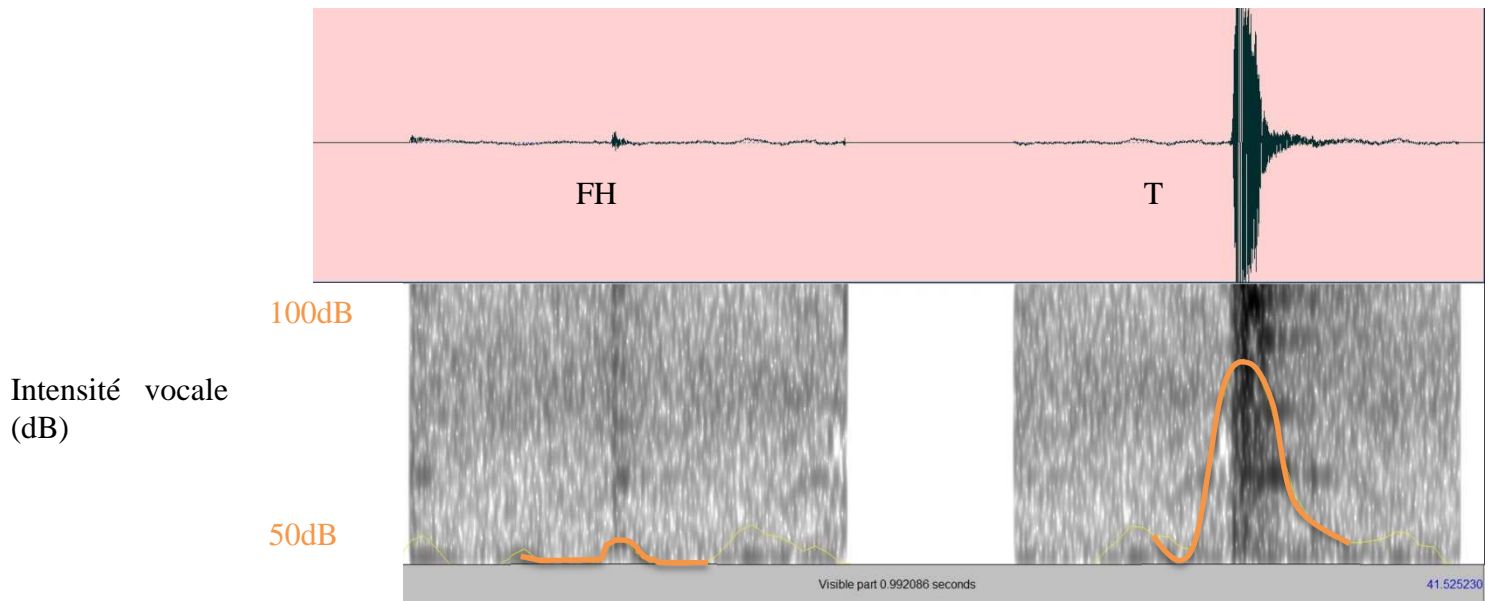
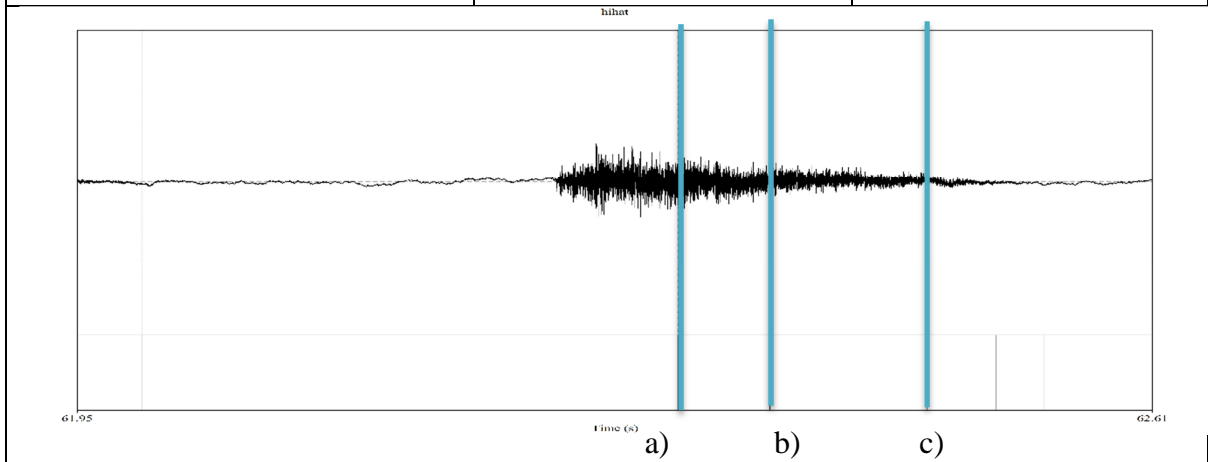
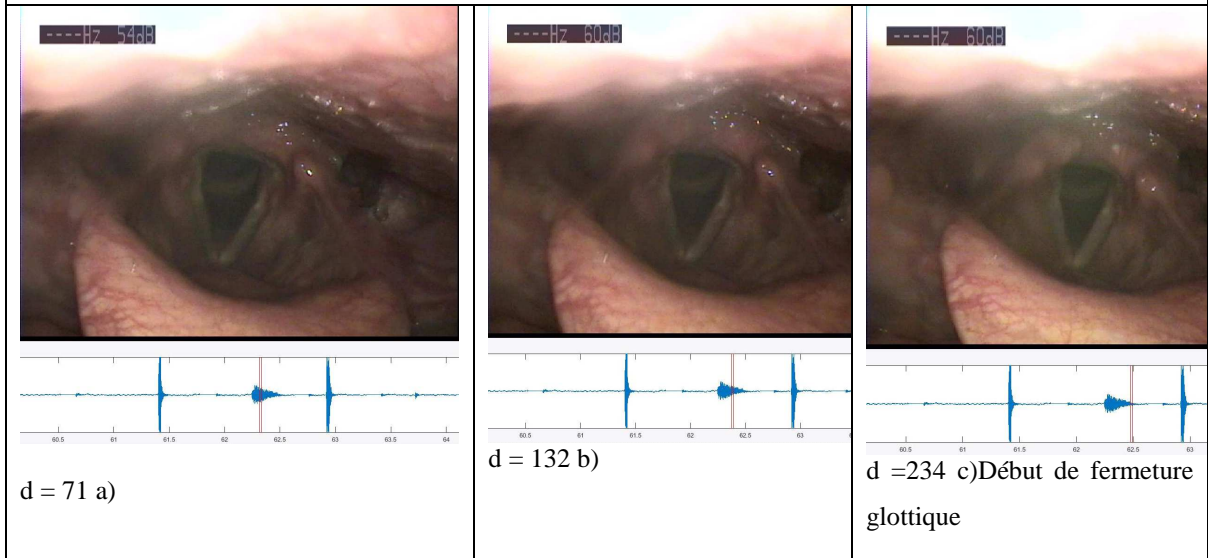
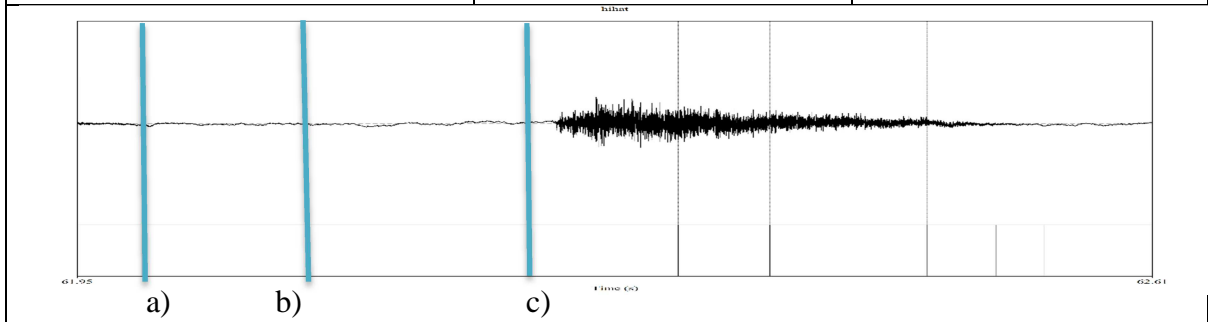
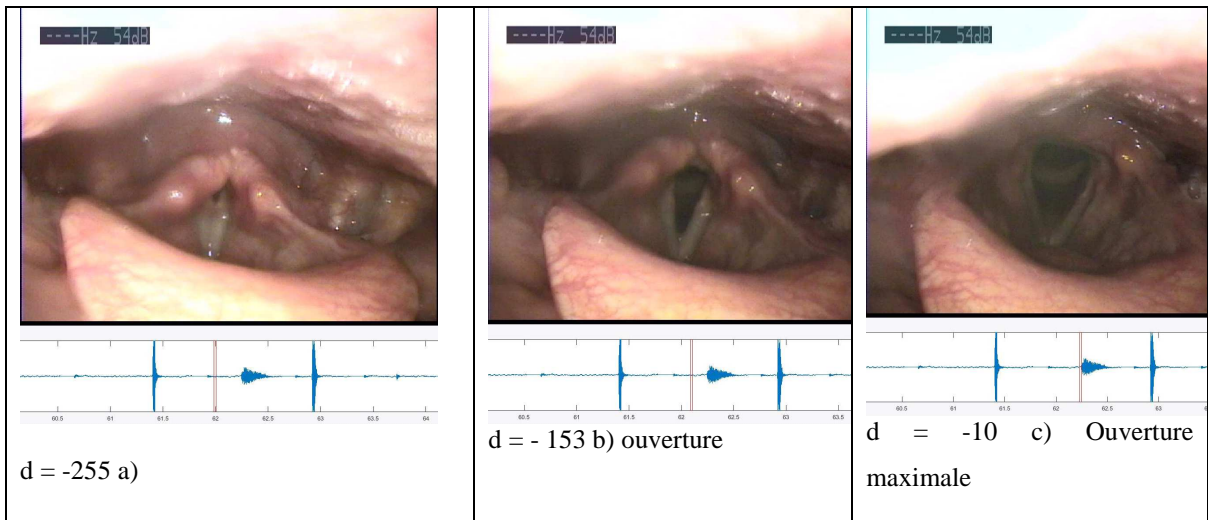


Figure 22 : signaux acoustiques et spectrogrammes (échelle de 0 à 5000 Hz) de Fast hi-hat (FH) vs classic hi-hat (T). La courbe pleine sur le spectrogramme présente l'intensité acoustique (non calibrée).

\_ inward hi-hat {^t}, produit pour donner un effet de cymbale long avec pédale. Il est réalisé en inspiration, les plis vocaux s'ouvrent de -250 ms à -10ms jusqu'en position O3 (ouverture maximale), puis se referment brusquement vers +200ms jusqu'à une occlusion complète glottique vers 310 ms correspondant à la fin de l'effet d'un point de vue acoustique.



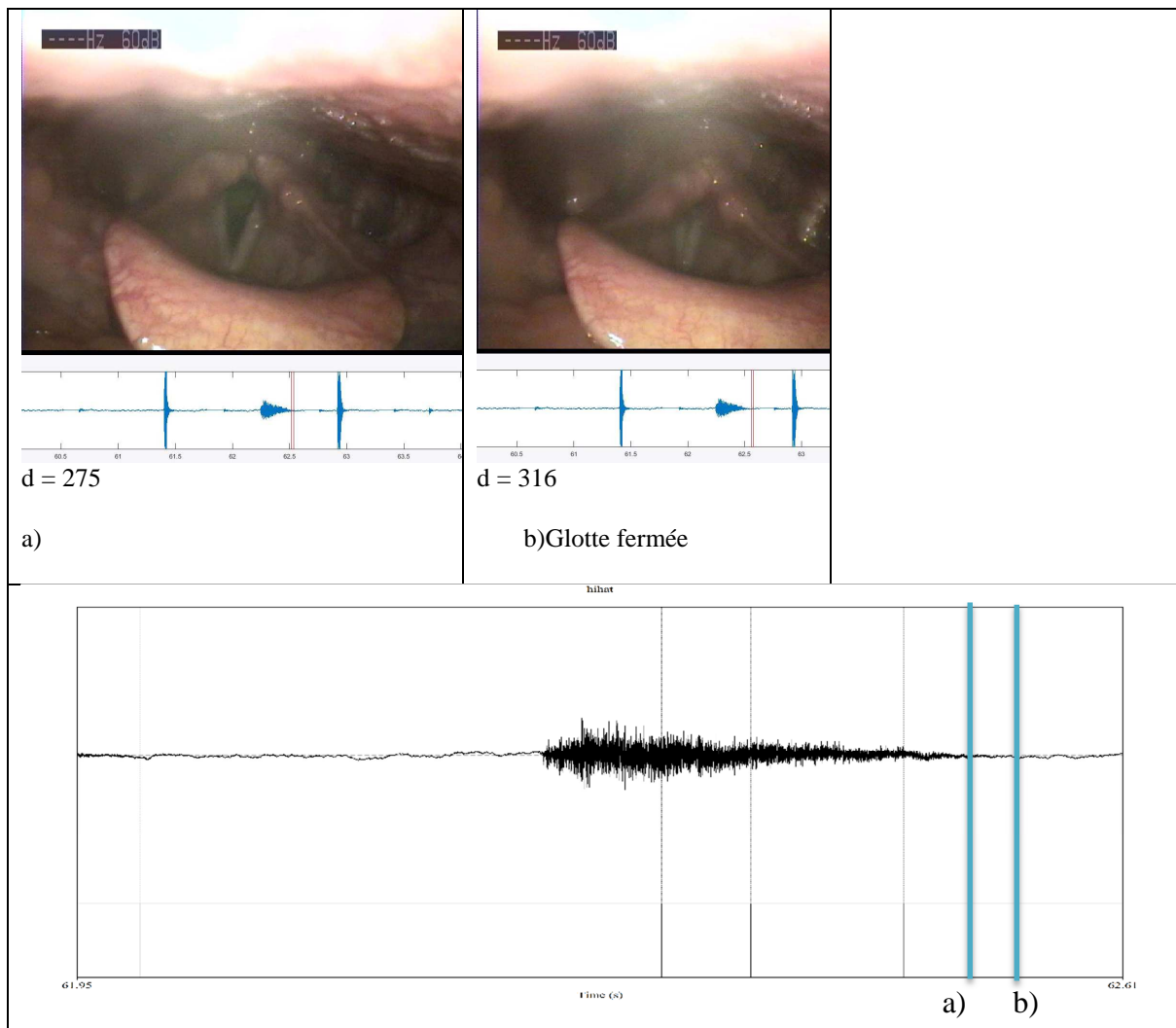


Figure 23 : Séquence laryngée d'un Inward hihat.

Ces variantes ont été réalisées à notre demande mais sont utilisées dans les séquences libres d'enchaînement de sons percussifs. Pour les fast hihat, dans notre corpus, nous les retrouvons systématiquement par groupe de deux et souvent situés entre deux temps rythmiques forts. Le Inward hihat est produit en fin de phrase rythmique en séquence libre ou pendant l'enchaînement d'accélération pour la reprise du souffle en milieu de phrase mais est dans ce cas de durée plus courte.

### 1.2.2. Sujet 2 : beatboxeur amateur

Pour le sujet 2, il n'y a pas de modification substantielle de la dynamique laryngée par rapport au sujet 1. Ce sont des éjectives pures avec fermeture glottique et ascension laryngée. Les hihat ont d'ailleurs une meilleure proximité acoustique avec les productions du sujet 1, que les kicks (cf au-dessus).



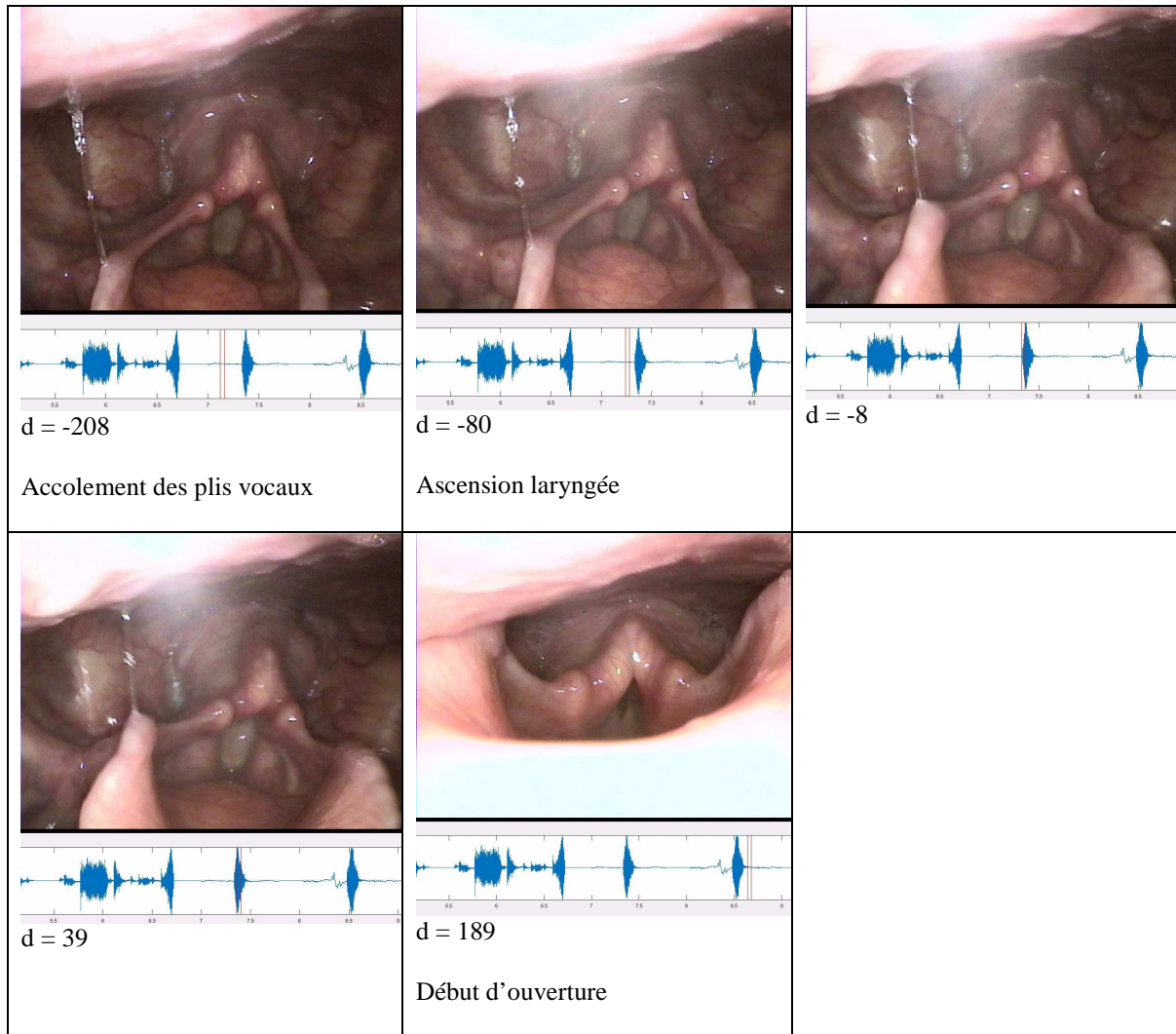


Figure 24 : illustration de la production d'un hihat par le sujet 2.

### 1.3. snare - /Ps/

#### 1.3.1. Sujet 1 : beatboxeur professionnel

Pour la parole, n'ayant pas dans le corpus de réel /p/ ou /ps/, nous avons vérifié que le /pes/ de « pesto » ou le /py/ de « puche » ne présentaient pas de variation articulatoire importante par rapport à la rapidité de production de l'énoncé (80/s, 120/s et 160/s). En dehors d'une abduction plus rapide après le burst des plis vocaux, +50ms au lieu

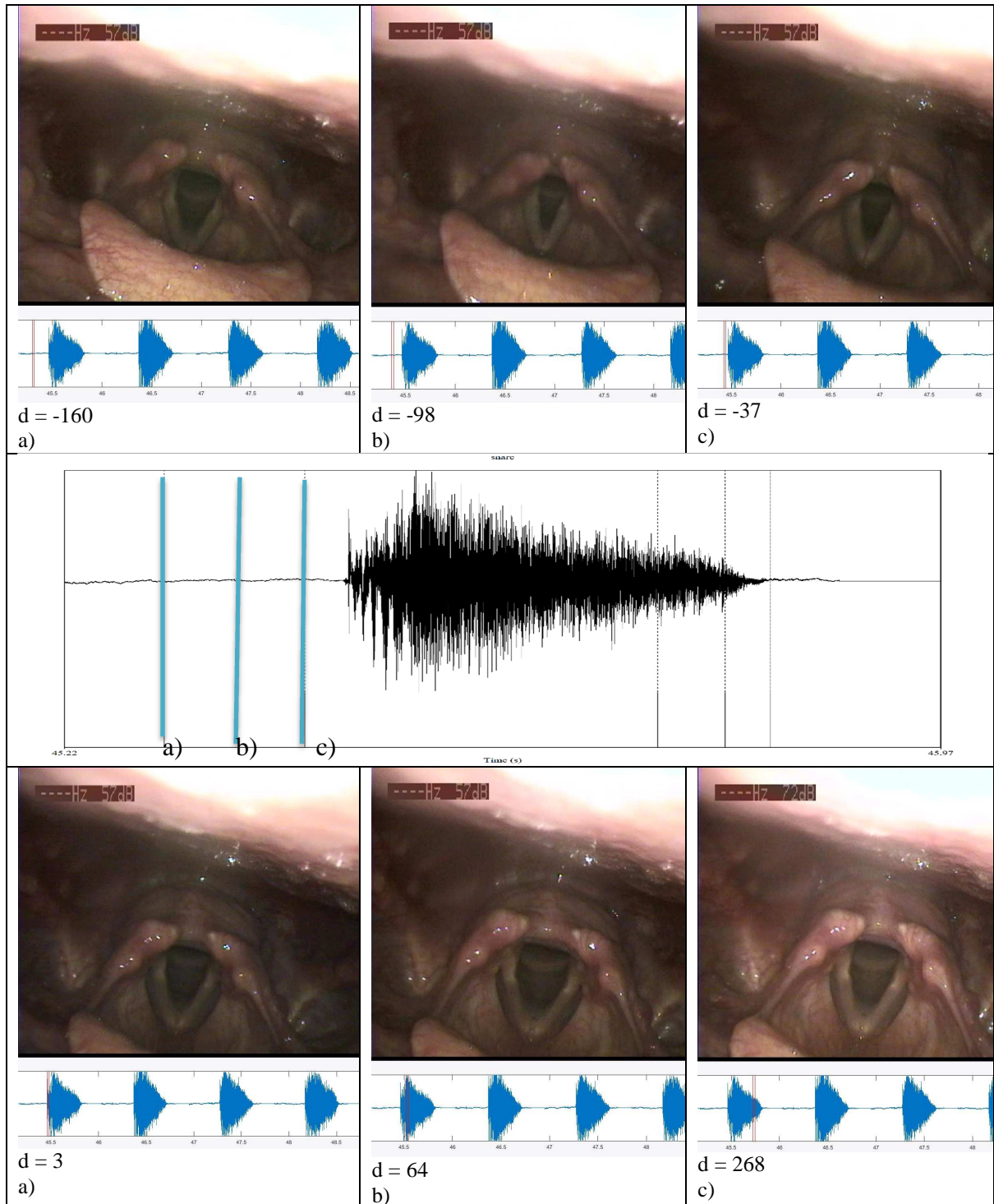
de +90ms correspondant à un raccourcissement en durée de la syllabe, nous n'avons pas relevé de différence par rapport aux gestes de production du /p/. De même au niveau acoustique il n'y a pas de modification des durées de burst et reprises de voisements (moyennes de VOT = 15.8ms et 12.9 ms respectivement pour les productions à 80/s et 160/s, test de Wilcoxon : p-value = 0,31). Le continuum entre parole et beatbox ne se fait pas de façon continue, probablement en raison de la voyelle intercalée entre la plosive et la fricative.

**En Beatbox**, l'effet snare semble être l'effet le plus variable de notre corpus au niveau de la gestuelle et acoustique selon l'intensité de la plosive et le type de snare (/ps/ /p/), pvalue = 0.03).

Les mesh-snares (/p/ ou /ps/ allongé) sont retrouvées les plus fréquemment, elles ressemblent acoustiquement à l'inward hihat, c'est-à-dire de type fricatif long, mais avec un bruit de plosion initial très prononcé et sont produites en expiration. Elles débutent vers -200ms par un mouvement de fermeture glottique, qui se finit environ 10ms avant le burst. L'occlusion semble incomplète (état O1) dans 60% des cas. En fin de fermeture glottique, il y a une ascension laryngée, suivie d'une ouverture très brutale d'environ 20ms, correspondant à l'intensité maximale acoustique jusqu'en ouverture O3. La glotte reste dans cet état d'ouverture maximale et en position haute avec un aspect de contre résistance au flux d'air : les plis vestibulaires sont arqués avec un ventricule particulièrement visible, des mouvements minimes des aryténoïdes sont perçus pendant la durée de la fricative (comme un cerf-volant dans la brise). Advient ensuite une fermeture de l'étage glottique correspondant à la décroissance de la fricative non systématiquement complète vers 430ms qui est la durée moyenne de l'effet toujours. Cette phase de fermeture se fait la plupart du temps avec l'aspect de contre résistance (ventricules ouverts). La dilatation de l'espace rétrocricoïde et hypopharyngé est particulièrement importante. L'espace rétrocricoïdien est visible en même temps que l'ascension laryngée. La position haute du larynx est de durée variable, tandis que l'hypopharynx est dilaté pendant la totalité de l'effet avec un effet rebond exagéré au moment de l'occlusion glottique.

Se pose la question de la part éjective dans ces productions : la glotte n'est pas fermée, pourtant le larynx passe en position haute, peut être sous la poussée de pression d'air pulmonaire ou un état stable « quantique » provoqué par la participation de la musculature extrinsèque laryngée.

En raison de la saturation du microphone utilisé pendant l'enregistrement, nous ne pouvons pas affirmer la synchronicité temporelle avec les maxima d'intensité.



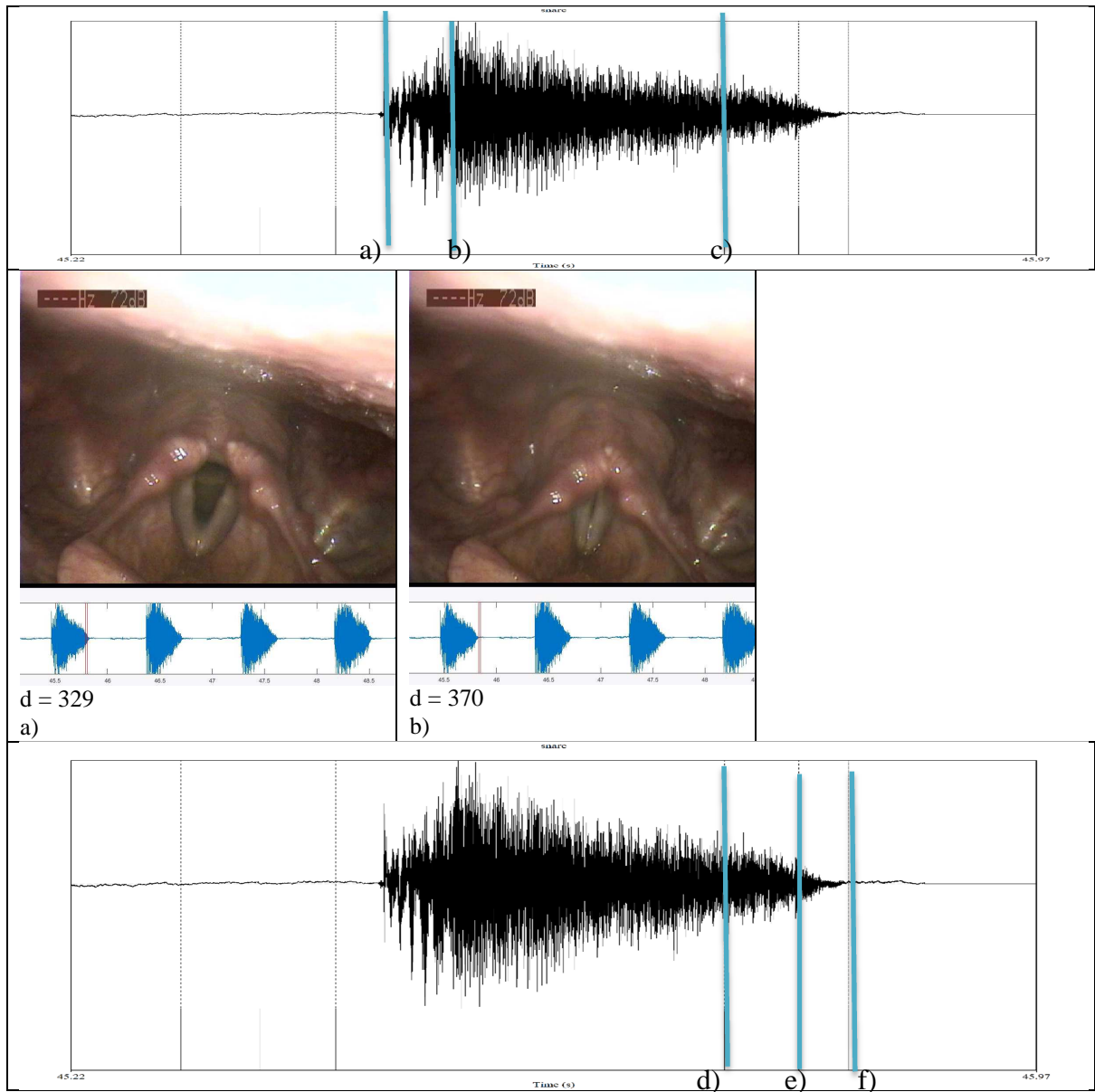
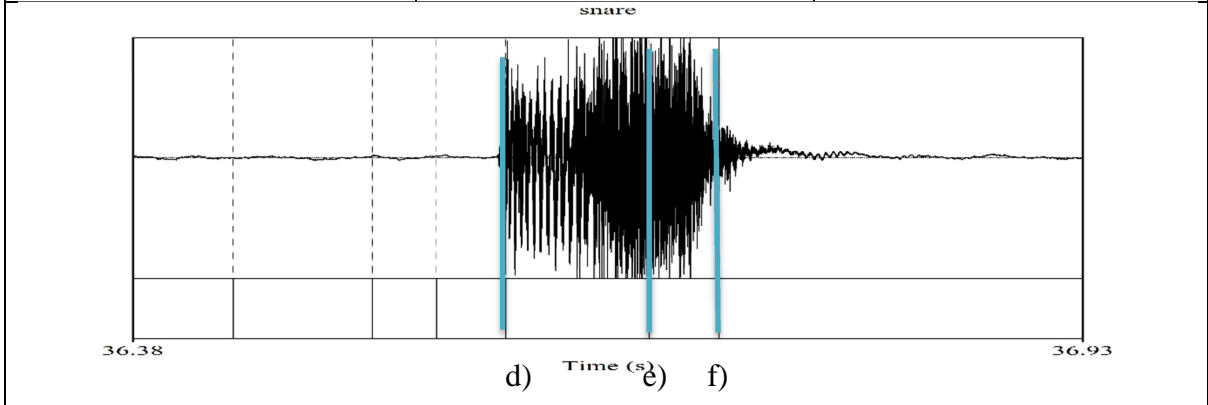
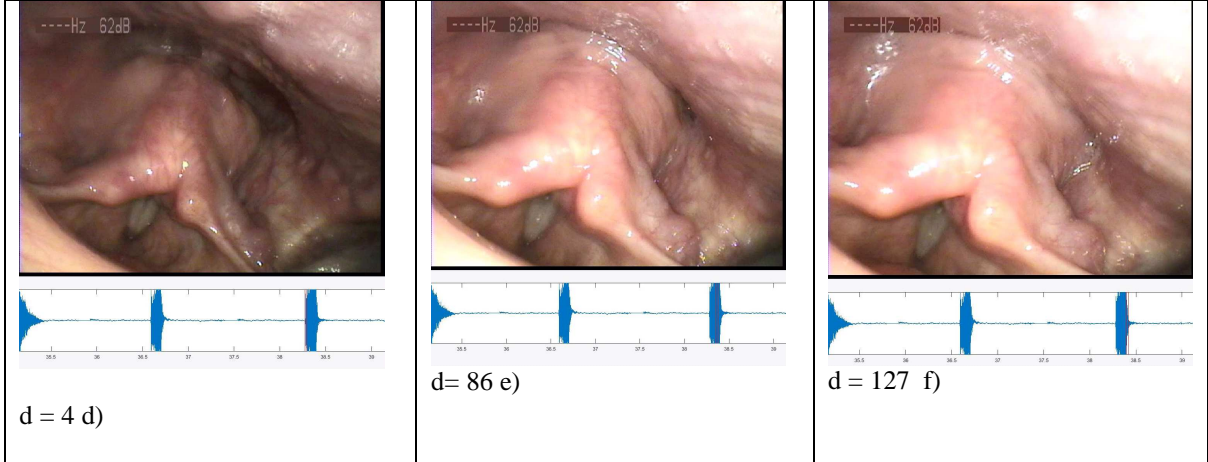
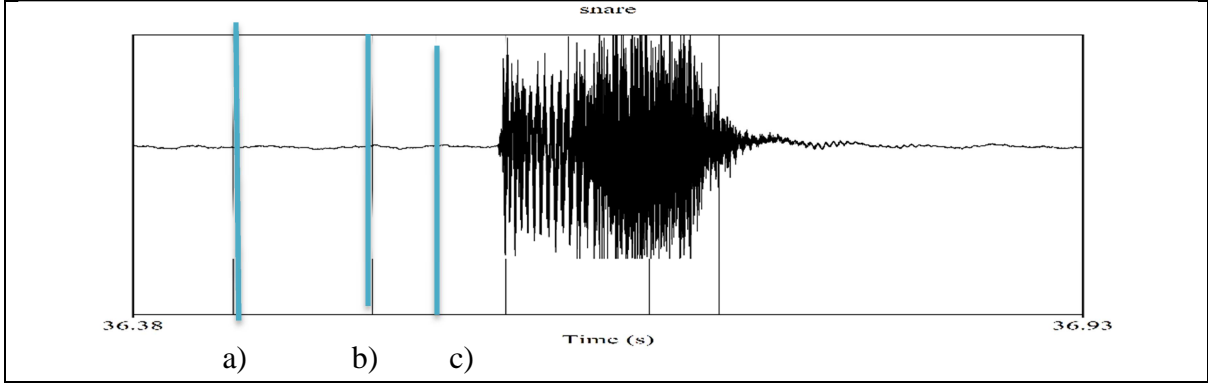
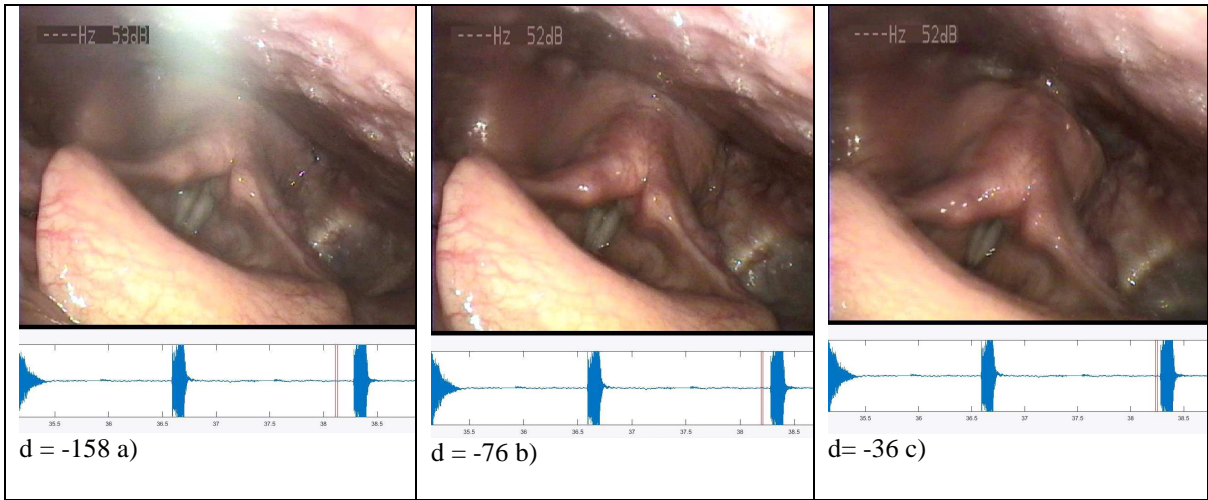


Figure 25 : production du Mesh snare et sa séquence gestuelle.

Nous notons la présence de constriction épilaryngée avec fermeture du sphincter aryépiglottique dans les effets /pf snare/ et un réel mécanisme éjectif avec ascension laryngée par exemple :



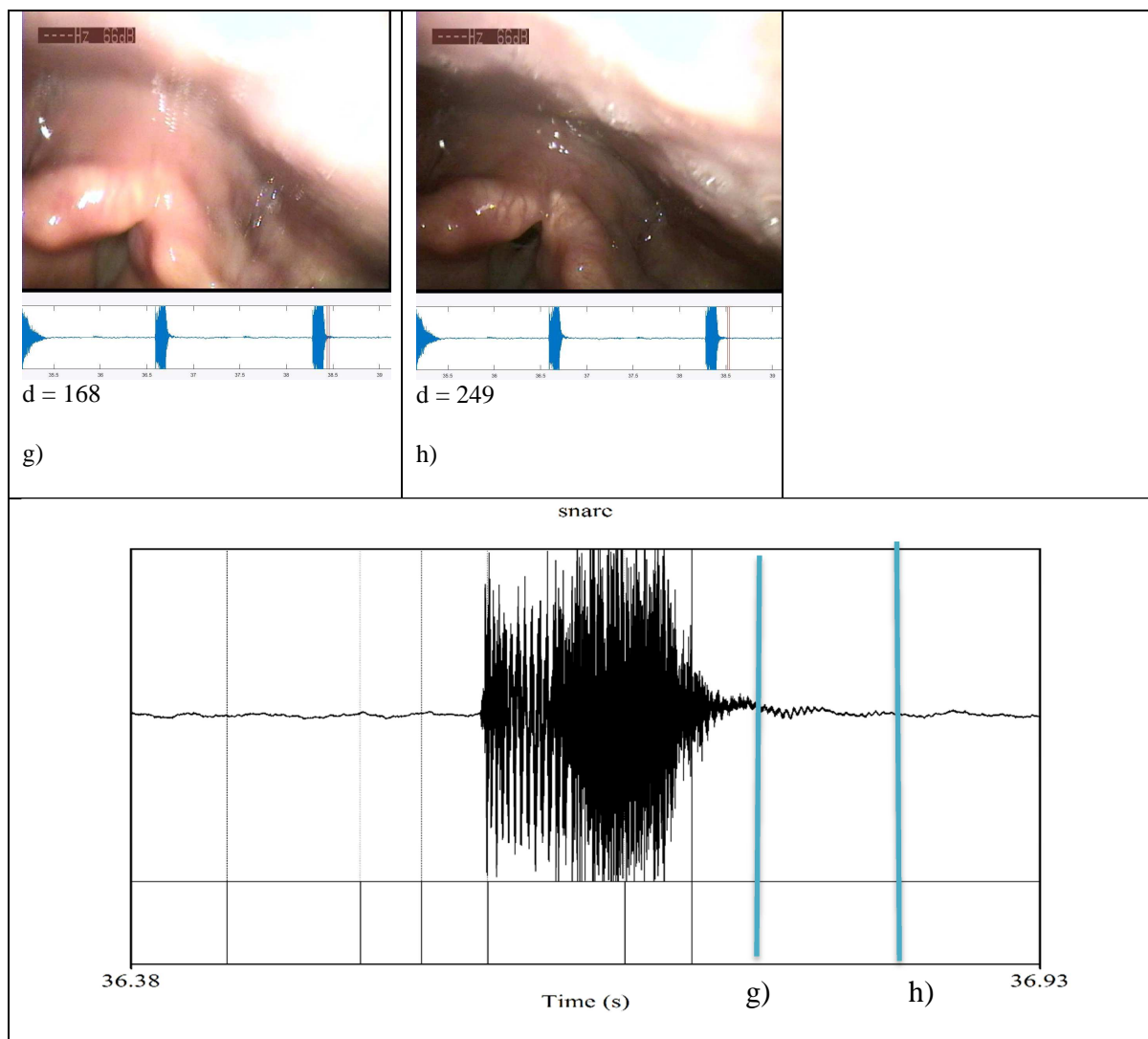


Figure 26 : /pf snare/ et séquence vidéo.

De plus nous tenons à signaler que malgré la consigne pour les variantes, le sujet n'a pas réalisé la variante inward (ingressive) qui a été cependant enregistrée sur la partie aérodynamique. Les différentes variantes étaient produites en corrélation avec la phrase de base qui était à suivre :

\_ /puch/-> /pch/

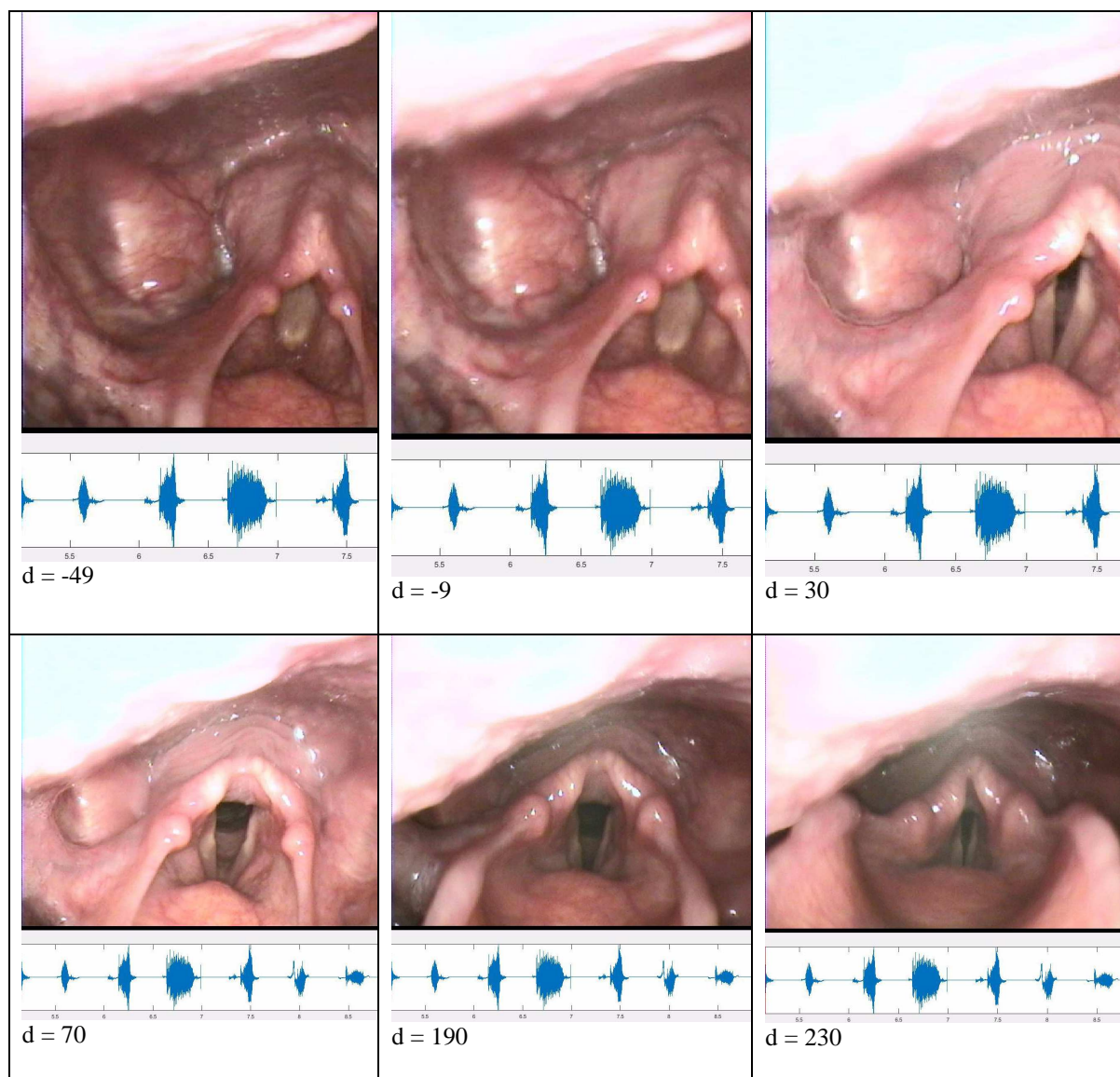
\_ /pes/ -> /ps/ ou /pf/

En production libre elles sont utilisées sans réelle influence sur les consonnes suivantes (fréquence relativement identique  $pvalue=0.65$ ). On note acoustiquement et visuellement sur le signal cependant qu'un certain nombre de productions paraissent plus « approximatives » que lors de la consigne d'effets seuls.

### 1.3.2. Sujet 2 : beatboxeur amateur

Notre deuxième sujet réalise principalement des snares « typiques » de type éjectif comme le sujet 1.

Pour les « mesh snares », on perçoit un bruit de plosion beaucoup plus important que le sujet 1 correspondant aux mouvements éjectifs, puis le larynx s'ouvre pour la partie fricative et elles se finissent plus abruptement par une constriction glottique et épilaryngée.



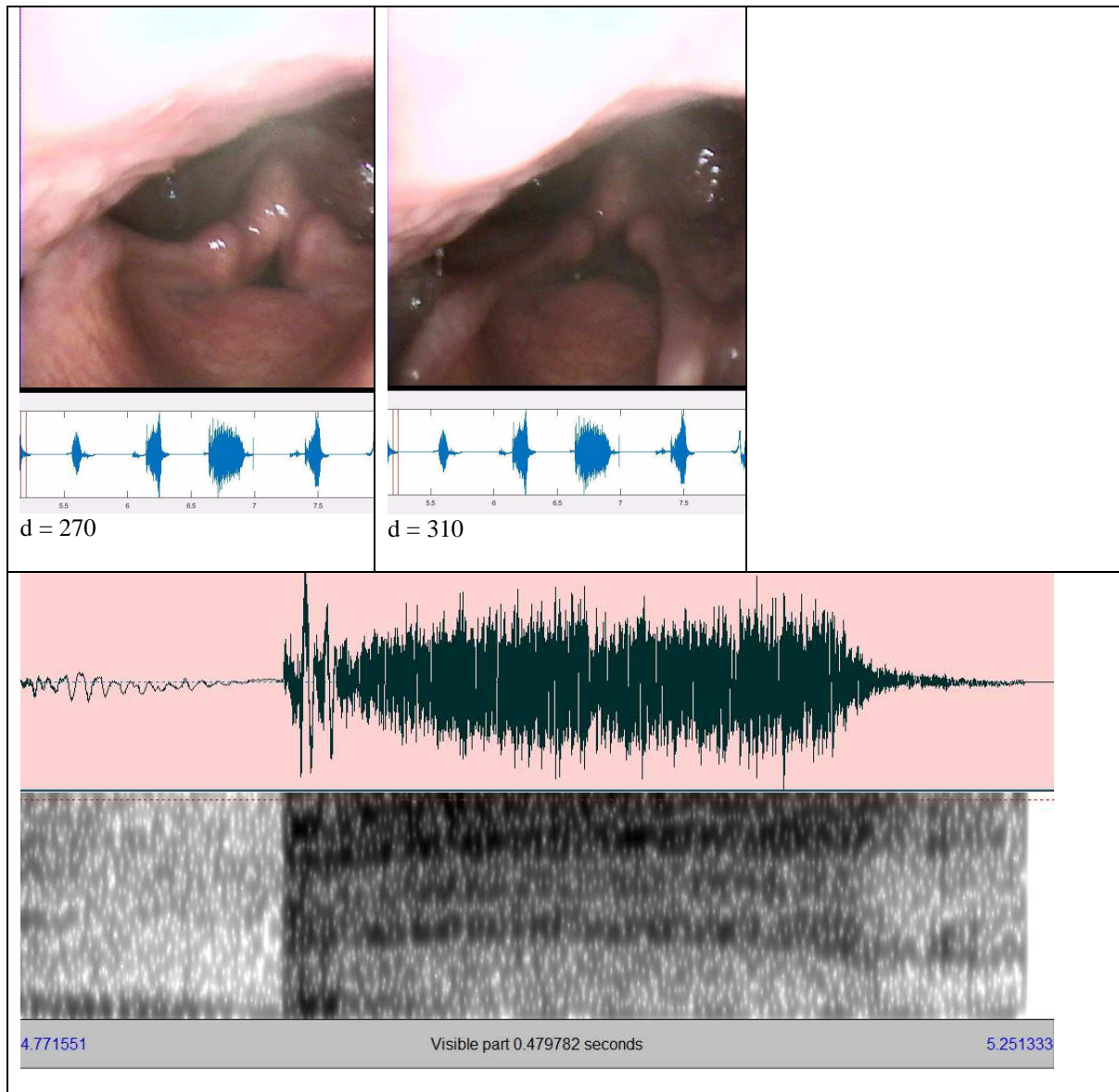


Figure 27 : Mesh snare du sujet 2.

#### 1.4. rimshot - /k/

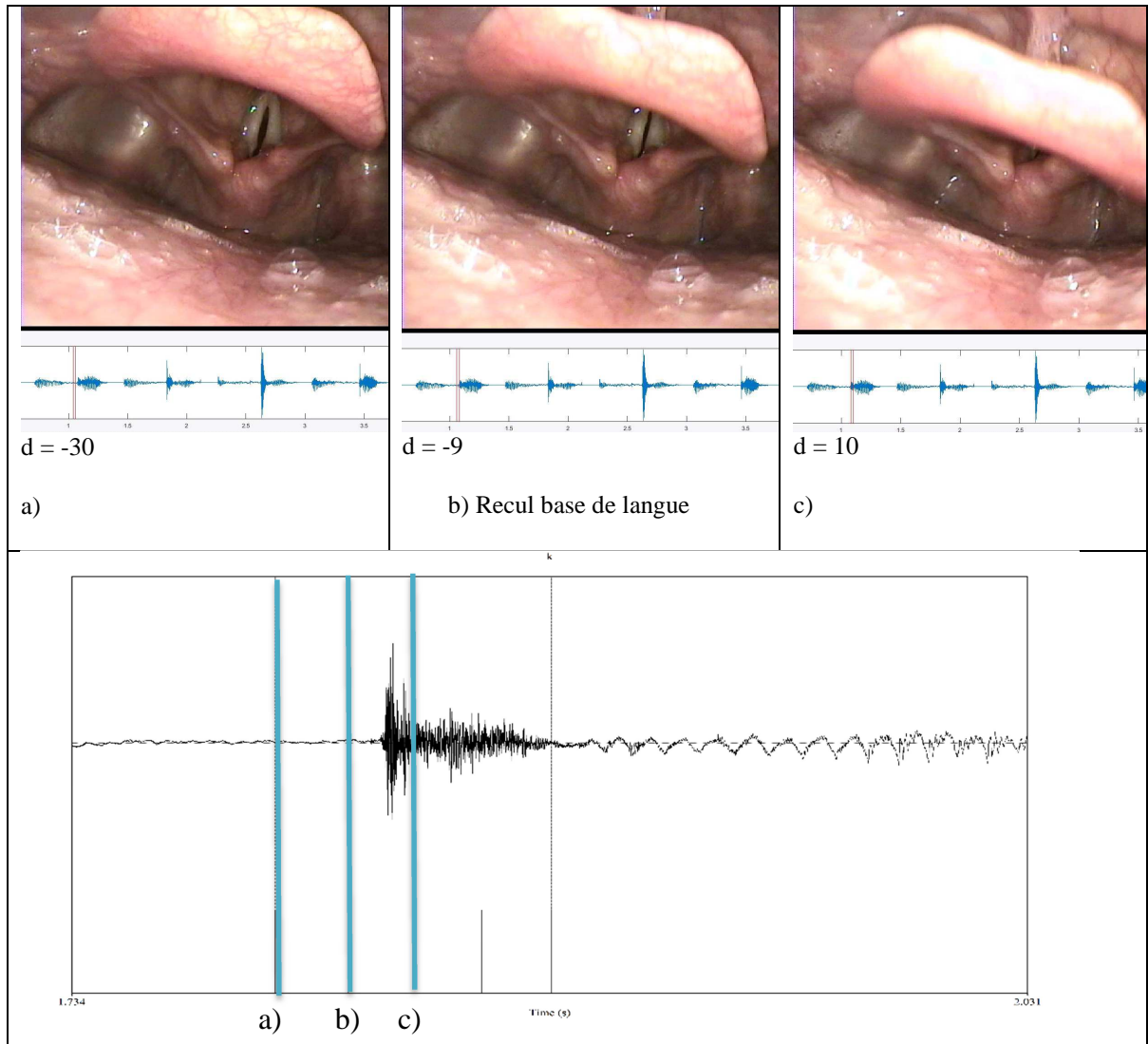
##### 1.4.1. Sujet 1 : beatboxeur professionnel

**En parole,** Le /k/ est ici prononcé dans des contextes précédents différents (/and/ /ta/ et /ti/) avec une voyelle suivante similaire (/a/).

Les plis vocaux semblent rester en position entrouverte tout au long de la consonne, la base de langue part en arrière dans les 20ms précédant la consonne, accompagnée d'une contraction des parois pharyngées latérales qui forment un tube étroit autour de l'épiglotte pendant la production de la voyelle (+165ms). La vue postérosupérieure ne nous permet pas de déterminer l'état glottique au moment de ces mouvements en raison de l'étroitesse



du tube ainsi créé. Il apparait de façon empirique que l'importance du recul de la base de langue et de la contraction pharyngée soit en rapport avec l'intensité du /k/.



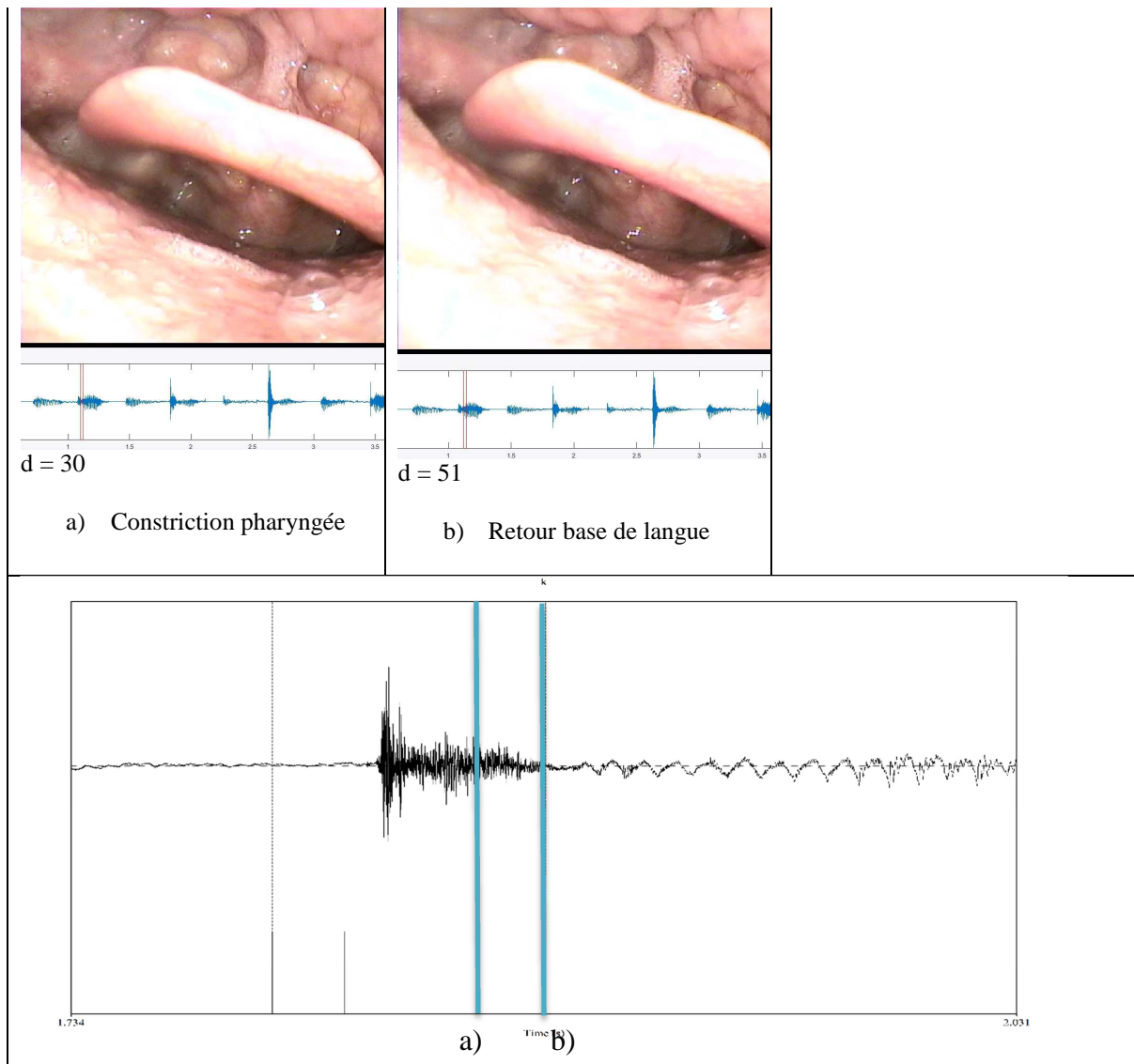
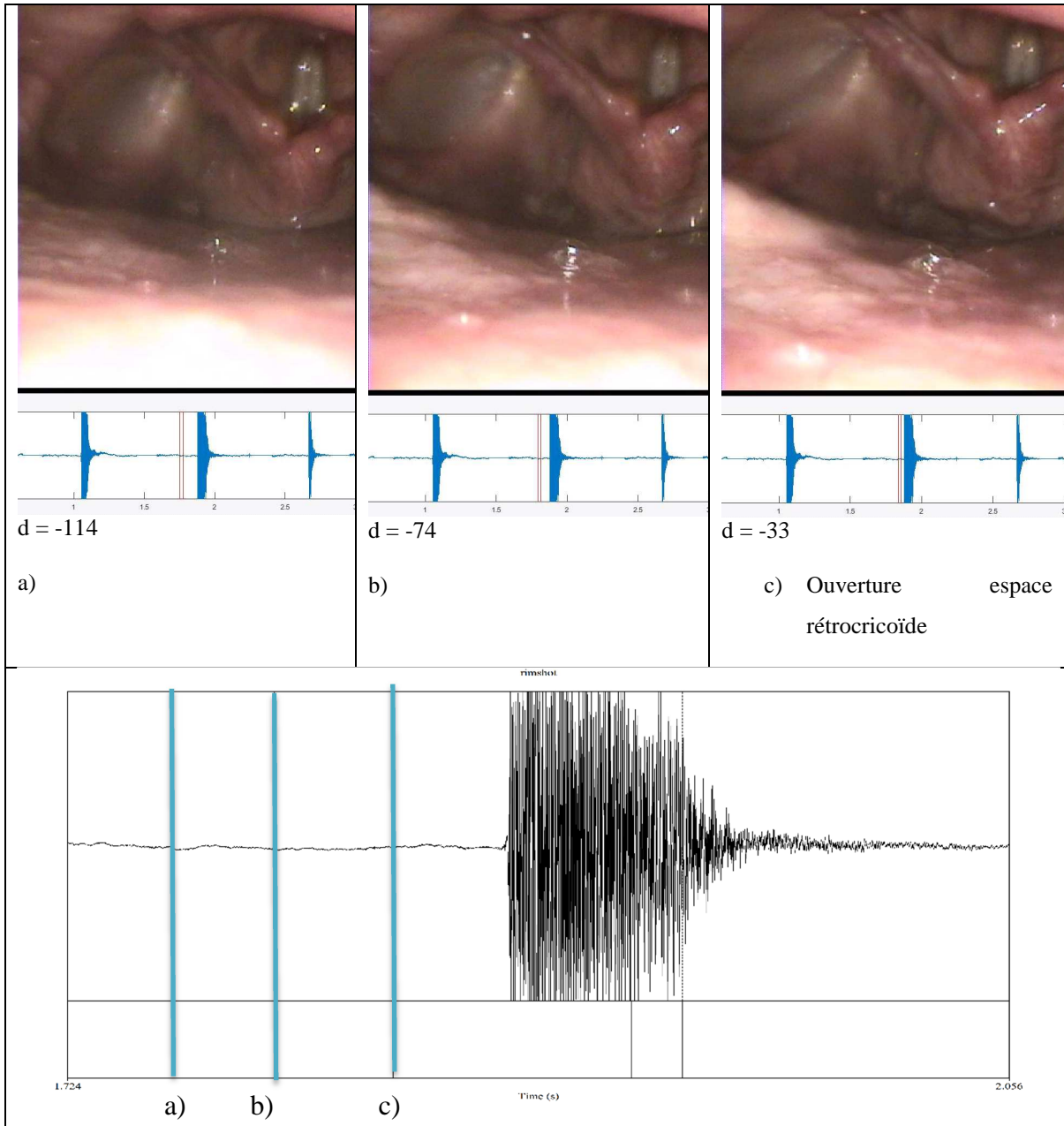
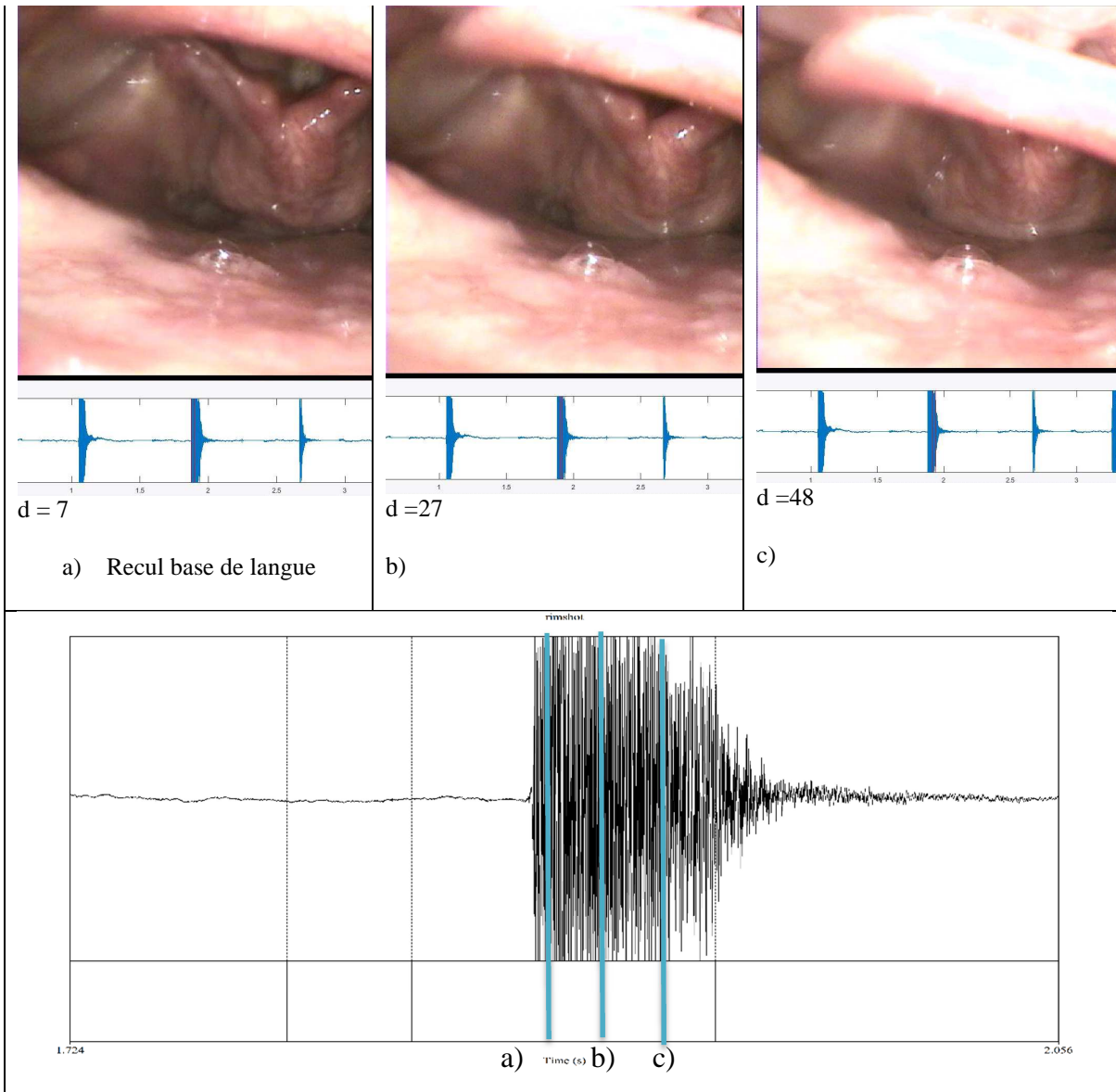


Figure 28 : séquence gestuelle laryngée du /k/ dans « poutika ».

Le Rimshot du **Beatbox**, est une exagération du /k/ avec des mouvements plus amples de la base de langue et de l'épiglotte, une contraction pharyngée plus importante provoquant une réelle constriction du sphincter pharyngoépiglottique. La phase de transition est relativement continue, bien que nous ne puissions pas réellement évaluer l'état glottique sur une majorité d'effets. Les plis vocaux en vue postérosupérieure sont en position fermée à -150ms, début de dilatation hypopharyngée, moment de la bascule épiglottique et du recul de la base de langue. La contraction pharyngée débute environ 50ms avant le burst et persiste jusqu'à 60ms, et est accompagné de la dilatation de l'espace rétrocricoidien. Il n'y a pas de modification de hauteur du larynx.





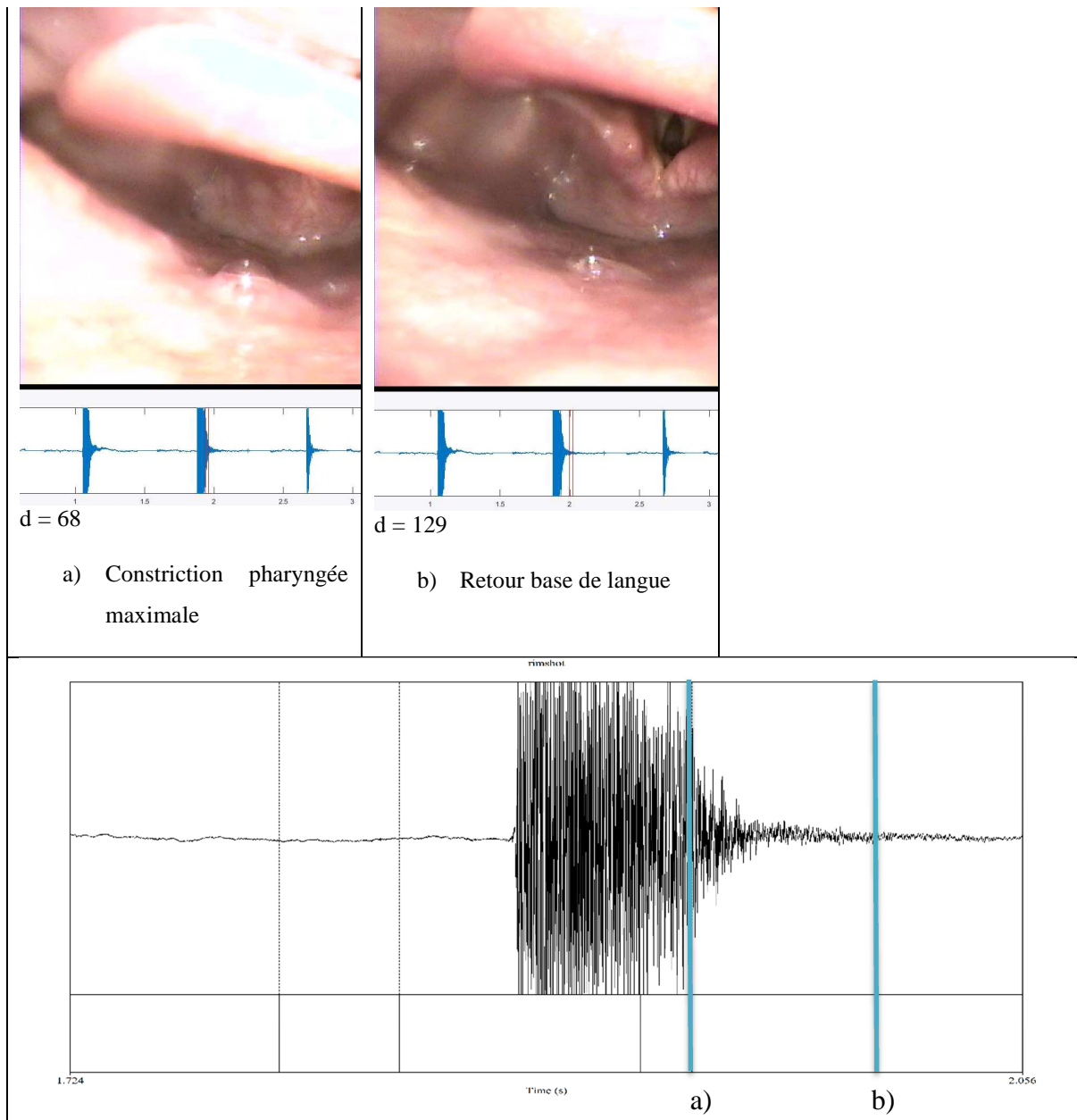


Figure 29 : Rimshot du sujet 1 en vue postérosupérieure.

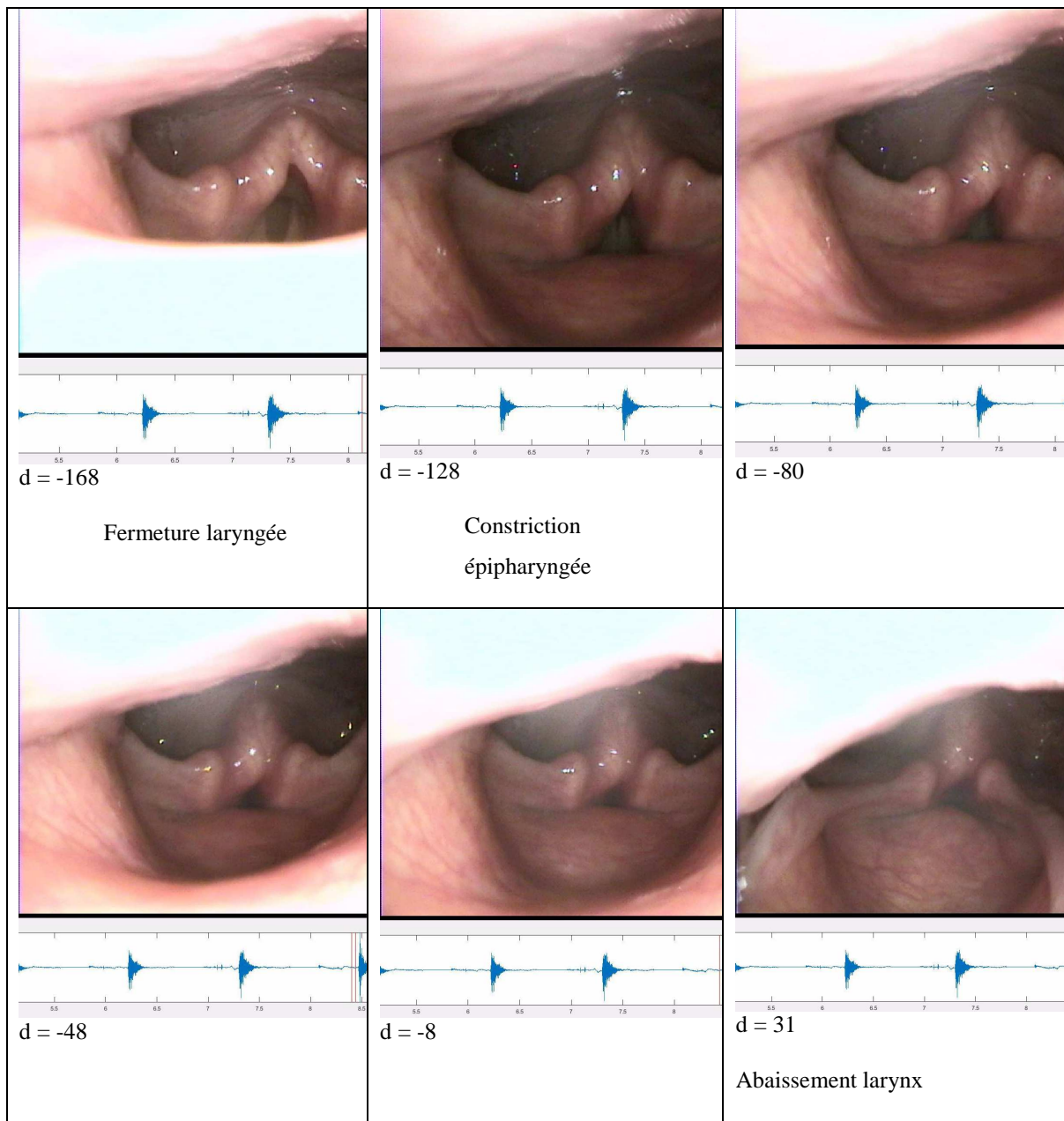
Comme pour les snares, les données acoustiques sont saturées et n'avons donc pas pu analyser les données acoustiques de façon plus poussée. Les variations portent sur le timbre allant d'un équivalent de /ko/ à /kri/. Nous ne notons pas de modification des patterns pharyngés, il semble essentiellement que cette variation soit d'origine buccolinguale.

#### 1.4.2. Sujet 2 : beatboxeur amateur

La même remarque concernant la tendance intempestive et difficile à supprimer à donner à ses consonnes un « type Beatboxé » (22 occurrences). La phrase « biscotte petite biscotte » qui nous a été donnée par le sujet a été ajoutée dans le corpus. De plus, le micro

n'étant pas le même entre les sessions d'enregistrement nous n'avons pas pu faire de comparaison des signaux pour savoir si l'intensité du signal était la même en intersujets ou était la cause des mouvements moins amples du sujet 2.

Les Rimshots produites quasi systématiquement par le sujet 2 sont injectives avec le larynx qui s'abaisse en position fermée avec constriction épilaryngée. Nous avons l'impression que la langue s'avance après la descente laryngée, entraînant l'épiglotte en avant et un étirement des replis aryépiglottiques. Le son produit est plus aigu et semble moins intense.



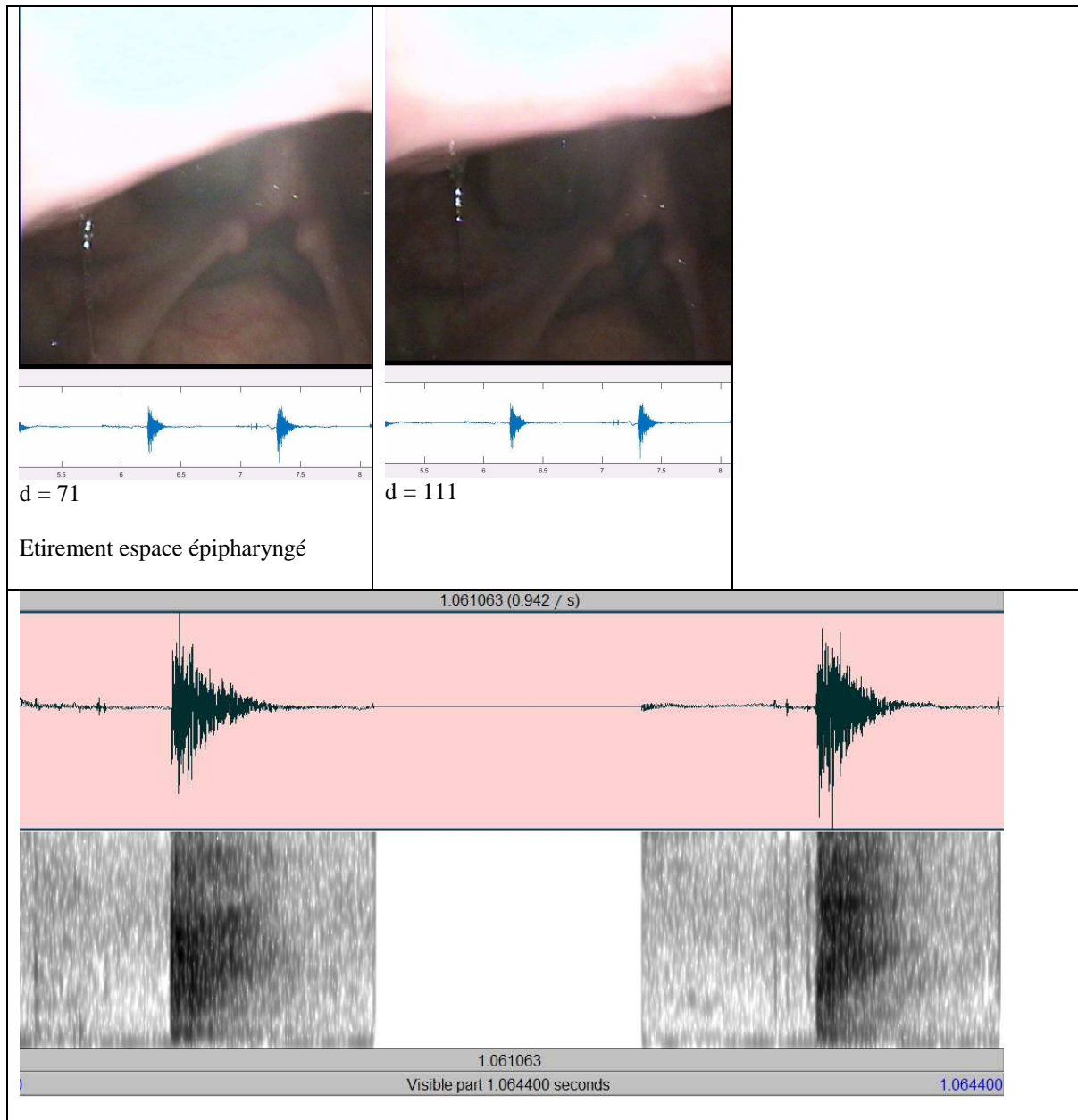


Figure 30 : Rimshot ingressive du sujet 2.

Mais lors de la production de variantes (probablement question de terminologie « ancienne » vs « nouvelle » école, le sujet 2 produit des Rimshots du même type que le sujet 1, c'est-à-dire éjectives avec constriction pharyngée (bien que moins amples) et dont le résultat acoustique est beaucoup plus proche des Rimshots du sujet 1 :

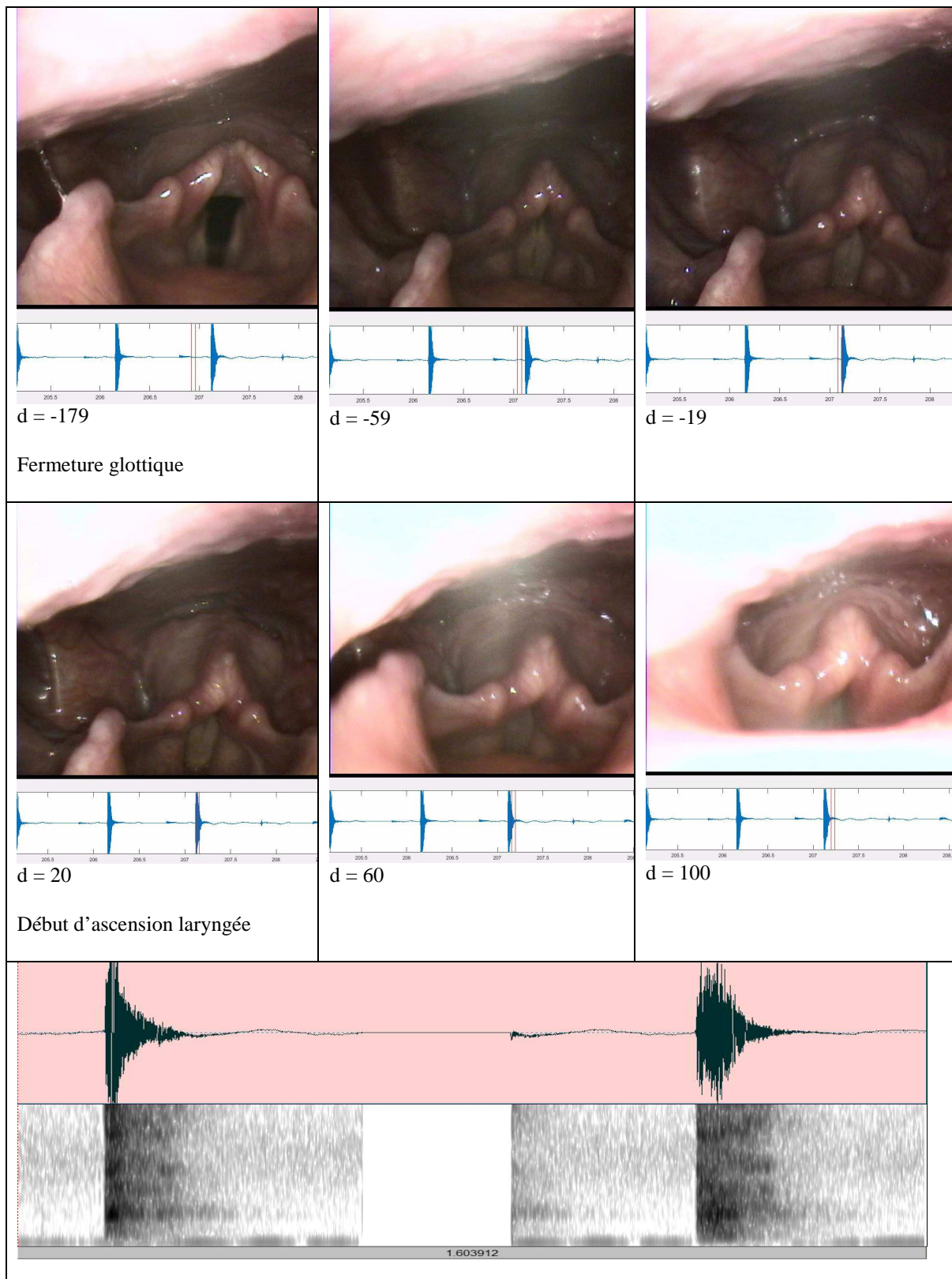


Figure 31 : séquence laryngée de Rimshot « variante » du sujet 2.



### 1.5. Tableaux résumés des données temporelles

		b	p	P	Ps	t	T	k	K
Aryténoïdes	Accolement			-56			-44		
Epiglotte	Bascule Ar			-30			-99	-47	-84
	Bascule Avt								
Retrocricoid	Ouverture			-61	-33		-49		-71
Pharynx haut	Contraction			-31			-13	-29	-59
	Relachement						90		
Hypopharynx	Dilatation	-71	-184	-76	-73		-104	-104	
Plis vestibulaires	Hypercontraction			-83					
Larynx	Ascension			-32	-33		-10		
Plis vocaux	Ouverture		-465	-243	-34		-222		-150
	Fermeture	-319	-228	-182	-242	-93	-181		-253
Etat glottique	Ouvert	-473	-252	-309	-24	-95			-272
	Fermé	-179	-48	-90		-4	-100	-108	-259

Figure 32 : données temporelles des débits de mouvements des structures par rapport au burst.

		b	p	P	Ps	t	T	k	K
Aryténoïdes	Accolement			70			89		
Epiglotte	Bascule Ar			74			143	153	142
	Bascule Avt								
Retrocricoid	Ouverture			72	199		93		101
Pharynx haut	Contraction			95			255	183	109
	Relachement						137		
Hypopharynx	Dilatation	240	209	134	446		141	122	
Plis vestibulaires	Hypercontraction			68					
Larynx	Ascension			99	85		102		
Plis vocaux	Ouverture		143	65	127		170		445
	Fermeture	130	150	117	224	122	136		135
Etat glottique	Ouvert	150	205	161	139	91,5			177
	Fermé	398	57	120		71,5	510	163	176

Figure 33 : Données de durée du mouvement à partir du début du mouvement.

Les tableaux en figure 32 et 33 résument les différents mouvements et les données temporelles et de durée sur une moyenne de 10 effets (6 pour les /b/). Il convient de constater quelques chevauchements sur des mouvements en raison de la présence de légères « variantes ». Il est intéressant de constater principalement la différence des structures utilisées en parole et en Beatbox. On peut noter la similarité temporelle et de durée d'un certain nombre de mouvements comme l'ascension laryngée, entre différents effets.

## **2. Etude aérodynamique et acoustique**

Nous présentons nos résultats sur les données du sujet 1 uniquement. Le but de ce travail étant la mise en relation des données aérodynamiques avec l'endoscopie entre la consonne en parole et son équivalent en Beatbox, certains points relevés lors des analyses seront présentés brièvement mais non détaillés. Nous détaillerons plutôt les différences d'aspect temporelles entre un type de consonne et le son percussif associé sauf mention contraire (variante).

Le type de voyelle produite avant ou après la consonne en parole n'a pas d'impact significatif sur nos données ( $p$ value=0.50), de même pour l'effet précédent ou suivant en Beatbox ( $p$ =0.25). Nous avons donc exclu ces variables.

En dehors du humming (sons percussifs associés à une mélodie produite par voisement) nous n'avons pas d'effet du type de séquence (continuum, enchaînement de percussives, effets type) sur nos résultats. Les effets produits en humming n'ont donc pas été analysés pour ce travail. Nous n'avons inclus que les effets décrits en endoscopie comme « classiques » ou les plus fréquents (kick, hihat, Rimshot, mesh-snare). Nous présentons une variante en figure 43 pour les snares.

### **2.1. Acoustique**

Au niveau acoustique, l'intensité maximale, l'intensité moyenne et la durée de la consonne/effet sont significativement différentes entre chacune des consonnes et son effet correspondant ( $p$ value<0.001, cf figure 32).

Le centre de gravité spectral permet seulement de différencier la paire /t/ -/T/ (2212 vs 3367 Hz,  $p$ =0.02) et le /k/ -/K/ (2182 vs 1542 Hz,  $p$ <0.001).

Le moment spectral de dissymétrie (skewness) est différent pour la paire /p/ -/P/ (1.0870 vs 2.369,  $p$ value<0.01) ainsi que la kurtosis (6.07 vs 10.47,  $p$ <0.001). Nous retrouvons de même pour la paire /k/ -/K/ (skewness deviation : 1.08 vs 1.87,  $p$ value<0.001 ; Kurtosis : 2.59 vs 6.62,  $p$ <0.001).

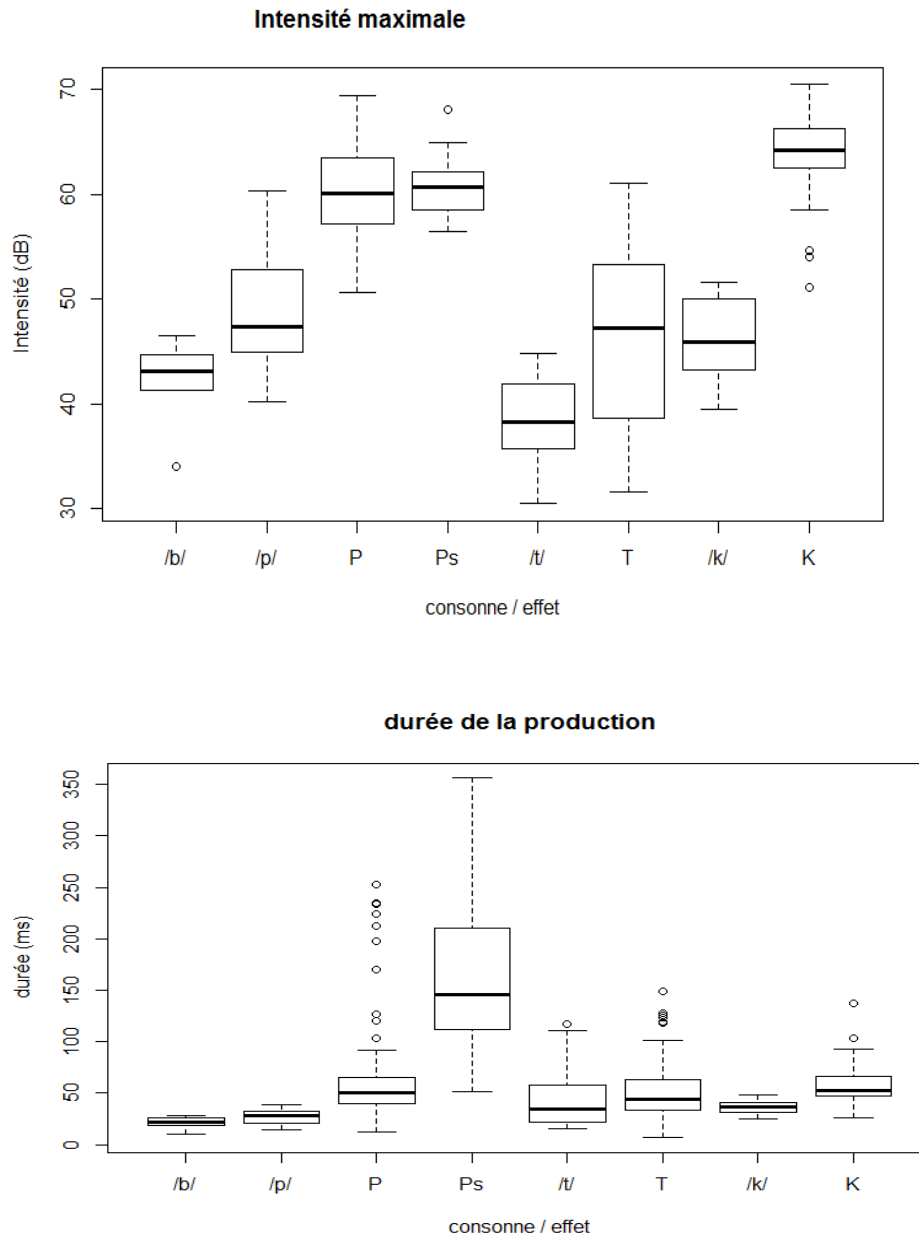


Figure 34 : Boxplots par effet des intensités maximales, des durées de production par consonne/effet : les effets du Beatbox sont notés en majuscule (P = kick, Ps = snare, T= hihat, K = Rimshot).

Il n'y a pas de différence significative des moments spectraux selon la variable « type de tache ».

Etant donné que les moments spectraux sont très corrélés à l'intensité du signal (moyenne et maximale) pour l'effet hihat (figure 35), nous avons un grand nombre de liaisons avec nos variables de PIO, OAF, dp/dT qui semblent expliquées par cette

corrélation, c'est pourquoi nous ne ferons pas d'autres analyses à partir de ces paramètres de spectre.

De plus nous avons réalisé un script de reconnaissance des effets à partir de ces moments, en plus de l'intensité moyenne, maximale et de la durée de la consonne/effet. Le pourcentage d'erreur était de 40% de reconnaissances fausses sur la totalité de la base pour la reconnaissance du type d'effet.

Le délai entre le burst et l'instant d'intensité maximale ne diffère pas significativement entre parole et Beatbox, quel que soit le type de couple consonne-effet.

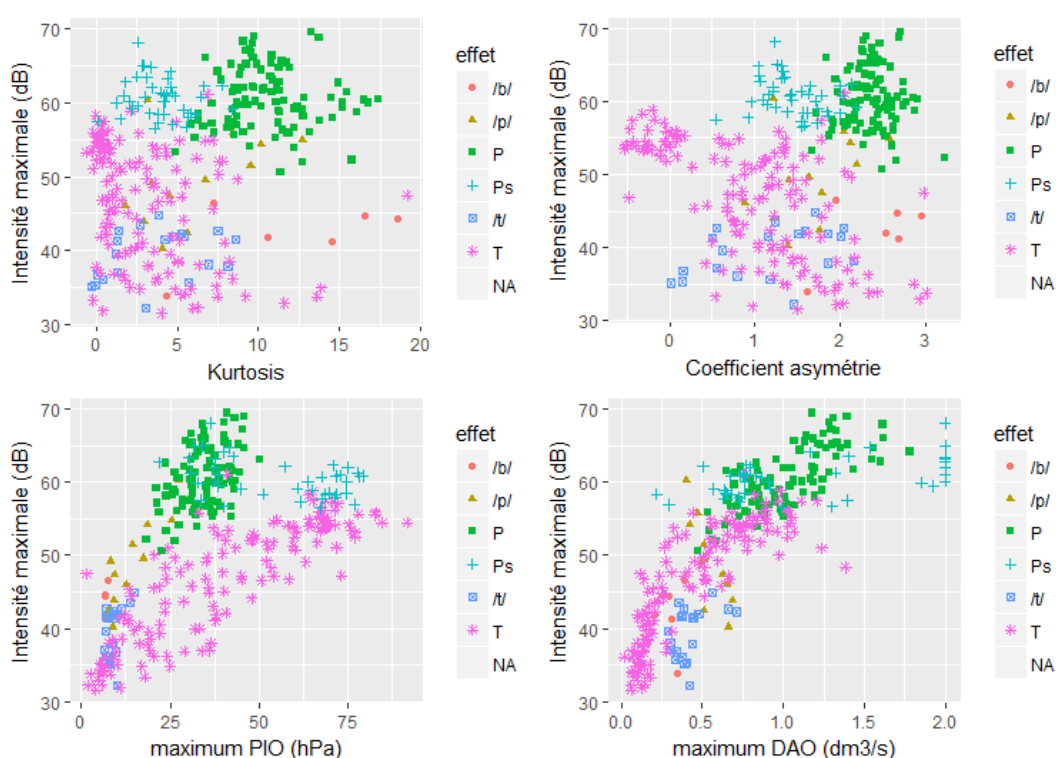


Figure 35 : Intensité maximale en fonction des moments spectraux, des maximums de PIO et DAO par effet.

## 2.2. PIO et débit d'air oral (DAO)

Le maximum de PIO, la pression au temps de relâchement (burst), les coefficients b et d permettent la distinction de chaque paire ( $p < 0.001$ ).

Cependant la pression moyenne d'occlusion ne distingue pas l'effet kick et l'effet snare du /p/ ( $p = 0.32$ ). De même la pression moyenne d'occlusion des effets du Beatbox par rapport à la PIO maximale des consonnes n'est statistiquement significative que pour le /t/-hihat.



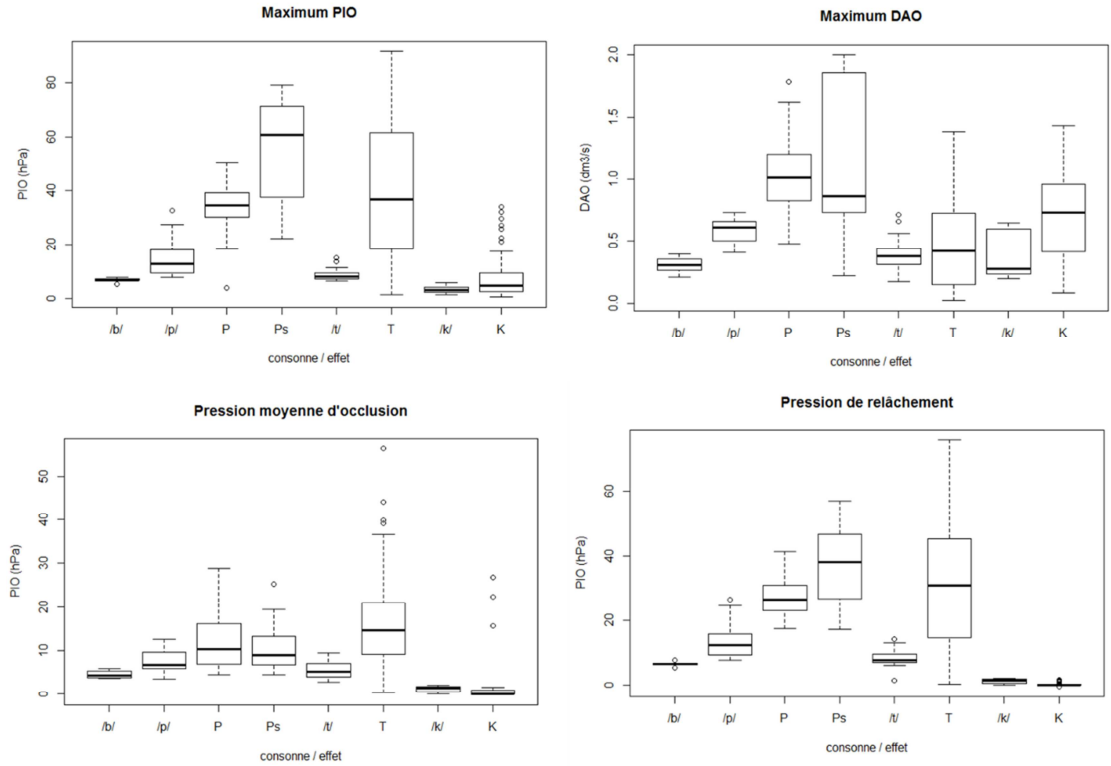


Figure 37 : Boxplots des différentes variables : maximas des PIO et DAO, pression moyenne d'occlusion et au moment du relâchement.

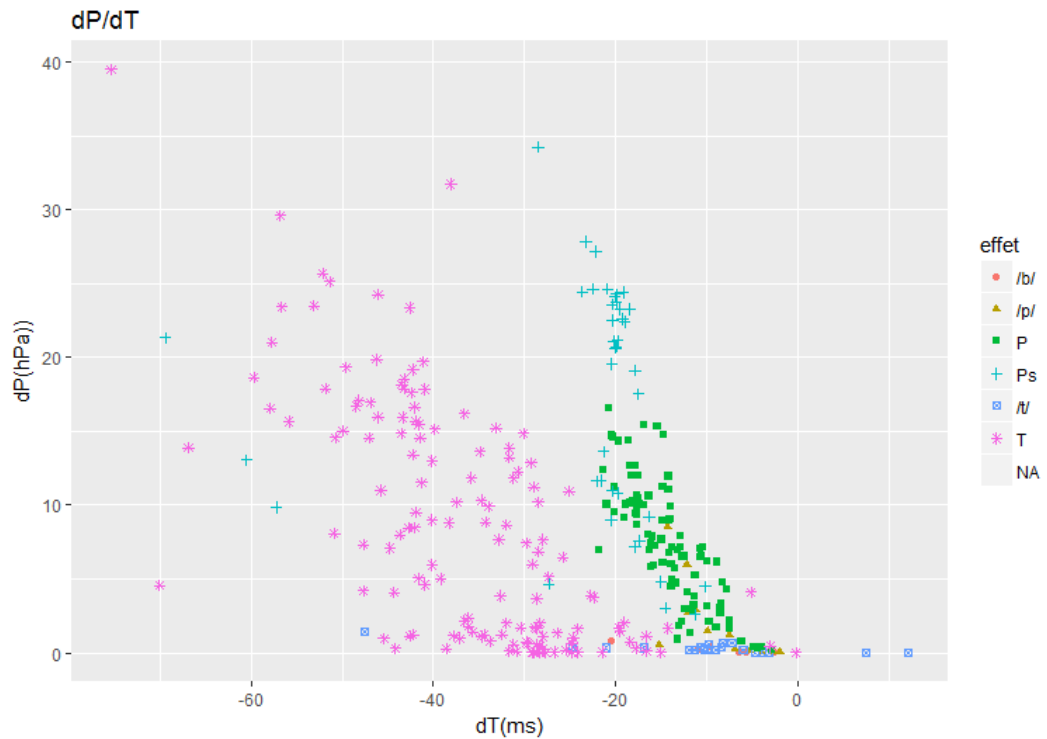


Figure 38 : relations linéaires entre dP et dT pour les effets du Beatbox en comparaison avec les consonnes.

En conclusion, les données de type temporel concernant les durées des phases ne sont pas très discriminantes pour une paire consonne-effet, à la différence des valeurs de pression et de débit d'air et donc des vitesses de celles-ci : il y a des changements d'état plus importants en un temps à peu près similaire. Le Human Beatbox par des effets de type éjectif entraîne une augmentation de la pression intraorale alors que l'occlusion en aval a été relâchée (bilabiale ou apicodentale) à la différence des consonnes de parole pour lesquelles la pression intraorale n'augmente pas après relâchement de l'occlusion. En revanche il se pose une interrogation concernant dT qui est censé être positif pour la parole normale (Koenig & Lucero, 2008; Muller & Brown, 1979).

Les graphes ci-dessous ont été réalisés sur Praat, avec les TextGrids en rapport à partir d'exemples prototypiques. Les données de PIO DAO et DAN ont été normalisées sur le maximum et le minimum de la totalité des effets entre -1 et 1. Les courbes visualisées sont en échelle automatique permettant une visualisation plus facile des courbes mais ne permettant pas leur comparaison visuelle. Sont affichés de haut en bas (channel 1 à 4) : le signal acoustique, la PIO, le DAO, le DAN.

Les légendes des intervalles des figures suivantes sont données ici :

B = intervalle de la consonne, V = voyelle, F= partie fricative

PIO/DAN- C = augmentation, PIO/DAN - D = décroissance,

DAO-C = closing, DAO-O = occlusion, DAO-R = release,

Tmo = temps d'occlusion moyenne, dT= délai début release-maxPIO, OAFM = temps maximum OAF, Tcp = Temps changement de courbure.

DAN-P = phase de plateau

## Consonnes :

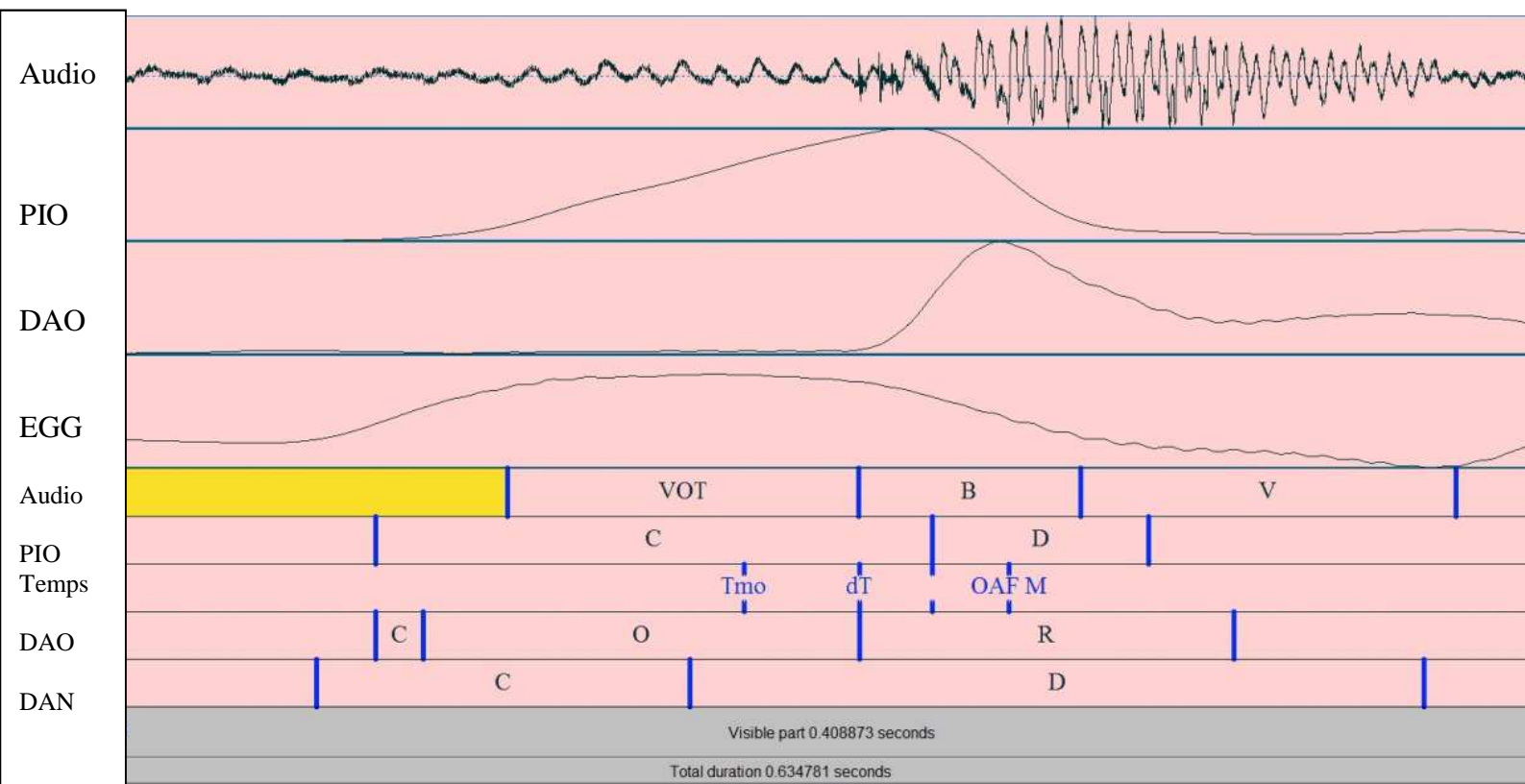


Figure 39 : production de /b/ en parole avec données aérodynamiques en rapport



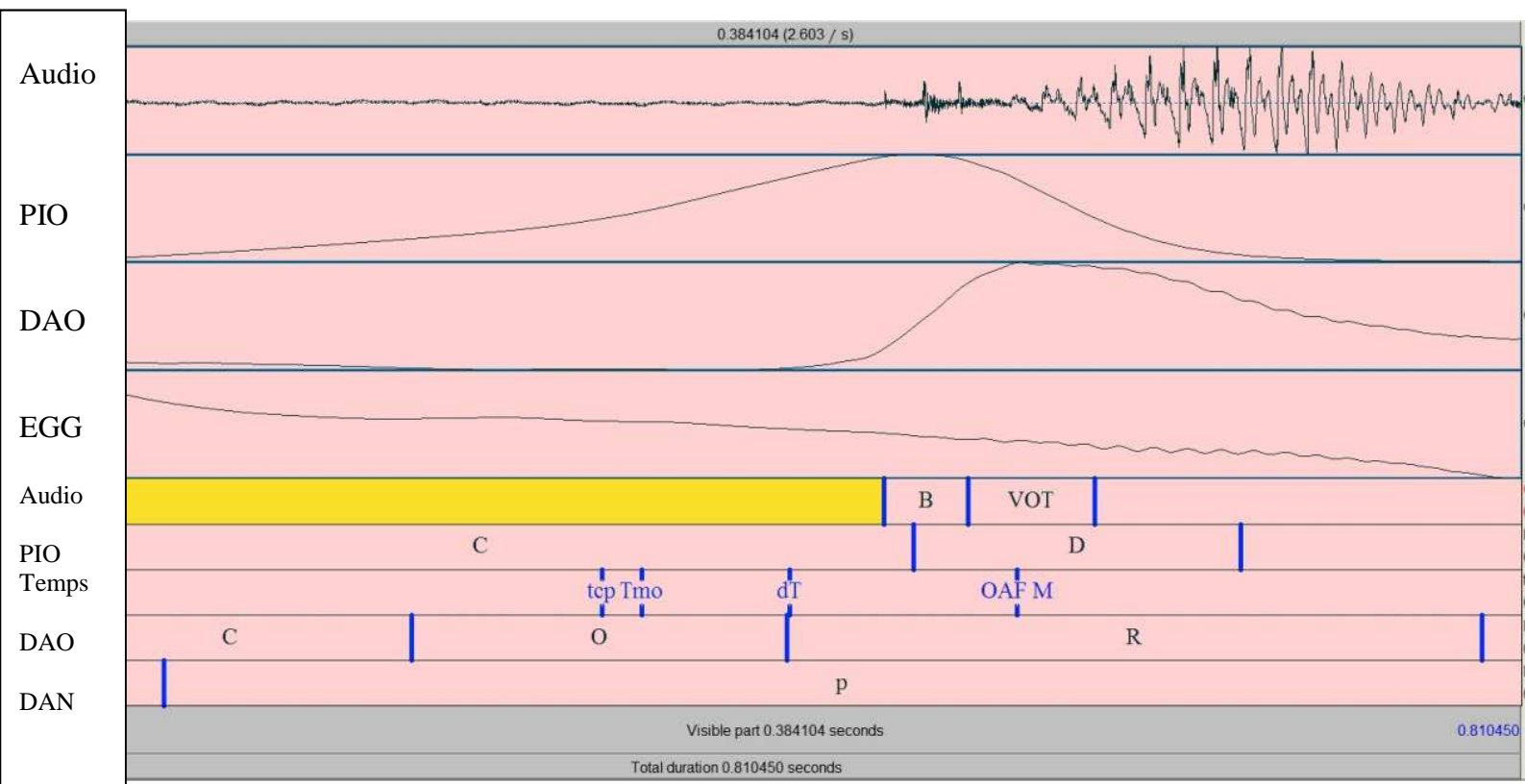


Figure 40 : production d'un /p/ en parole et données aérodynamiques.

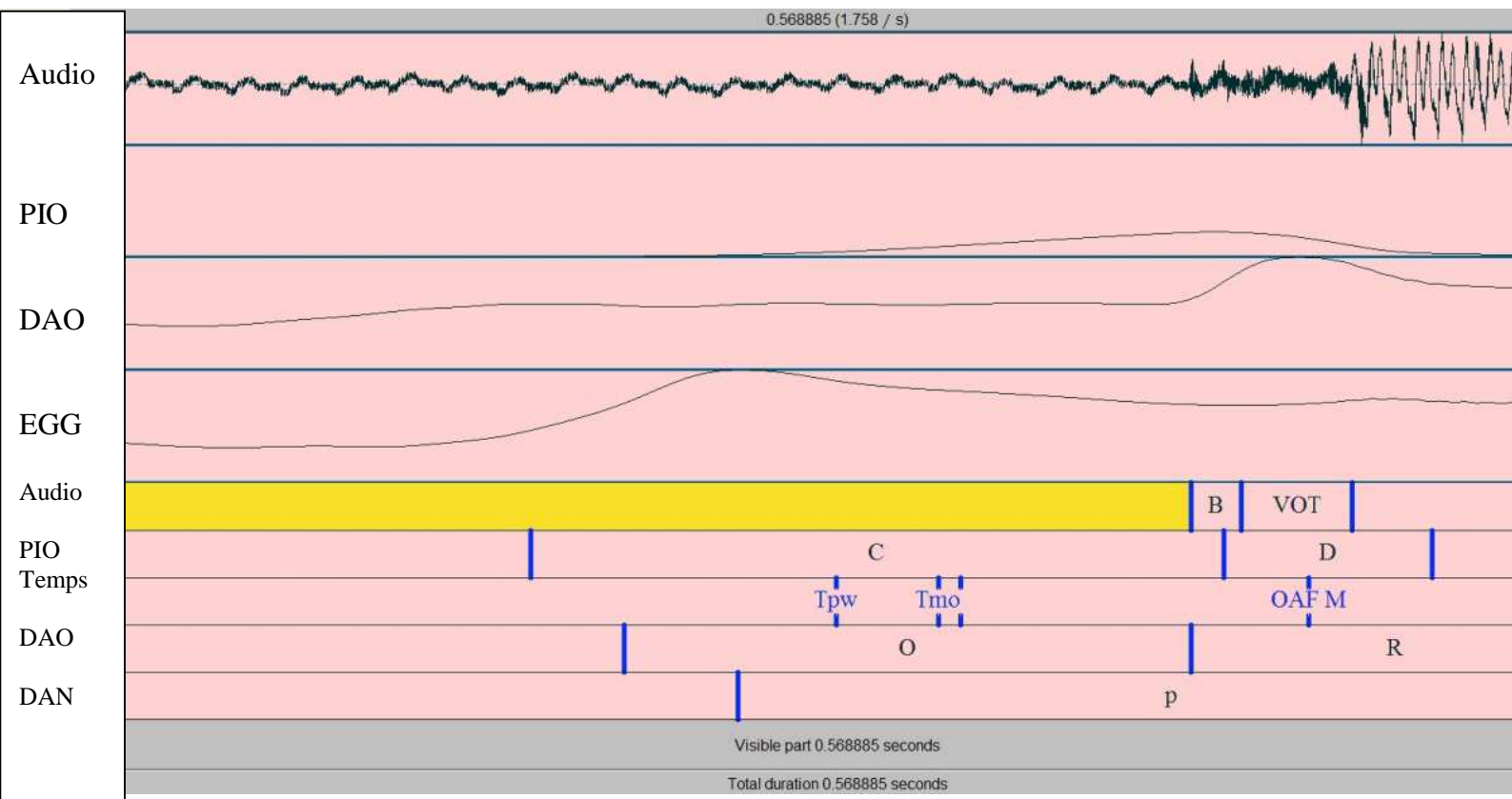


Figure 41 : production d'un /t/ et données aérodynamiques en rapport.

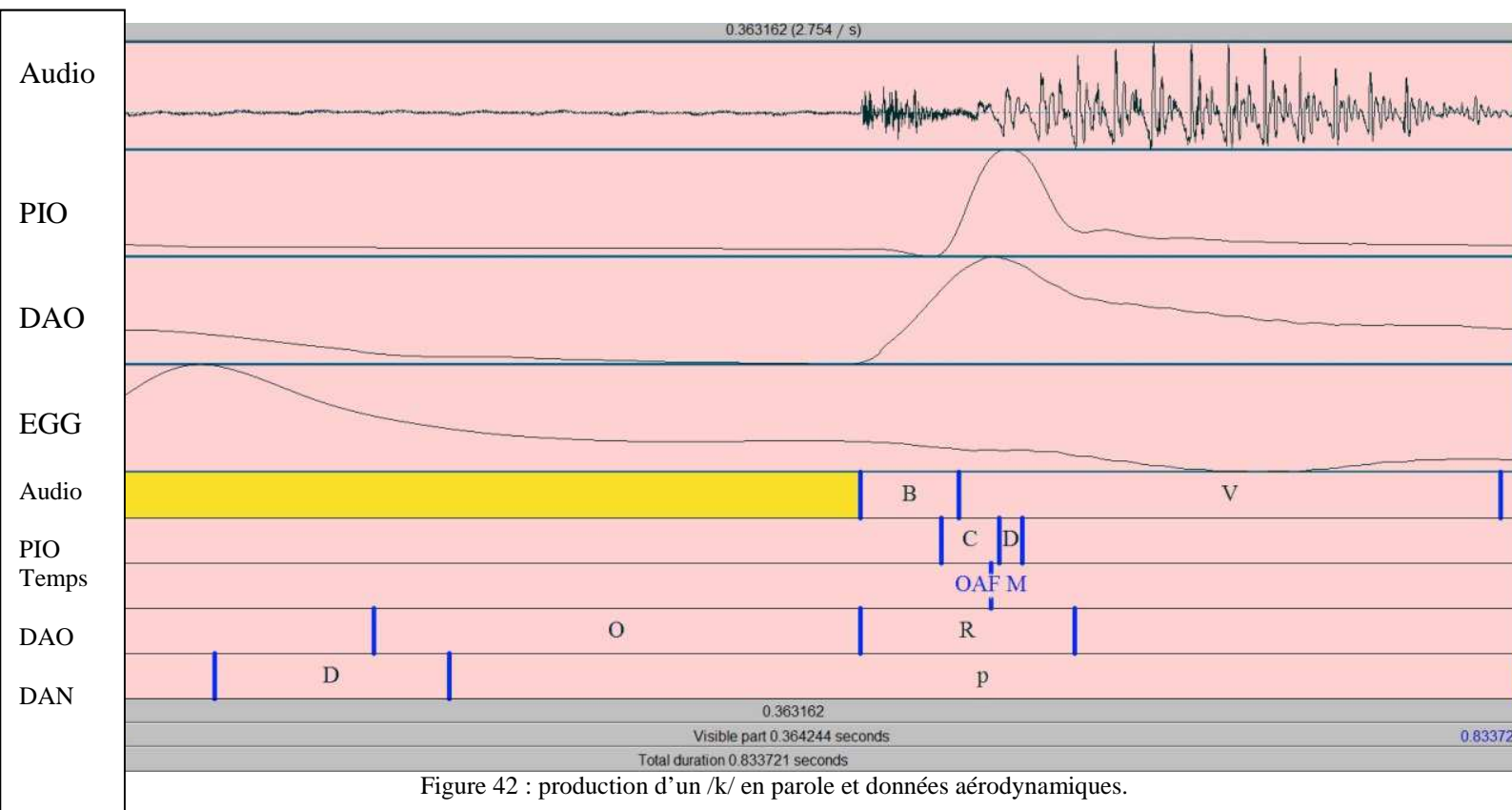


Figure 42 : production d'un /k/ en parole et données aérodynamiques.

**Beatbox :**

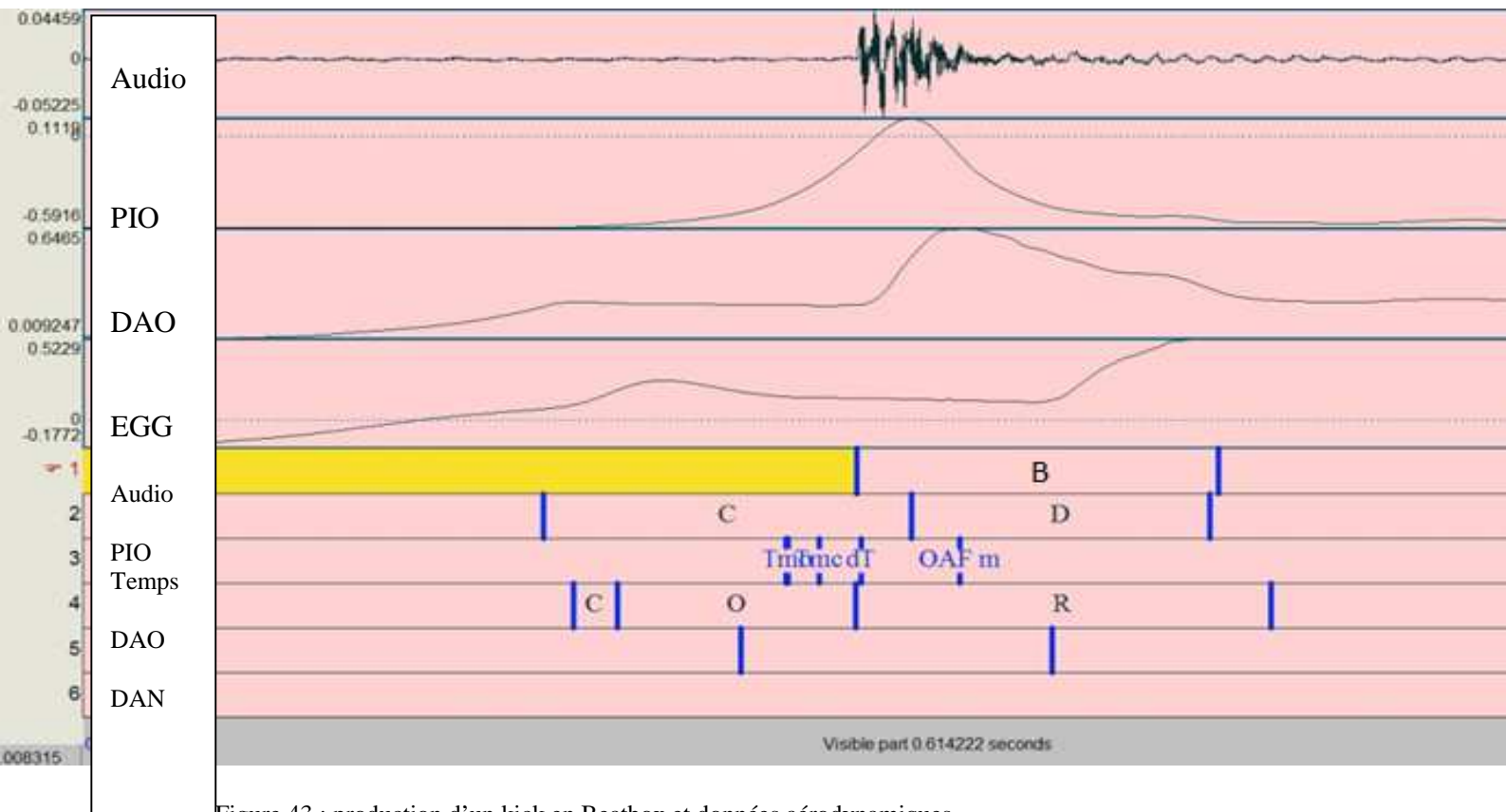


Figure 43 : production d'un kick en Beatbox et données aérodynamiques.

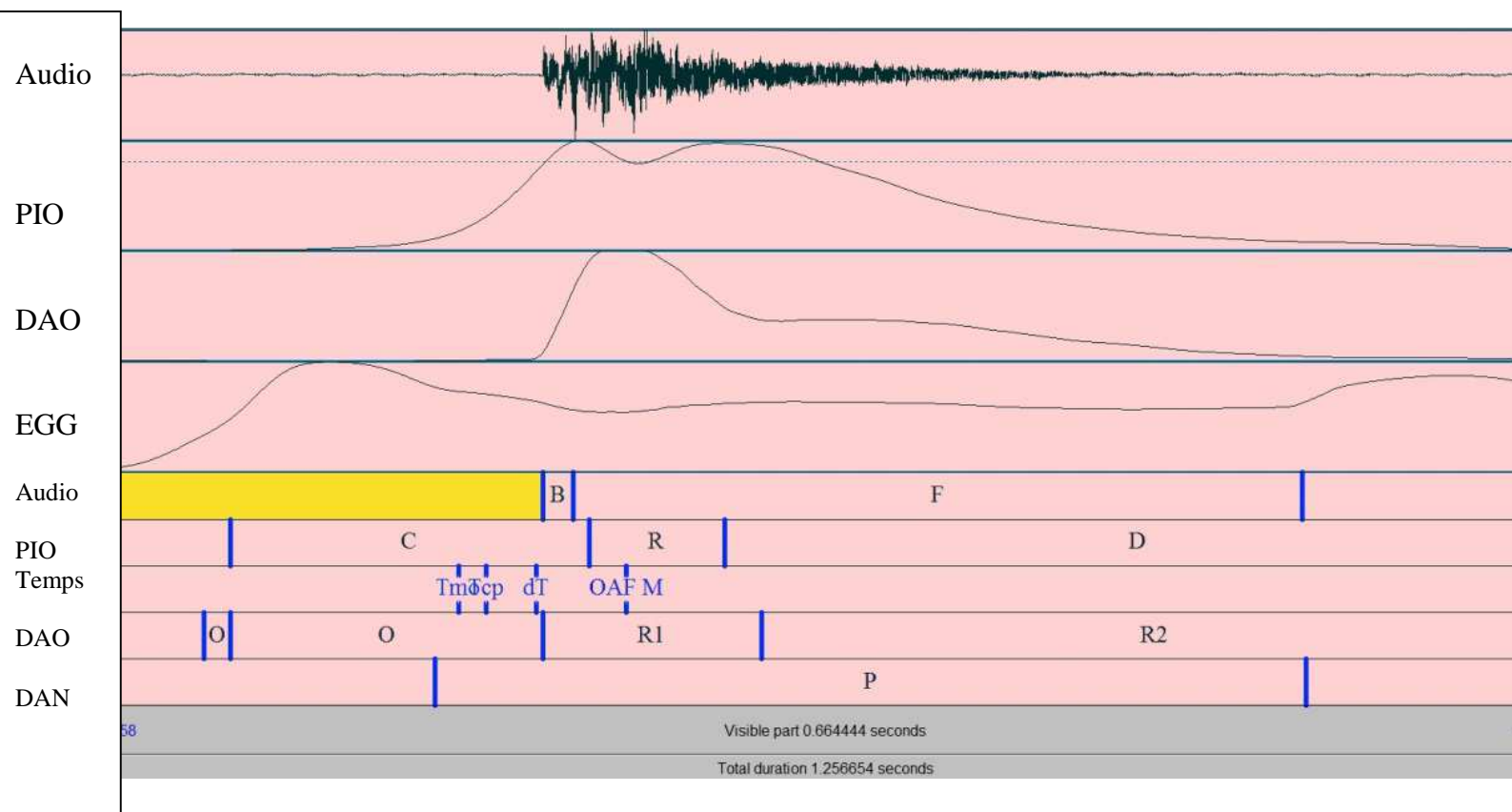


Figure 44 : production d'un snare en Beatbox et données aérodynamiques.

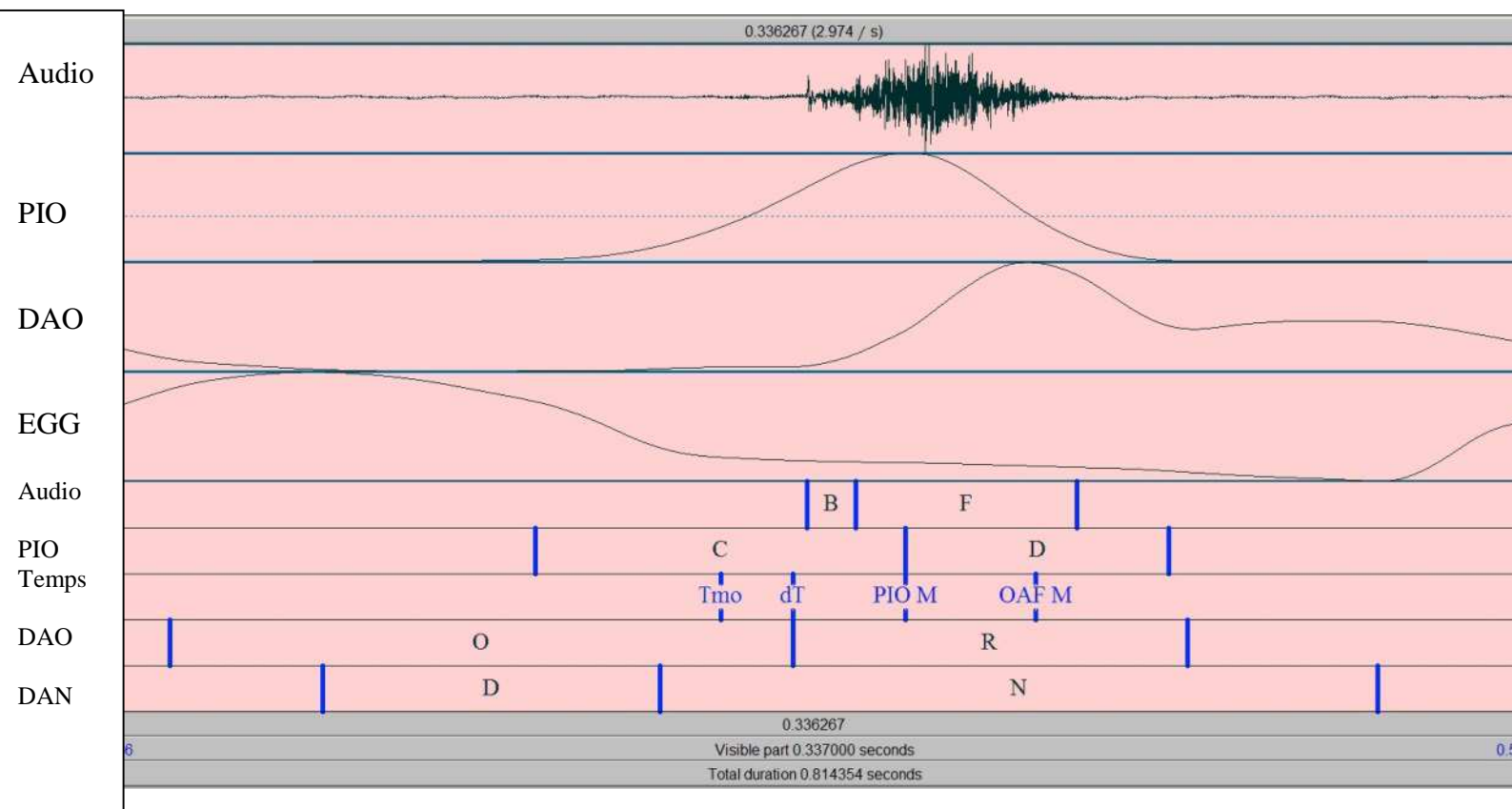


Figure 45 : production d'un hihat et données aérodynamiques.

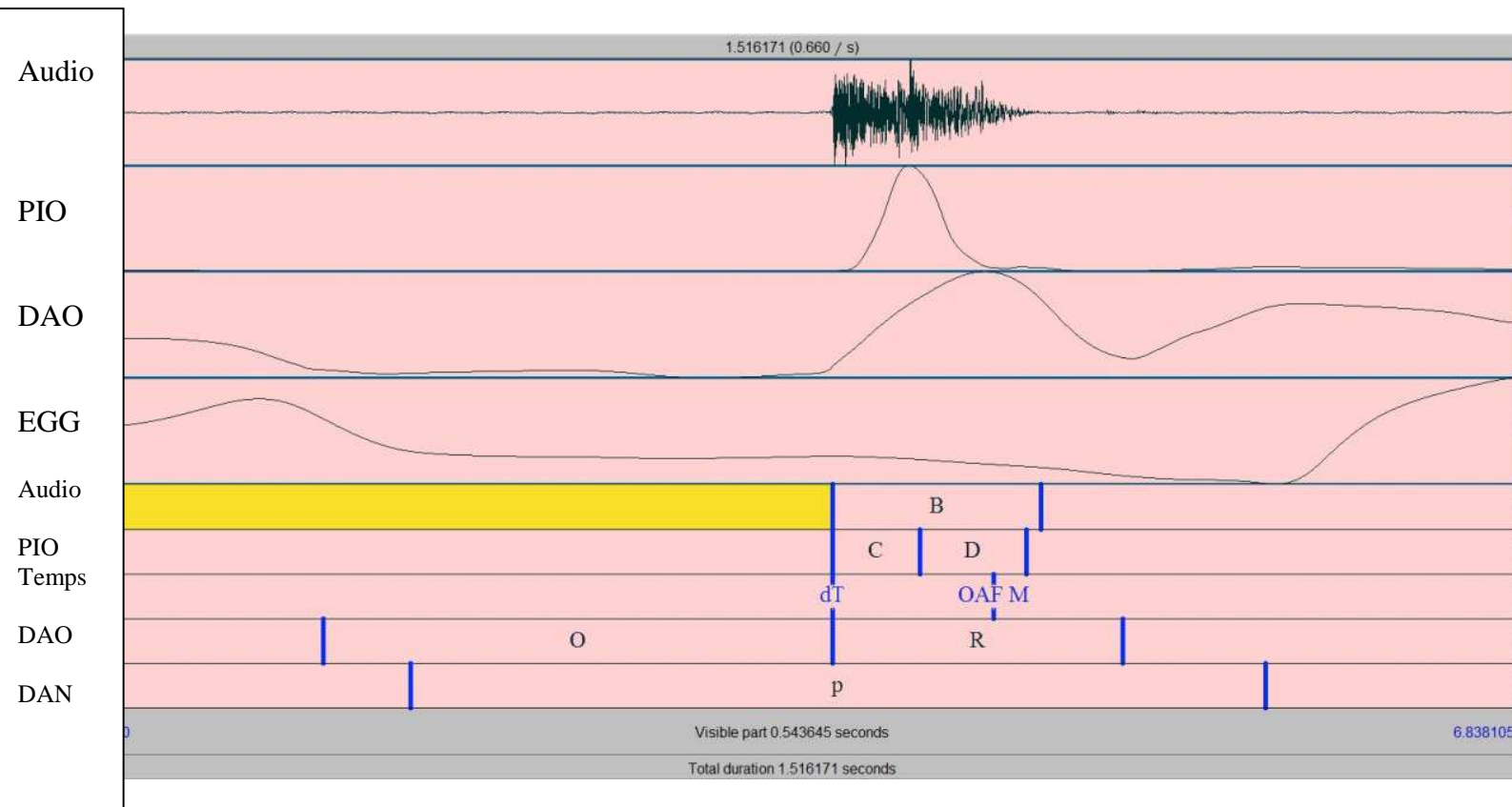


Figure 46 : production d'un Rimshot et données aérodynamiques.

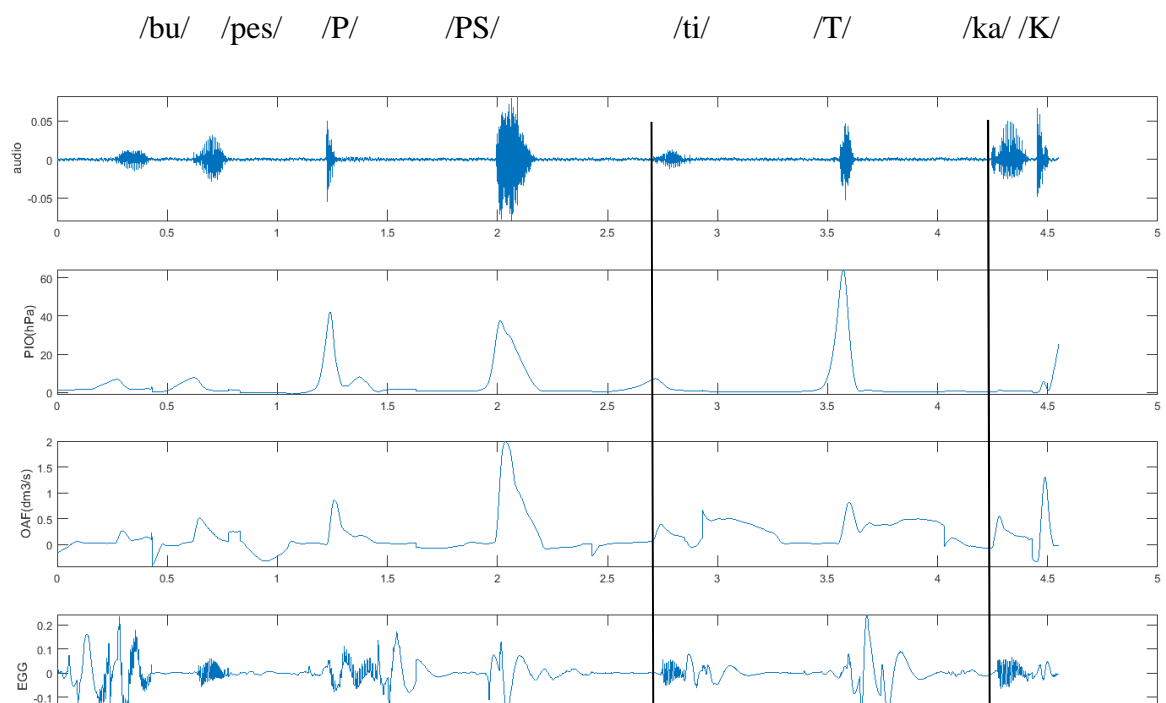


Figure 47 : Consonnes et sons percussif du Beatbox. Les données sont en valeur absolue facilitant la comparaison entre une consonne et son effet correspondant en Beatbox

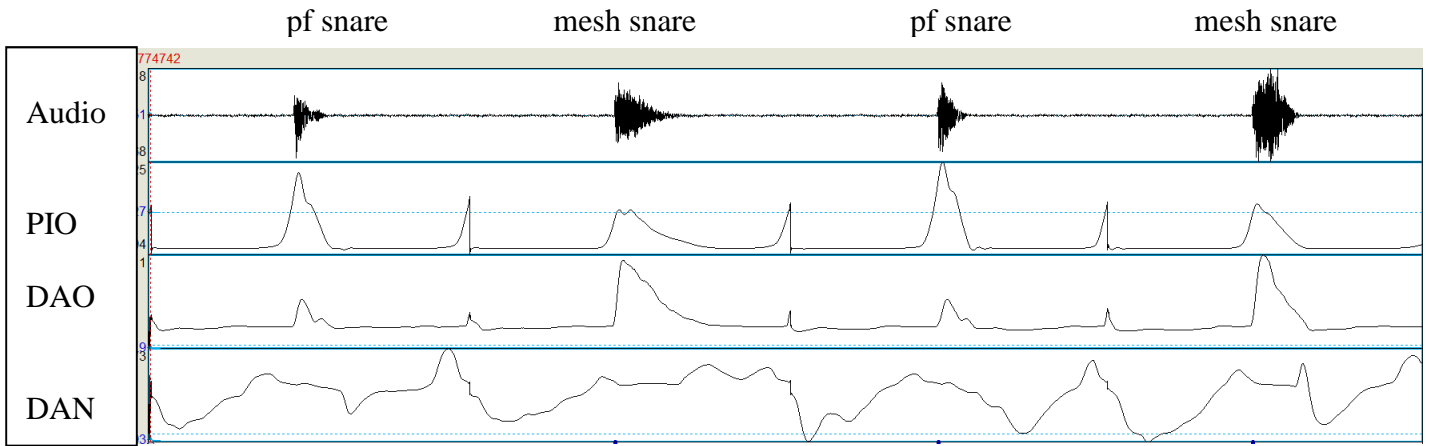


Figure 48 : exemple de variantes de l'effet snare : alternance de /pf snare/ et/mesh snare/.

### 2.3. Débit d'air nasal (DAN)

En dehors d'une différence forte de profil et de maximum de DAN entre la production de Beatbox avec et sans humming ( $p < 0.001$ ), il n'a pas été retrouvé de différence significative du DAN entre les consonnes et les effets « non hummés ». Globalement le débit d'air nasal devient quasi nul à partir du début de l'augmentation de la PIO, et forme un plateau pendant la production de la consonne/effet. Nous ne détaillerons donc pas ce point.

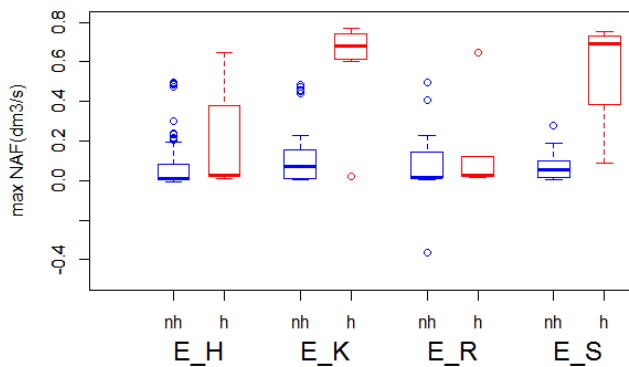


Figure 49 : maximum de débit d'air nasal selon la composante « Humming » pour les quatre effets étudiés du Beatbox. Nh = version classique de l'effet, h = version humming

### 2.4. EGG

Pour la parole, on retrouve le voisement produit après le maximum de PIO pour les consonnes sourdes correspondant à la voyelle : le signal EGG présente

des oscillations brèves autour de la ligne de base et d'amplitude relativement peu importante.

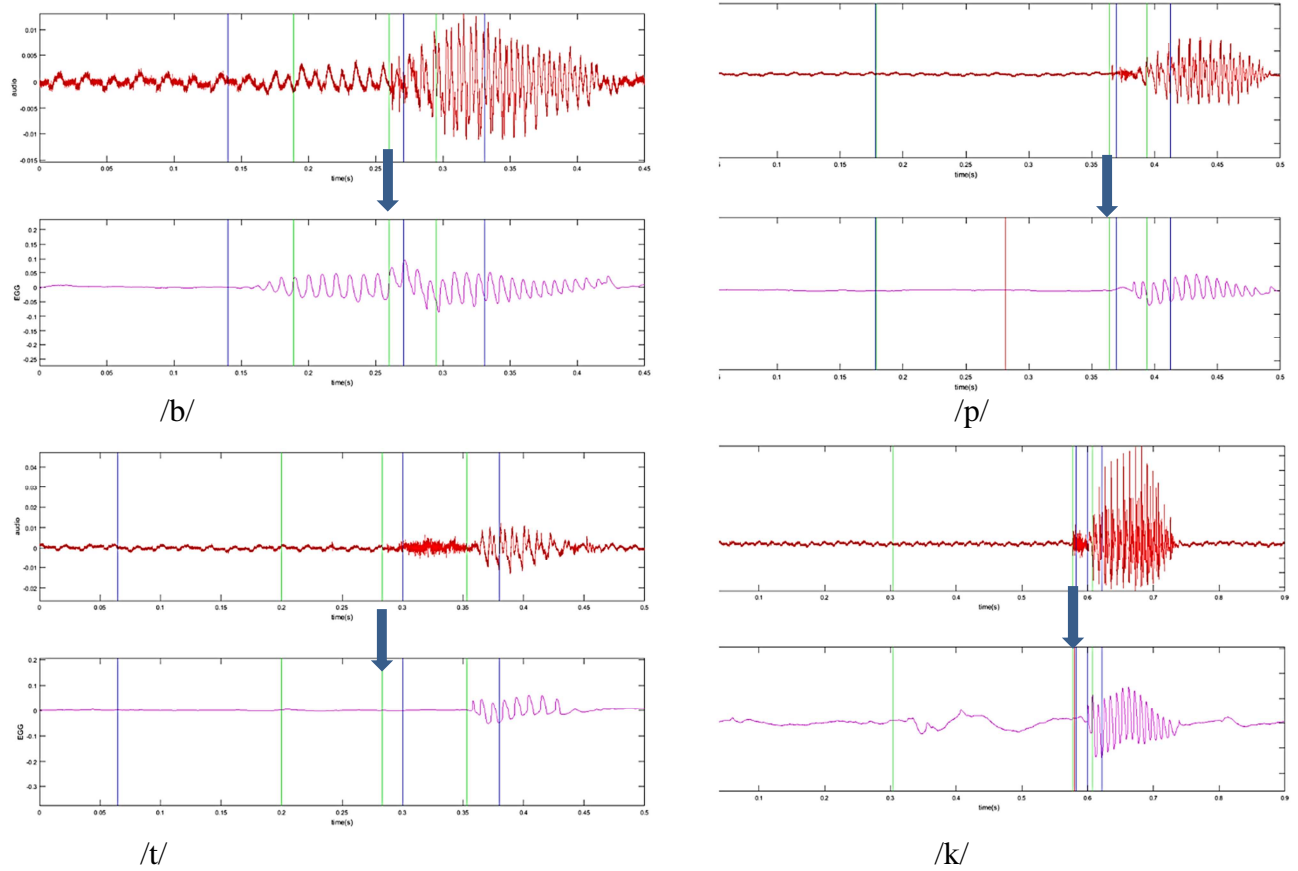


Figure 50 : données acoustiques et EGG des consonnes en parole normale. Les flèches correspondent au Burst

A l'inverse pour les effets du Beatbox, on retrouve un voisement minime voire inexistant en production de percussives simples « classiques ». Mais il y a des variations très amples et lentes de la ligne de base, assez reproductibles que l'on pourrait mettre en rapport avec des contractions de la musculature extrinsèque du larynx plutôt qu'en rapport avec des mouvements d'ascension laryngée, puisque nous avons le même profil de courbe pour les Rimshot alors la hauteur laryngée n'est pas modifiée.

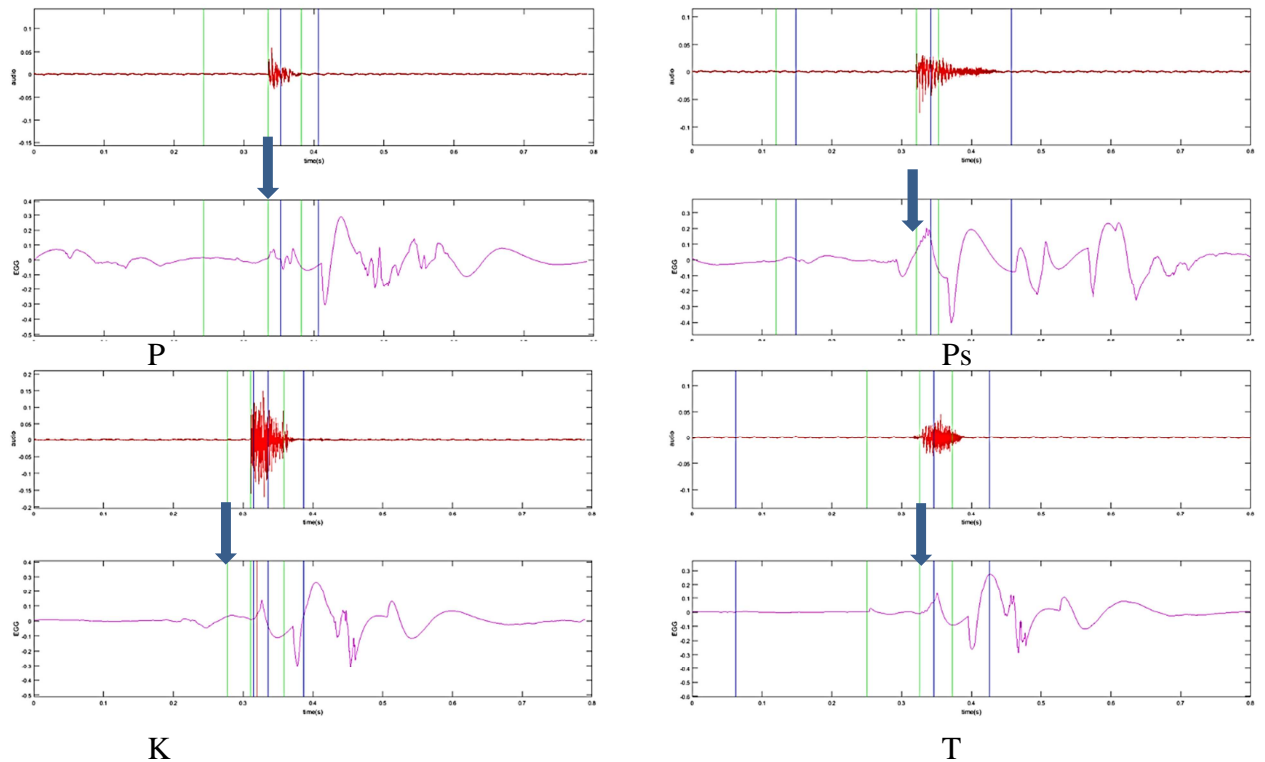


Figure 51 : EGG et données de PIO pour les effets du Beatbox. Les flèches correspondent aux instants du burst.

### 3. *Corrélations entre les données endoscopiques et EVA2*

Etant donné que les microphones utilisés pour les deux sessions d'enregistrement ne sont pas calibrés, nous avons une différence moyenne d'intensité (maximale et moyenne) de 25dB environ ( $pvalue < 0.001$ ). Après normalisation des signaux, l'analyse statistique entre les données récoltées en EVA et en endoscopie ne montre pas de différence significative entre les deux signaux pour l'intensité, les moments spectraux et le délai de maximum d'intensité. Nous pouvons donc bien utiliser nos deux séquences de signal acoustique pour corréler nos données aérodynamiques et endoscopiques.

Nous avons recherché les sons paraissant les plus proches perceptivement, et avons vérifié l'absence de différence significative sur les moments spectraux et en terme d'intensité (en dehors du facteur multiplicatif) et en durée d'effet, afin de reconstruire une « séquence type » artificielle de mise en parallèle des données aérodynamiques et endoscopiques. Nous présentons ces séquences ci-dessous par effet (nous avons laissé les données temporelles telles quelles ce qui explique les décalages). Nous avons d'abord regardé le ratio de durée « D » et « C » hypopharyngé pour les effets qui ont pu être

annotés de cette façon. Nous avons un rapport de 2/3 entre D et C (0.665 IC95%(0.58 0.75)) pour les kicks et les hihats et avons appliqué cette correction sur les effets annotés « D » sur la globalité du mouvement.

Nous avons effectué des comparaisons des moyennes temporelles gestuelles et des temps des phénomènes aérodynamiques par des tests non paramétriques de comparaison multiple étant donné le faible nombre d'échantillons analysés ainsi que la concomitance de certains gestes.

Nous avons traité les mouvements selon leur temporalité par rapport au burst. Nous n'avons pas réussi à obtenir de corrélation entre aérodynamique et états glottiques pour les états peu fréquents d'ouverture ou de fermeture en raison de l'impossibilité de s'assurer de la présence ou non de cette conformation lors de la production pendant la session EVA.

Nous présentons en figure 51 et 52 des exemples de textgrid recomposés à partir des données aérodynamiques et gestuelles mis en rapport avec le signal acoustique.

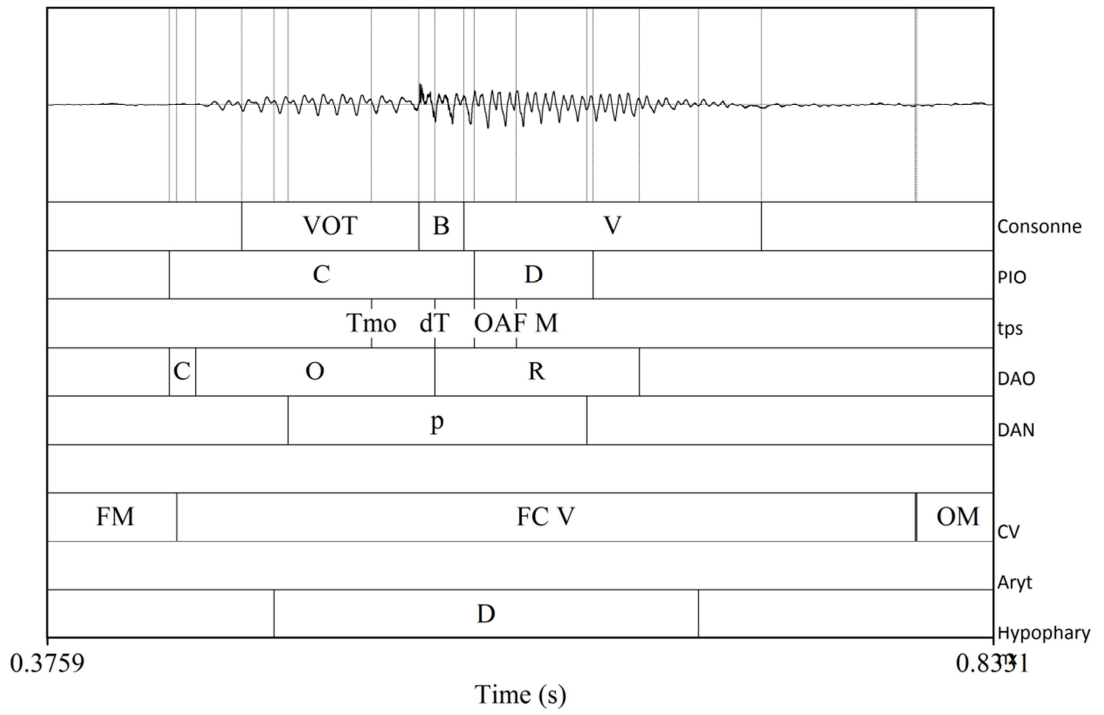
Les données aérodynamiques sont annotées de la même manière qu'en figure 28 à 45.

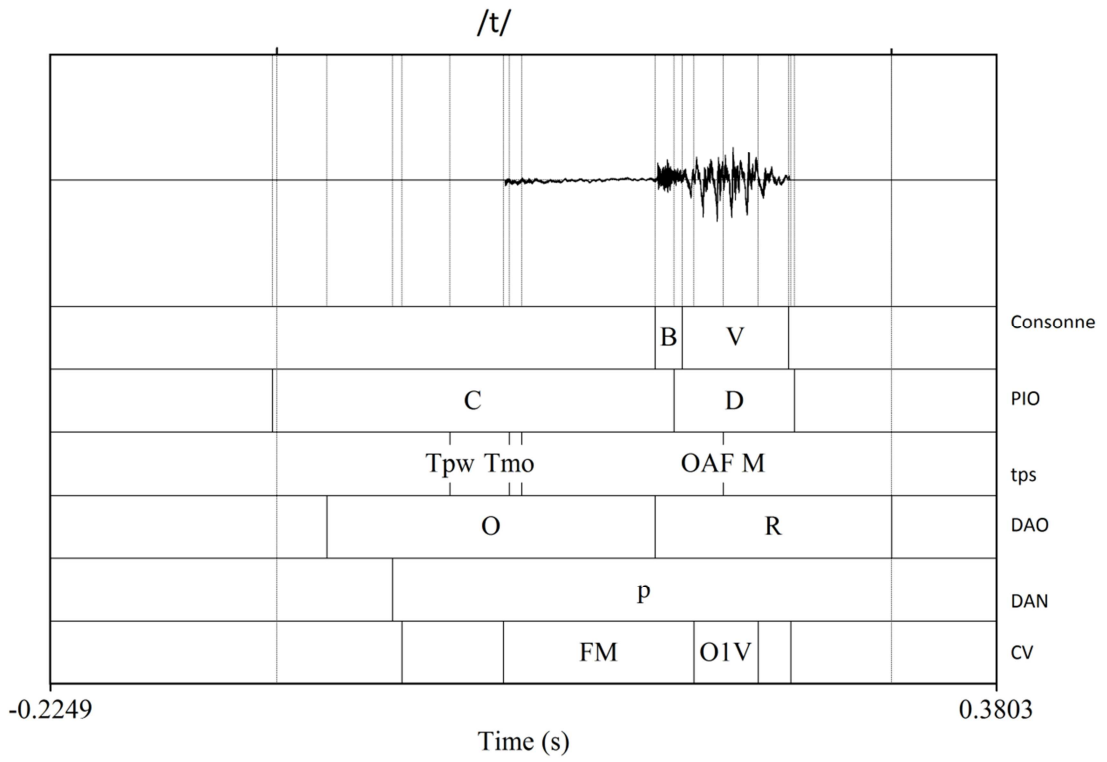
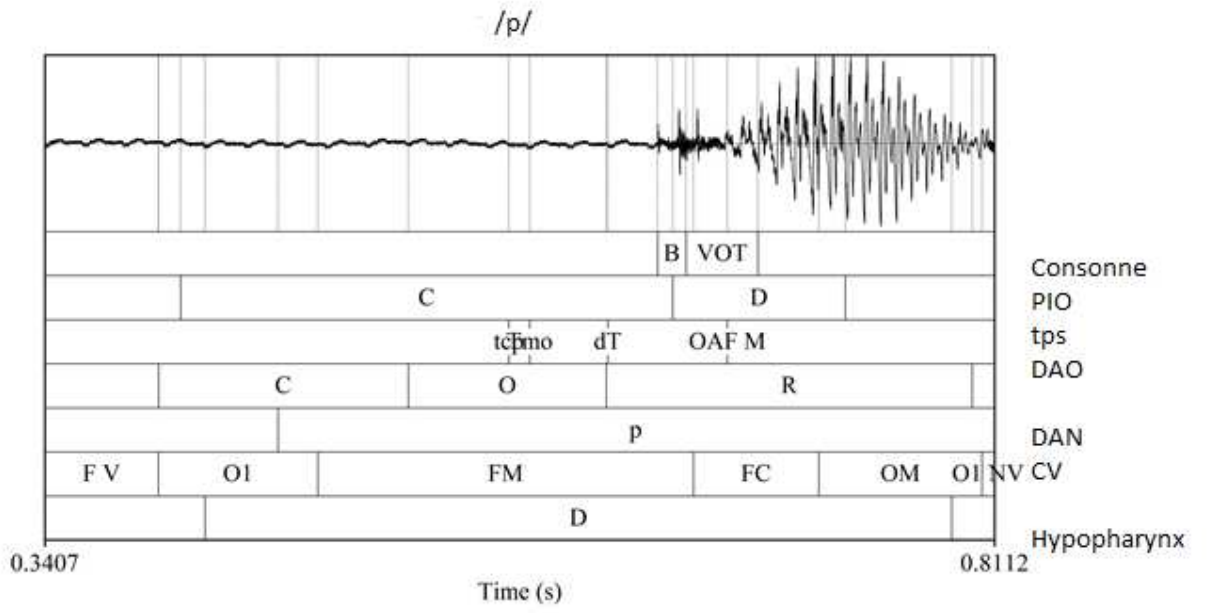
Les données endoscopiques comportent : FM : mouvement de fermeture, OM mouvement d'ouverture, C : contraction/constriction, D : dilatation, R et BAR: bascule arrière, A et BAV: avancée ou bascule avant, O1 à O3 : état d'ouverture de minime à maximal, FI : fermeture incomplète, FC : fermeture complète, larynx-A : ascension.

Pour les **consonnes** :



/b/





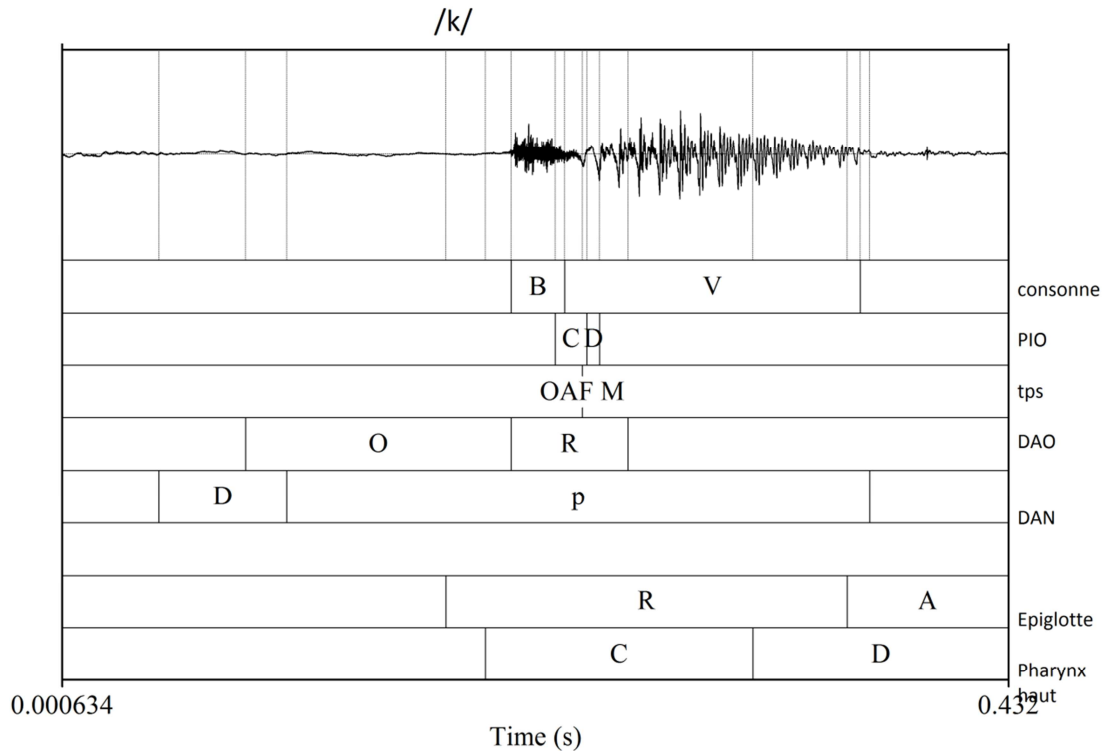


Figure 52 : /b/, /p/, /t/, /k/ en séquences reconstruites : gestes laryngés et données aérodynamiques mises en parallèle

Le début d’augmentation de la PIO est très relié de visu au début du mouvement de fermeture glottique pour tous les effets ( $p < .01$ ).

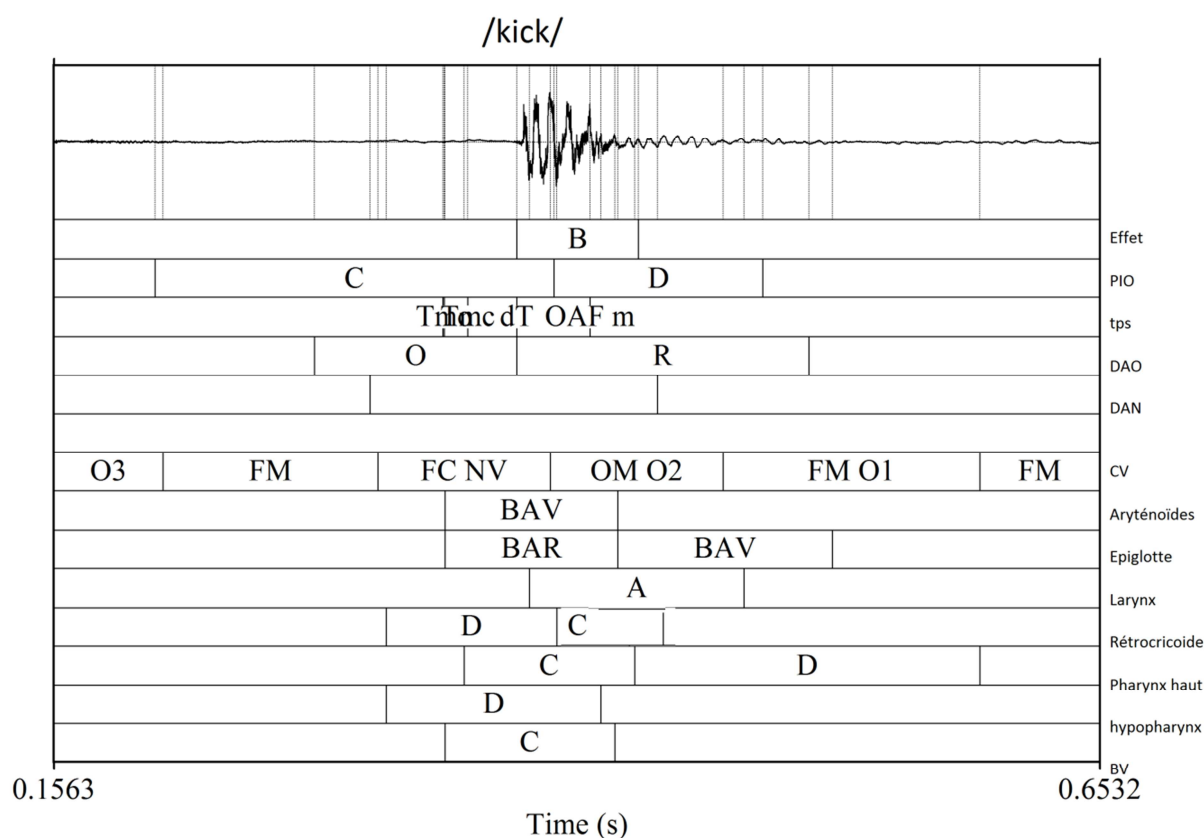
La dilatation hypopharyngée survient à peu près au temps moyen d’occlusion ( $p = 0.034$ ), le pic de PIO précède le temps d’intensité maximal ( $p = 0.04$ ). Pour les /k/ la phase de relâchement/burst survient à la suite de la contraction pharyngée avec un délai de 20ms.

Pour les effets du **Beatbox** kick, nous retrouvons une tendance ( $p$ value=0.054 et  $p = 0.051$ ) à l’égalité des moyennes des temps moyens d’occlusion avec la dilatation pharyngée et la visualisation de l’espace rétrocricoïde, ces temps moyens d’occlusion sont en revanche assez reliés à la contraction du haut pharynx ( $p$ value<0.001). Le délai dT est sensiblement égal au début de l’ascension laryngée ( $p$ value=0.02). Le temps du maximum de DAO est statistiquement lié à l’ouverture glottique ( $p$ value<0.001) et à la fin de contraction du pharynx haut ( $p = 0.035$ ). Le maximum de PIO est atteint en fin de dilatation pharyngée ( $p$ value =0.02) et en fin de visualisation de l’espace rétrocricoïdien ( $p = 0.04$ ).

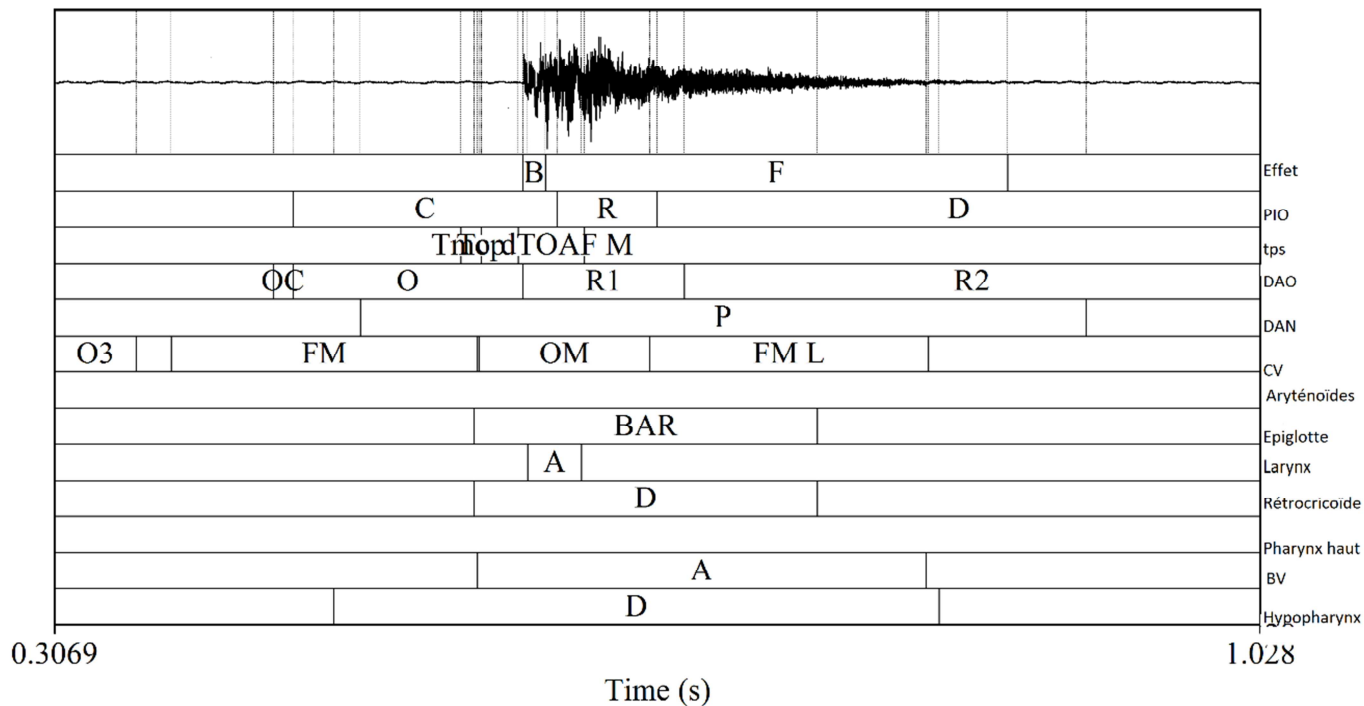
Les snares nous retrouvons : l'ascension laryngée contemporaine du premier pic de PIO et du pic d'intensité ( $p=0.01$ ), nous ne retrouvons aucune donnée susceptible d'expliquer le deuxième pic, témoignant d'un phénomène probable de double articulation par un obstacle buccolingual pour la réalisation de la partie fricative de l'effet. La fermeture lente glottique semble corrélée avec la diminution lente de la PIO mais il faudrait des mesures de type kymographiques pour confirmer cette évolution. Les plis vestibulaires se contractent au moment du temps de pression moyenne d'occlusion ( $p=0.034$ ) en même temps que la dilatation hypopharyngée. La fin de la dilatation de l'hypopharynx se situe en fin de consonne après retour à la normale de la PIO et du DAO.

Les hihats : la dilatation hypopharyngée survient dès le début du temps d'occlusion, avec la visualisation de l'espace rétrocricoïdien intervenant aux environs du temps de pression moyenne d'occlusion ( $p<0.05$ ), l'ascension laryngée se fait au moment du maximum d'intensité ( $p<0.01$ ) qui est statistiquement le même que le temps de maximum de PIO ( $p<0.03$ ). La phase de relâchement intervient au moment de l'ouverture glottique ( $p=0.045$ ). Le retour à la normale hypopharyngé débute en fin de consonne au moment de la fin du pic de PIO ( $p<0.01$ ).

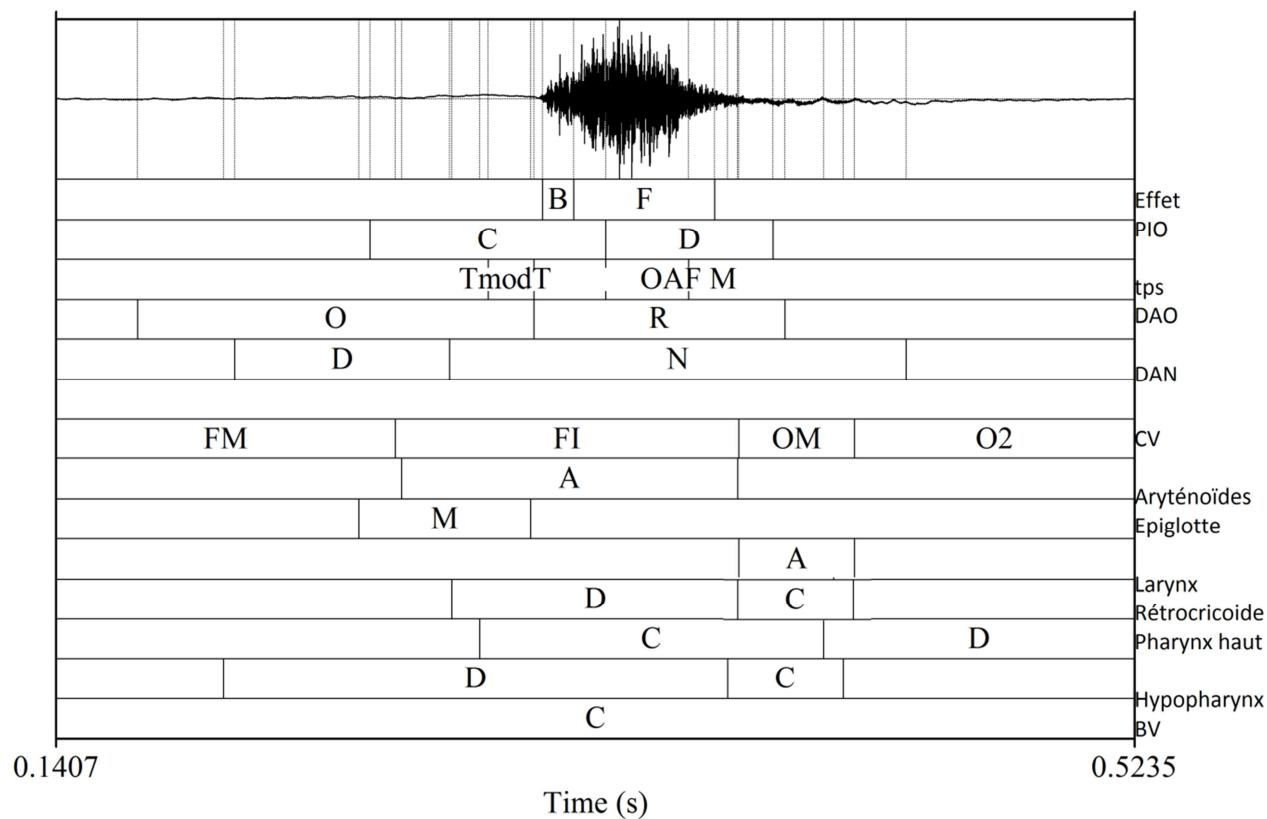
Pour les Rimshot, nous avons seulement trouvé que la fin de la contraction pharyngée haute survient en même temps que le temps du maximum de DAO. Le pic d'intensité a une tendance à être similaire avec le temps de fin de recul de la base de langue. ( $p=0.1$ ).



/Snare/



/hihat/



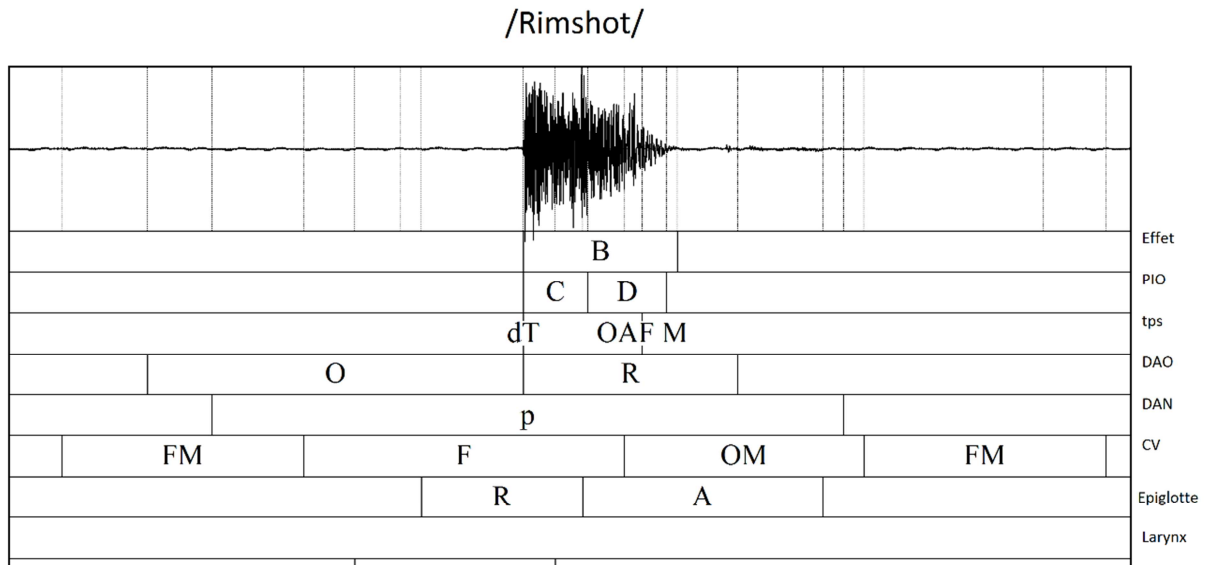


Figure 53 : Relations temporelles entre gestuelle laryngée et données aérodynamiques pour les effets du Beatbox

0.4392

Time (s)

1.01

#### 4. *Echographie*

Etant donné les difficultés pour objectiver l'ascension laryngée, nous avons essayé de la mesurer plus précisément grâce à l'échographie laryngée. Malheureusement comme nous l'avons signalé les données sont difficilement exploitables en raison du délai initial et des mouvements de la sonde et nous n'avons pas encore pu trouver de moyens de mieux les exploiter notamment de façon quantitative (sur le plan vertical).

Nous présentons quelques images échographiques du larynx du sujet 1 :

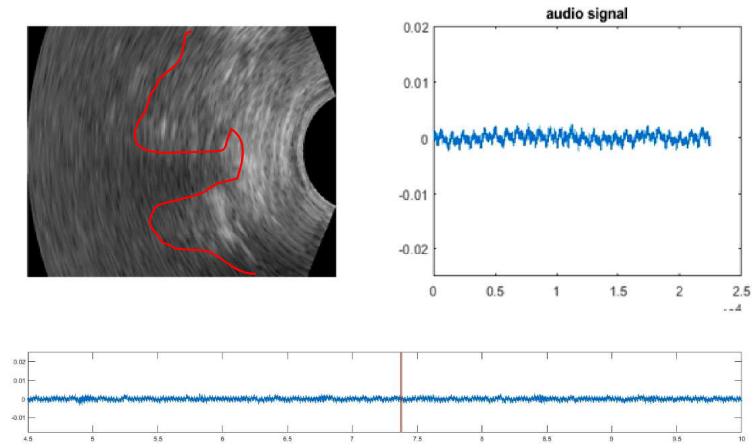
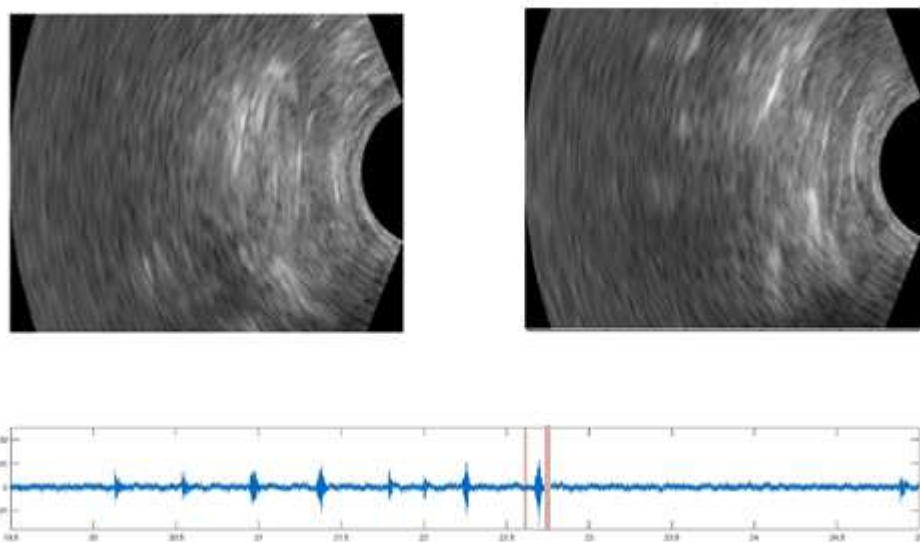


Figure 54 : Exemple de visualisation des données échographiques à partir de notre script : en haut à gauche une image échographique de profil du larynx de profil au repos, on voit aisément la bande ventriculaire et le pli vocal en dessous (profil en rouge). Le cadran en haut à droite montre le détail du signal audiométrique correspondant au temps de l'image, et en bas nous avons la position de l'image dans le signal audio.

Nous avons essayé de mesurer la différence de hauteur du larynx sur le peu de données semblant en rapport avec le signal audio. Il semble visuellement que nous avons une durée d'ascension laryngée mieux objectivée à l'échographie qu'en vidéoendoscopie bien que non statistiquement démontrée (moyenne de durée de l'ascension sur la totalité des effets : 120ms vs 50ms  $p=0,01$ ).



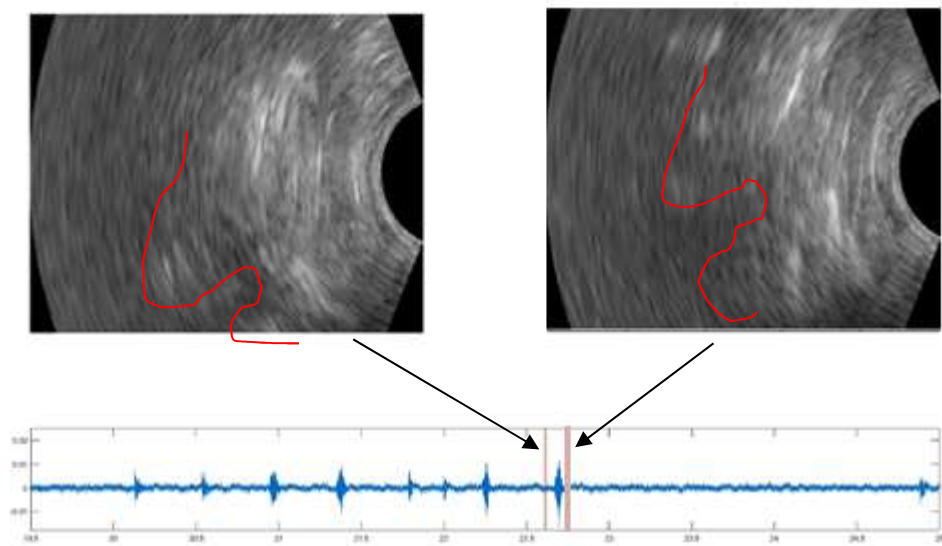


Figure 55 : images échographiques correspondantes au temps notés par les barres verticales sur le signal audiométrique. Les traits de visualisation des structures laryngées montrent une ascension laryngée pendant la production d'un « kick » à la différence des /p/.



## **Partie 3**

-

### **Discussion et mises en perspective**

## Chapitre 6. Discussion

### 1. Discussion

Nous notons dans ce travail une grande différence entre la parole et le Beatbox, déjà retrouvé dans des travaux précédents (de Torcy et Clouet, 2010 ; de Torcy et al., 2014): les patterns laryngés du Beatbox sont beaucoup plus complexes que ceux des consonnes de parole, avec la mise en jeu de structures normalement non utilisées (plis vestibulaires, hypercontraction pharyngée, états d'ouverture glottique inexistant dans la langue française). L'utilisation de la quasi-totalité de ces structures de manière très synchronisée pour l'atteinte de la cible acoustique inexistante dans la langue maternelle des sujets nécessite un apprentissage par ceux-ci. Nos données endoscopiques montrent des conformations pharyngolaryngées extrêmes avec des constriction des différents sphincters laryngés épilaryngés et pharyngés en accord avec les travaux endoscopiques de Torcy et Clouet (2010), de Saphavee et al. (2014) et les données IRM de Proctor et al. (2013) et Patil et al. (2017), notamment pour les effets de kick, snare et rimshot.

Le beatboxeur a l'habitude de s'entraîner ou de mémoriser des séquences rythmiques à partir de sons de sa langue, expliquant l'approche phonétique de plusieurs études du Beatbox. Il en modifie l'articulation, qu'elle soit laryngée ou supralaryngée (Paroni et al. 2017), afin d'obtenir une meilleure imitation de la percussion. Il ne nous semble pas y avoir de réelle continuité entre parole et Beatbox à partir de notre corpus, tant d'un point de vue acoustique, que d'un point de vue articulaire. Cette « disruption » acoustique et gestuelle semble en phase avec la théorie quantique de Stevens sur la présence d'états stables laryngés (« quantum larynx ») (Moisik, 2014; Moisik & Gick, 2017). De plus nous avons pu mettre en évidence des états stables laryngés pour lesquels nous suspectons une participation de la musculature extrinsèque qui serait simulable numériquement.

La sollicitation des espaces hypopharyngé et rétrocricoïde est retrouvée systématiquement dans les effets du Beatbox par rapport à la parole. Ces constatations sont en accord avec Torcy et Clouet (2010) qui relie ce mouvement à la fermeture glottique et retrouvent de la même façon une réelle dissociation entre pharynx musculaire et mouvements laryngés. Les sinus piriformes se dilatent au moment de l'augmentation de la PIO et retournent à un état non dilaté après le relâchement avec un temps de latence plus long pour les effets du Beatbox que pour les consonnes bilabiales. Nous suspectons ainsi

que les sinus piriformes se comportent comme une membrane élastique ou un ballon de baudruche en suivant la loi de Laplace : peu dilaté pour des pressions intraorales « normales » lors de la production de consonnes occlusives, au-dessus d'un seuil qui reste à déterminer, lorsque la pression augmente fortement, il se met à gonfler brusquement et exagérément puis se dégonfle doucement avec une certaine latence notamment pour les parties fricatives des hihat et snare allongées en suivant la décroissance douce de la pression intraorale contrôlée par le flux d'air pulmonique. Kitamura (2005) relie la variation du volume hypopharyngé mesurée par IRM pour des productions voisées à des variations interindividuelles dans la zone du formant du chanteur (2.5kHz). Il serait intéressant de déterminer sa qualité de résonateur dans la production de consonnes et du Beatbox en fonction de la variation de son volume.

L'espace rétrocricoïde n'est jamais visible en production de consonnes dans la parole normale, même en voix forte. Le mouvement d'ouverture de cette zone est rapide et apparaît plus particulièrement coordonné à l'ascension laryngée. Nous pouvons formuler l'hypothèse d'une contribution du positionnement des articulateurs (larynx et langue) et de la musculature extrinsèque laryngée pour cette « dilatation ».

Nous retrouvons sur nos données aérodynamiques des caractéristiques de type éjectif pour les sons percussifs. C'est-à-dire que le beatboxeur crée deux points d'occlusion glottique et labial ou apico-dental, et génère une plosion au relâchement due à la compression de l'air sous l'impulsion de l'ascension du larynx. Ces constatations avaient déjà été faites en vidéoendoscopie laryngée par de Torcy et al. (2014) et en IRM par Proctor et al. (2013) et en articulographie électromagnétique par Paroni et al. (2017).

Les données temporelles aérodynamiques de changement d'état ou de phase sont relativement similaires entre consonnes et Beatbox, mais les amplitudes et vitesses de variation sont beaucoup plus importantes en Beatbox, tant pour la PIO que pour le DAO. Les pressions intraorales en Beatbox sont très importantes, et rejoignent quasiment les valeurs de pressions intraorales retrouvée chez les joueurs d'instrument à vent (Fuks & Sundberg, 1999), bien que la durée de ces fortes pressions soient beaucoup plus faibles (durée de l'effet au maximum). Nous retrouvons des valeurs de débit d'air oral similaires à Bourdin et Navion (2013) et des variantes ingressivees telles que décrites par Garrigues (2008). Le rendement vocal important est lié à un débit d'air oral et une pression intra-orale plus importants en Beatbox que pour la parole normale (Bourdin, 2013) mais nous

ne retrouvons pas réellement d'effet du débit ou de la pression intra-orale sur les moments spectraux (Paroni, 2016).

Comme nous l'avons vu, une grande partie des phénomènes aérodynamiques peuvent être expliqués par le déplacement des structures pharyngolaryngées. Nous pouvons mettre en corrélation les gestes laryngés avec des variations aérodynamiques au niveau pharyngolaryngé (PIO) mais aussi avec le débit d'air oral. La fermeture glottique et le comportement de piston laryngé semble particulièrement relié à l'augmentation rapide et importante de la PIO, de même nous associons l'augmentation de la PIO se poursuivant après la phase de relâchement de l'occlusion avec ce mécanisme de piston éjectif laryngé.

De façon moins attendue, nous retrouvons des mouvements de type éjectifs (piston laryngé) à glotte ouverte pour notre premier sujet pour la production de certaines snares. Cela rejoint une observation de Patil et al. (2017) que la caractéristique éjective n'est pas indispensable dans la production d'un effet. Mais Patil et al. avaient aussi montré en IRM que des variations existaient entre les sujets : des beatboxeurs professionnels ont une glotte ouverte ou en état d'ouverture intermédiaire en production d'effets produits normalement à glotte fermée, ce qui pouvait être la résultante d'un meilleur contrôle laryngé ou du flux d'air pulmonaire, permettant la création d'effets stylistiques chez le beatboxeur professionnel.

Les changements de hauteur laryngée s'observent sur des données EGG par Garrigues (2008), bien que nous n'ayons pas fait d'analyse approfondie électroglottographique, nous retrouvons des profils EGG reproductibles avec des variations amples et lentes de la ligne de base du signal EGG pour les rimshots et un pattern ressemblant à celui des snares alors que la hauteur du larynx ne change pas pour les rimshots. De plus amples analyses seraient nécessaires pour tester si les variations EGG des effets seraient liées à l'état pharyngolaryngé dans son ensemble ou à une contamination du signal par les mouvements amples de la langue et de la musculature cervicale.

Nos données ne montrent pas de variation intersujet importante dans la production d'un effet particulier. Les variations d'amplitude/force articulatoire des mouvements laryngés, non mesurée objectivement, semblent corrélées à l'intensité vocale. En revanche, la variation en termes de type de mouvement des structures laryngées génère

une variante acoustique de l'effet initial (distinction par le beatboxeur de l'effet type). Il y a une grande similarité de comportement laryngé entre nos deux sujets pour un effet donné (Clouet et Torcy, 2010).

En tant que producteur expert, le beatboxeur choisit le son qu'il veut imiter de manière consciente lorsqu'on lui demande ou lorsqu'il le décide. Dans les séquences rythmiques accélérées ou libres de notre protocole il a choisi plutôt une certaine forme « typique » à lui, devenue sa référence dans la famille d'effet, probablement celle qui lui demande le moins d'effort ou qu'il est le plus habitué à utiliser, à moins d'être contraint pour une raison de fin d'expiration ou de style musical d'utiliser une variante.

Nous avons par ailleurs noté pour la tâche de parole une certaine difficulté à se détacher du Beatbox pour les prononcer en parole « normale ». Cela pourrait être dû au type rythmé des phrases du corpus ou au fait que ce sont des phrases classiques utilisées pour la mémorisation de séquences rythmiques par nos sujets. Il est possible que le pattern acquis initialement de façon consciente, et répété de multiples fois devienne automatique au niveau cognitif et influe sur la parole : par exemple nous n'expliquons pas vraiment le délai inversé entre le moment du relâchement de l'occlusion et le maximum de PIO en parole phénomène plutôt caractéristique des percussives en Beatbox.

## ***2. Limites et perspectives***

Nous avons rencontré un certain nombre d'obstacles qui peuvent avoir une influence sur nos conclusions.

Tout d'abord, nos données aérodynamiques ne concernent qu'un seul sujet, et ne permettent donc pas de tirer des conclusions générales sur la gestion aérodynamique dans la production de Beatbox.

Il aurait été intéressant d'enregistrer l'EKG pendant les enregistrements endoscopiques afin d'obtenir une meilleure corrélation entre l'analyse des données aérodynamiques et fibroscopiques. De plus sur les données EKG de notre corpus, nous avons seulement réalisé une comparaison rapide, alors qu'il y a la présence d'un certain nombre de variations temporelles semblant corrélées à des mouvements retrouvés en vidéoendoscopie.

La fréquence d'échantillonnage des prises d'images endoscopiques n'étant pas tout à fait stable pour notre premier sujet, la variation de nos données temporelles, bien que diminuée par le rééchantillonnage de la fonction Bootstrap utilisée pour nos analyses, diminue certainement la puissance de nos résultats. L'enregistrement vidéoendoscopique du deuxième sujet avec un framerate de 25 images/seconde ne nous paraît pas suffisant pour une bonne évaluation temporelle de la gestuelle laryngée puisque une seule image équivaut à peu près à la durée d'un effet de Beatbox. Etant donné la rapidité et la puissance des mouvements pharyngolaryngés en production de Beatbox, ou pour la consonne de parole /k/, certaines structures sont parfois invisibles et l'on ne peut que conjecturer sur leur état. L'utilisation de la vue postérosupérieure en nasofibroscopie améliore cependant la visualisation de ces structures bien que l'on perde de l'information sur le reste des données. Les mouvements très fins, qui ont une probable influence acoustique comme la vibration des replis aryépiglottiques, ne sont pas visualisables en raison des capacités du matériel utilisé, des gestes plus importants tels que l'ascension laryngée sont difficilement caractérisables du fait des mouvements globaux des structures étudiées et de la caméra en l'absence de point de référence fixe. Nous n'avons annoté qu'une petite partie des effets, nous obligeant à utiliser des tests non paramétriques, peu puissants. Pour la suite de ce travail nous allons continuer les annotations, afin d'analyser la gestuelle laryngée par des tests de type séries temporelles.

Nous nous sommes intéressés uniquement au « pharyngo-larynx » en partant du principe que les mouvements très reproductibles en production de Beatbox sont actifs donc probablement acquis de façon consciente au cours de l'entraînement du Beatbox. Mais il semble y avoir aussi une très grande coordination entre les structures cervicales basses que nous avons étudiées, la partie pulmonaire et les structures orales, notamment pour les snares, nous pouvons suspecter que certaines conformations laryngées sont « réactionnelles » ou « passives » au flux d'air sur une certaine durée car non possibles d'un point anatomo-physiologique. Nous ne pouvons que supputer en l'état actuel de nos travaux sur l'influence des sphincters supralaryngés et de la soufflerie sur ces conformations laryngées.

Les paramètres acoustiques étudiés ne semblent pas suffisamment puissants pour la distinction des effets et de leurs variantes, ce qui confirme les analyses des travaux précédents (Paroni, 2016).

Il n'a malheureusement pas pu être fait d'analyse quantitative sur l'influence des vitesses des mouvements des structures, et nous n'avons pas réussi à déterminer des paramètres acoustiques suffisamment robustes pour la discrimination des effets.

Le travail d'annotation vidéo sous la forme que nous avons utilisée est fastidieux et reste imprécis. Nous n'avons pas réussi à mettre en place des algorithmes de tracking (de type optical flow) et de reconnaissance des structures de façon efficace en raison de la variation très faible de couleur des structures (globalement roses). Mais nous pensons que la variation de couleur produite par le niveau de hauteur par rapport à la source de lumière, appliquée à un modèle 3D du larynx (créé à partir de données IRM ou à partir d'images dans 2 plans relativement orthogonaux comme la vue supérieure et postérosupérieure), contraint sur les formes qu'il peut prendre (musculature, limites anatomiques), pourrait permettre une annotation automatique et permettre le suivi des structures et leurs variations absolues et relatives (Schenk, Aichinger, Roesner, & Urschler, 2015). La création d'un tel modèle en trois dimensions aurait un certain nombre d'application tant sur le plan de la recherche en phonétique que sur le plan médical pour l'évaluation fonctionnelle et chirurgical en comparant simulants numériquement le retentissement de différentes techniques sur la voix et la respiration. Le Beatbox par ses gestes pharyngolaryngés très divers, dans des conformations ou des états souvent extrêmes nous semble une très bonne base pour la création d'un tel modèle numérique.

### **3. Conclusion**

Dans ce travail nous avons décrit des effets du Beatbox en endoscopie et en enregistrements aérodynamiques, en rapport avec des consonnes occlusives du français. Les structures pharyngolaryngées impliquées dans la production du Beatbox ne sont pas les mêmes que dans la tâche de parole. Nous avons ainsi mis en évidence la présence de mécanismes éjectifs et de contractions pharyngées particulièrement importants dans le Beatbox puisque ces mécanismes participent à l'augmentation forte de la Pression intraorale qui est reliée à l'intensité acoustique.

Nous avons pu mettre en évidence sur nos données aérodynamiques que les différentes données temporelles ne sont pas particulièrement différentes entre

la parole et le Beatbox à la différence de la forme des courbes et des variations d'amplitude tant sur la pression intraorale que sur le débit d'air oral.



## Bibliographie

- Abdelli-Beruh, N. B. (2004). The stop voicing contrast in French sentences: contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, 61(4), 201-219. <https://doi.org/10.1159/000084158>
- Aron, M., Kerrien, E., & Laprie, Y. (2007). Coupling electromagnetic sensors and ultrasound images for tongue tracking : acquisition set up and preliminary results. 7th International Seminar on Speech Production-ISSP'06.
- Atherton, M. (2018). Rhythm speak : mnemonic, language play or song? International Conference on Music Communication Science (p. 15)
- Bourdin, D., & Navion, A. (2013). Mesure de l'efficacité vocale au sein d'une population de chanteurs de Human Beatbox : Analyse acoustique, aérodynamique et observation. Mémoire d'orthophonie, Paris VI
- Branderud, P., Lundberg, H.-J., Lander, J., & Djamshidpey, H. (1998). X-ray analyses of speech: methodological aspects. *Swedish phonetics conference*
- Brunner, J., Fuchs, S., & Perrier, P. (2005). The influence of the palate shape on articulatory token-to-token variability., *ZAS Papers in linguistics* p. 42:43-66.
- Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, 136(5), 2762-2772.
- Contesse, A. (2015). Vocal Grammatics - Une écriture pour le beatbox. Mémoire de fin d'étude de design, ESED Dijon
- de Torcy, T., Clouet, A. (2010). Le Human Beatbox : étude qualitative acoustique et en vidéo-fibroscopie. Mémoire d'Orthophonie, Paris 6 Faculté Pitié Salpêtrière..
- de Torcy, T., Clouet, A., Pillot-Loiseau, C., Vaissière, J., Brasnu, D., & Crevier-Buchman, L. (2014b). A video-fiberscopic study of laryngopharyngeal behaviour in the

*human beatbox. Logopedics Phoniatrics Vocology*, 39(1), 38-48.

<https://doi.org/10.3109/14015439.2013.784801>

Dejonckere, P. (1985). Le concept oscillo-impédantiel de la vibration laryngée. *Rev Laryngol* 106: 275-282.

Deliyski, D. D., Hillman, R. E., & Mehta, D. D. (2015). Laryngeal High-Speed Videoendoscopy: Rationale and Recommendation for Accurate and Consistent Terminology. *Journal of Speech, Language, and Hearing Research : JSLHR*, 58(5), 1488-1492.

Deliyski, D. D., Powell, M. E., Zacharias, S. R., Gerlach, T. T., & de Alarcon, A. (2015). Experimental Investigation on Minimum Frame Rate Requirements of High-Speed Videoendoscopy for Clinical Voice Assessment. *Biomedical signal processing and control*, 17, 21-28. <https://doi.org/10.1016/j.bspc.2014.11.007>

Demolin, D. (2004). Dynamics and articulatory control in Amharic ejectives, *Journal of the Ac Soc of America*. V. 115.

Echternach, M., Sundberg, J., Markl, M., & Richter, B. (2010). Professional Opera Tenors' Vocal Tract Configurations in Registers. *Folia Phoniatrica et Logopaedica*, 62(6), 278-287. <https://doi.org/10.1159/000312668>

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, 23(02), 157-191. <https://doi.org/10.1017/S095267570600087X>

Esling, J. H. (1999). The IPA categories « pharyngeal » and « epiglottal »: laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech*, 42 ( Pt 4), 349-372. <https://doi.org/10.1177/00238309990420040101>

Esling, John H. (2012). The articulatory function of the larynx and the origins of speech. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 38, p. 121–149).

- Esling, John H., & Moisik, S. R. (2011). Multimodal observation and measurement of larynx height and state during pharyngeal sounds. In *Proceedings of the 16th International Congress of Phonetic Sciences* (p. 643–646).
- Ewald, J. R. (1898). Die Physiologie des Kehlkopfes und der Luftröhre. Stimmbildung.
- Fabre, P. (1957). Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. Premiers résultats, p. 66-6.
- Fant, G. (1960). Acoustic theory of speech production.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84(1), 115-123.
- 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 pp:1234-1251.
- Fuks, L., & Sundberg, J. (1999, avril). Blowing Pressures in Bassoon, Clarinet, Oboe and Saxophone [Text]. Consulté 5 juin 2018, à l'adresse <http://www.ingentaconnect.com/content/dav/aaua/1999/00000085/00000002/art00014>
- Garrigues, L. (2015). Gestion des sons non pulmonaires et de la phonation inversée en en Human Beatbox. Mémoire d'orthophonie Paris VI.
- Granqvist, S., & Lindestad, P. A. (2001). A method of applying Fourier analysis to high-speed laryngoscopy. *The Journal of the Acoustical Society of America*, 110(6), 3193-3197.

- Habermann, W., Jiang, J., Lin, E., & Hanson, D. G. (2000). Correlation between glottal area and photoglottographic signal in normal subjects. *Acta Oto-Laryngologica*, 120(6), 778-782.
- Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean., *Journal of Phonetics*, p. 263-271.
- Hassan, Z. M., Esling, J. H., Moisiuk, S. R., & Crevier-Buchman, L. (2011). Aryepiglottic trilled variants of /ʕ, ħ/ in Iraqi Arabic. *Proc. XVII, edited by WS Zee and E. Zee (Hong Kong)*, 831-834.
- Hertegård, S., Gauffin, J., & Lindestad, P.-Å. (1995). A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice*, 9(2), 149-155. [https://doi.org/10.1016/S0892-1997\(05\)80248-6](https://doi.org/10.1016/S0892-1997(05)80248-6)
- Hipke, K., Toomim, M., Fiebrink, R., & Fogarty, J. (2014). BeatBox: end-user interactive definition and training of recognizers for percussive vocalizations (p. 121-124). *ACM Press*. <https://doi.org/10.1145/2598153.2598189>
- Kaneko, K., Watanabe, T., Kihara, C., Sakaguchi, K., & Takahashi, H. (2017). High-speed digital imaging system with a transnasal flexible fiberoptic endoscope for observing vocal fold vibration. *Acta Medica Nagasakiensia*, 61(1), p.45-47; 2017.
- Kapoor, A., & Allen, J. B. (2012). Perceptual effects of plosive feature modification. *The Journal of the Acoustical Society of America*, 131(1), 478-491. <https://doi.org/10.1121/1.3665991>
- Kitamura, T., Honda, K., & Takemoto, H. (2005). Individual variation of the hypopharyngeal cavities and its acoustic effects. *Acoustical Science and Technology*, 26(1), 16-26. <https://doi.org/10.1250/ast.26.16>

- Klatt, D. H. (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech, Language, and Hearing Research, 18*(4), 686-706. <https://doi.org/10.1044/jshr.1804.686>
- Kob, M., & Frauenrath, T. (2009). A system for parallel measurement of glottis opening and larynx position. *Biomedical Signal Processing and Control, 4*(3), 221-228. <https://doi.org/10.1016/j.bspc.2009.03.004>
- Koenig, L. L., & Lucero, J. C. (2008). Stop consonant voicing and intraoral pressure contours in women and children. *The Journal of the Acoustical Society of America, 123*(2), 1077-1088. <https://doi.org/10.1121/1.2828065>
- König, W. (1994). Atlas zur deutschen Sprache.
- Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. *Oxford: Blackwell.*
- Ladefoged, P. (1979). [Review of *Review of Fundamental Problems in Phonetics*, par J. C. Catford]. *Language, 55*(4), 904-907. <https://doi.org/10.2307/412751>
- Laver, J. (1980). The Phonetic description of voice quality. *Cambridge: Cambridge University Press.*
- Lederer, K. (2005). the phonetics of beatboxing. étude non publiée, disponible sur internet
- Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD, 20*(3), 384-422.
- Löfqvist, A., Carlborg, B., & Kitzing, P. (1982). Initial validation of an indirect measure of subglottal pressure during vowels. *The Journal of the Acoustical Society of America, 72*(2), 633-635. <https://doi.org/10.1121/1.388046>

- Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Swedish obstruent clusters. *The Journal of the Acoustical Society of America*, 68(3), 792-801.  
<https://doi.org/10.1121/1.384774>
- Lubker, J. F., & Parris, P. J. (1970). Simultaneous Measurements of Intraoral Pressure, 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(2B), 625-633. <https://doi.org/10.1121/1.1911938>
- Mamadou M, S., Ghio, A., Espesser, R., Bernard, T., Dramé, M., & Viallet, F. (2011). Relevance of Aerodynamic Evaluation in Parkinsonian Dysarthria. <https://doi.org/10.5772/23443>
- Marasek, K. (1996). Glottal correlates of the word stress and the tense/lax opposition in German. In , *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings* (Vol. 3, p. 1573-1576 vol.3).  
<https://doi.org/10.1109/ICSLP.1996.607920>
- Martino, R. (2009). Le Human Beatbox et ses pratiquants. Mémoire de sociologie, Université de Grenoble Alpes.
- Moisik, S. (2014). Towards a theory of Quantum Phonology: The search for quantal states of the larynx. Presentation
- Moisik, Scott R., Czaykowska-Higgins, E., & Esling, J. H. (2011). The epilaryngeal articulator: A new conceptual tool for understanding lingual-laryngeal contrasts. In *meeting for Phonology in the 21st Century: In Honour of Glyne Piggott (MOT)*, Montreal, Quebec.
- Moisik, Scott Reid, & Gick, B. (2017). The Quantal Larynx: The Stable Regions of Laryngeal Biomechanics and Implications for Speech Production. *Journal of*

*Speech Language and Hearing Research*, 60(3), 540.

[https://doi.org/10.1044/2016\\_JSLHR-S-16-0019](https://doi.org/10.1044/2016_JSLHR-S-16-0019)

- Muller, E., & Brown, W. S. (1979). Variations in the Supraglottal Air Pressure Waveform and Their Articulatory Interpretation. *Speech and language, advances in basic research and practice* pp323:396
- Ohala, J. J., & Riordan, C. J. (1979). Passive vocal tract enlargement during voiced stops, p. 89-92.
- Orlikoff, R. F. (1991). Assessment of the dynamics of vocal fold contact from the electroglottogram: data from normal male subjects. *Journal of Speech and Hearing Research*, 34(5), 1066-1072.
- Paroni, A. L. (2014). How do beatboxers play with their tongue and lips? An ultrasound and high-speed imaging exploration.
- Paroni, A. L. (2016). Production de sons plosifs : Comparaison du beatbox et de la parole.
- Patil, N., Greer, T., Blaylock, R., & Narayanan, S. S. (2017). Comparison of Basic Beatboxing Articulations Between Expert and Novice Artists Using Real-Time Magnetic Resonance Imaging (p. 2277-2281). ISCA. <https://doi.org/10.21437/Interspeech.2017-1190>
- Perello, J. (1973). Théorie muco-ondulatoire de la phonation. *Ann. Oto-rhino-laryngol.*
- Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, 92(6), 3078-3096. <https://doi.org/10.1121/1.404204>
- Picart, B., Brognaux, S., & Dupont, S. (2015). Analysis and automatic recognition of human beatbox sounds: A comparative study. In *Acoustics, Speech and Signal*

- Processing (ICASSP), 2015 IEEE International Conference on* (p. 4255–4259).  
IEEE.
- Poupineau, C. (2018). Elaboration d'un outil de rééducation orthophonique basé sur le Human beatbox pour des patients présentant une immobilité laryngée unilatérale. Mémoire de Master 2 en Orthophonie.
- 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, *J. Acoust. Soc. Am.* p. 1043-1054.
- Prosek, R. A. (1975). Intraoral air pressure as a feedback cue in consonant production. *Journal of Speech and Hearing Research*, 18(1), 133-147.
- Ridouane, R., Audibert, N., & Minh Nguyen, V. (2012). Les ajustements laryngaux en français. <https://doi.org/10.13140/2.1.3336.5764>
- Roubeau, B. (1993). Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale. Thèse Université Paris XI, Orsay.
- Russel, G. O. (1928). The vowel, its psychological mechanism, as shown by x-ray.
- Sapthavee, A., Yi, P., & Sims, H. S. (2014). Functional Endoscopic Analysis of Beatbox Performers. *Journal of Voice*, 28(3), 328-331. <https://doi.org/10.1016/j.jvoice.2013.11.007>
- Schenk, F., Aichinger, P., Roesner, I., & Urschler, M. (2015). Automatic high-speed video glottis segmentation using salient regions and 3D geodesic active contours. *Annals of the BMVA, 2015*, 1-15.
- Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31(1), 26-35. [https://doi.org/10.1016/0093-934X\(87\)90058-7](https://doi.org/10.1016/0093-934X(87)90058-7)



- Sinyor, E., Rebecca, C. M., Mcennis, D., & Fujinaga, I. (2005). Beatbox classification using ACE. In *Proceedings of the International Conference on Music Information Retrieval*.
- Stevens, K., Manuel, S. Y., & Matthies, M. (1999). REVISITING PLACE OF ARTICULATION MEASURES FOR STOP CONSONANTS: IMPLICATIONS FOR MODELS OF CONSONANT PRODUCTION. *Proceedings of the International Congress of Phonetic Sciences*, 1117–1120.
- Stone, M. (1983). Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. *J Phon*, 13, 207-218.
- Stowell, D., & Plumbley, M. D. (2008). Characteristics of the beatboxing vocal style. *Queen Mary, University of London, Technical Report*, 4.
- Sussman, H. M., Fruchter, D., & Cable, A. (1995). Locus equations derived from compensatory articulation. *The Journal of the Acoustical Society of America*, 97(5 Pt 1), 3112-3124.
- Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *The Journal of the Acoustical Society of America*, 94(3 Pt 1), 1256-1268.
- Suthau, E., Birkholz, P., Mainka, A., & Simpson, A. P. (2016). Non-invasive photoglottography for use in the lab and the field. *Speech Communication*, 5.
- Švec, J. G., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2), 201-205.  
[https://doi.org/10.1016/S0892-1997\(96\)80047-6](https://doi.org/10.1016/S0892-1997(96)80047-6)
- Trullinger, R. W., & Emanuel, F. W. (1983). Airflow Characteristics of Stop-Plosive Consonant Productions of Normal-Speaking Children. *Journal of Speech*,

*Language, and Hearing Research*, 26(2), 202-208.

<https://doi.org/10.1044/jshr.2602.202>

TyTe, & Splinter. (2002). Standard Beatbox Notation (SBN). [Online]. Available:

<https://www.humanbeatbox.com/articles/standard-beatbox-notation-sbn/>.

Van den Berg, J. (1958). Myoelastic aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1(3), 227-244.

Westbury, J., Milenkovic, P., Weismer, G., & Kent, R. (1990). X-ray microbeam speech production database. *The Journal of the Acoustical Society of America*, 88(S1), S56-S56. <https://doi.org/10.1121/1.2029064>

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America*, 73(4), 1322-1336.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech Language and Hearing Research*, 48(3), 543. [https://doi.org/10.1044/1092-4388\(2005/037\)](https://doi.org/10.1044/1092-4388(2005/037))

## Table des illustrations

Figure 1 : Signal acoustique de /bu/ et son spectrogramme.....	7
Figure 2 : Signal acoustique de /pu/ et spectrogramme.....	8
Figure 3 : lieux d'articulation des consonnes occlusives et éjectives des langues du monde.....	9
Figure 4 : Consonnes affriquées.....	10
Figure 5 : faux palais d'électropalatographie. (image trouvée sur le site du Laboratoire de Phonétique et Phonologie (CNRS Sorbonne Nouvelle).....	12
Figure 6 : Exemple de radiographie cervicale dans le plan sagittal prise dans un contexte médical (CHU de Grenoble). Les structures annotées sont : base de langue (BdL), os hyoïde (OH), épiglotte (Ep), cartilage cricoïde (CC), Repli aryépiglottique (RAE), cartilages aryténoïdes (CA) Vertèbre cervicale C1, corps étranger intra oesophagien (CIE, arrête de poisson).....	13
Figure 7 : principe de la pharyngoscopie rigide.....	16
Figure 8 : Principe de la nasofibroscope.....	16
Figure 9 : Vidéokymogramme d'oscillation glottique.....	18
Figure 10 : Vues prises en fibroscopie souple : 1) vue classique supérieure, 2) vue postérosupérieure haute.....	31
Figure 11 : 1) sujet avec le tube de mesure de Pression Intra Orale en position, il est glissé par le sujet au niveau du sillon amygdalogue. 2) Recueil de mesure avec le système EVA2.....	32
Figure 12 : exemple d'extraction de la durée totale d'un effet, ici hihat dans une production de Beatbox dans la tâche « pâtes au pesto », s'étendant de la fin du kick au début du snare.....	34
Figure 13 : Mise en relation de la PIO avec le DAO (tirée de Muller et Brown, 1979), la courbe de PIO est située en haut, la courbe de débit d'air oral en bas de la figure.....	38
Figure 14 : Illustration d'une séquence laryngée en parole pour l'enchaînement de /bu/ (phrase « boots and cats »). d (ms) se réfère en burst.....	42
Figure 15 : Illustration d'une séquence laryngée pour l'enchaînement de /pes/ (phrase pâtes au pesto »). d (ms) se réfère en burst.....	44
Figure 16 : séquence temporelle d'un kick voisé, les flèches montrent l'espace rétrocricoïdien ouvert, et l'hypercontraction des plis vestibulaires avec le verrouillage aryténoïdien.....	47
Figure 17 : illustration de la production d'un kick par le sujet 2.....	49
Figure 18 : comparaison des signaux acoustiques du kick du sujet 1 et 2.....	49
Figure 19 : illustration de la production d'un /t/ du sujet 1 dans /to/ de « pâtes au pesto ».....	51
Figure 20 : Séquence temporelle d'un hihat.....	53
Figure 21 : Séquence de hihat en vue postéro supérieure, on voit bien le verrouillage aryténoïdien en figure 2 et 3. On peut noter le petit triangle ouvert postérieur mentionné ci-dessus.....	55
Figure 22 : signaux acoustiques et spectrogrammes de Fast hihat (FH) vs classic hihat (T). La courbe pleine sur le spectrogramme présente l'intensité acoustique.....	56
Figure 23 : Séquence laryngée d'un Inward hihat.....	58
Figure 24 : illustration de la production d'un hihat par le sujet 2.....	59
Figure 25 : production du Mesh snare et sa séquence gestuelle.....	62
Figure 26 : /pf snare/ et séquence vidéo.....	64
Figure 27 : Mesh snare du sujet 2.....	66
Figure 28 : séquence gestuelle laryngée du /k/ dans « poutika ».....	68
Figure 29 : Rimshot du sujet 1 en vue postérosupérieure.....	71
Figure 30 : Rimshot ingressive du sujet 2.....	73
Figure 31 : séquence laryngée de Rimshot « variante » du sujet 2.....	74
Figure 32 : données temporelles des débuts de mouvements des structures par rapport au burst....	75
Figure 33 : Données de durée du mouvement à partir du début du mouvement.....	75
Figure 34 : Boxplots par effet des intensités maximales, des durées de production par consonne/effet : les effets du Beatbox sont notés en majuscule (P = kick, Ps = snare, T= hihat, K = Rimshot).....	77
Figure 35 : Intensité maximale en fonction des moments spectraux, des maximums de PIO et DAO par effet.....	78

Figure 36 : Boxplots des coefficients a, b d et a_oaf respectivement de gauche à droite et haut en bas.....	79
Figure 37 : Boxplots des différentes variables : maximas des PIO et DAO, pression moyenne d’occlusion et au moment du relâchement. ....	80
Figure 38 : relations linéaires entre dP et dT pour les effets du Beatbox en comparaison avec les consonnes. ....	80
Figure 39 : production de /b/ en parole avec données aérodynamiques en rapport.....	82
Figure 40 : production d’un /p/ en parole et données aérodynamiques.....	83
Figure 41 : production d’un /t/ et données aérodynamiques en rapport. ....	83
Figure 42 : production d’un /k/ en parole et données aérodynamiques.....	84
Figure 43 : production d’un kick en Beatbox et données aérodynamiques.....	84
Figure 44 : production d’un snare en Beatbox et données aérodynamiques. ....	85
Figure 45 : production d’un hihat et donnés aérodynamiques. ....	85
Figure 46 : production d’un Rimshot et données aérodynamiques. ....	86
Figure 47 : Consonnes et sons percussif du Beatbox. Les données sont en valeur absolue facilitant la comparaison entre une consonne et son effet correspondant en Beatbox.....	86
Figure 48 : exemple de variantes de l’effet snare : alternance de /pf snare/ et/mesh snare/.....	87
Figure 49 : maximum de débit d’air nasal selon la composante « Humming » pour les quatre effets étudiés du Beatbox. Nh = version classique de l’effet, h = version humming .....	87
Figure 50 : données acoustiques et EGG des consonnes en parole normale. Les flèches correspondent au Burst.....	88
Figure 51 : EGG et données de PIO pour les effets du Beatbox. Les flèches correspondent aux instants du burst.....	89
Figure 52 : /b/, /p/, /t/, /k/ en séquences reconstruites : gestes laryngés et données aérodynamiques mises en parallèle .....	93
Figure 53 : Relations temporelles entre gestuelle laryngée et données aérodynamiques pour les effets du Beatbox.....	96
Figure 54 : Exemple de visualisation des données échographiques à partir de notre script : en haut à gauche une image échographique de profil du larynx de profil au repos, on voit aisément la bande ventriculaire et le pli vocal en dessous (profil en rouge). Le cadran en haut à droite montre le détail du signal audiométrique correspondant au temps de l’image, et en bas nous avons la position de l’image dans le signal audio.....	97
Figure 55 : images échographiques correspondantes au temps notés par les barres verticales sur le signal audiométrique. Les traits de visualisation des structures laryngées montrent une ascension laryngée pendant la production d’un « kick » à la différence des /p/.....	98
Figure 56 : Enregistrement synchronisé au moyen de la station EVA et de la caméra Mikrotron	122
Figure 57 : Enregistrements EVA avec et sans introduction de l’endoscope.....	123

## **Table des annexes**

Annexe 1 Elaboration des techniques expérimentales .....	1250
Annexe 2 Moyennes des temps vidéoendoscopiques.....	122
Annexe 3 Liste des abréviations.....	123

## Chapitre Annexe. Elaboration des techniques expérimentales

Nous décrivons dans ce chapitre la mise en place techniques des dispositifs de mesure et leur évaluation sur sujet sain.

### *4.1. Cinématographie haute vitesse et limites*

Il était prévu initialement un enregistrement de données par vidéo-nasofibroscopie souple au moyen d'un nasofibroscope de marque Xion et d'une caméra MotionBLITZ® EoSens Cube6 couleur (MIKROTRON GmbH, Allemagne).

Après montage d'une plateforme (photo) permettant la bonne stabilisation de la caméra et de l'endoscope, différents essais ont été menés sur l'auteur de ce mémoire afin de tester ce système. Un dispositif de synchronisation (« trigger ») permettant le déclenchement de la prise d'image a été branché en parallèle à l'entrée EGG de la station EVA2 (détails ci-dessous) pour synchroniser les données filmographiques et acoustiques.

Des enregistrements à différentes vitesses avec une ouverture maximale de l'objectif ont été réalisés.

Les données cinématographiques ont été exportées en format brut (.raw) afin de subir un post-traitement grâce à différents logiciels (MATLAB\_R2018a et ImageJ).

Malgré un certain nombre de tentatives de traitement d'image à l'aide d'algorithmes que nous ne détaillerons pas ici, il a été déterminé que les caractéristiques de la caméra en termes de sensibilité lumineuse (2000 ISO) ne permettent pas d'enregistrement de qualité suffisante à plus de 100 images par seconde in vivo avec l'endoscope souple.

Un essai de montage avec un endoscope rigide garantit une qualité d'image supérieure mais est impropre à l'étude des consonnes (comme nous l'avons mentionné plus haut) et la tolérance de l'endoscope fixé sur plateforme et placé en endobuccal est très mauvaise (réflexe nauséux).

Un enregistrement de données endoscopiques sur l'auteur et une volontaire a aussi été réalisé au CHU dans les mêmes conditions que décrites ci-dessous. Plusieurs scripts sous Matlab et ImageJ de traitements d'image ont été faits afin de permettre une

automatisation de l'analyse, notamment un script d'apprentissage automatique, mais sans succès actuellement en raison d'une variation importante de luminosité, de pertes de visualisation des structures qui sont souvent cachées par d'autres dans la production des consonnes, et de l'absence de point de référence fixe. Nous avons testé pour référence le centre de l'aire glottique, mais lors de la fermeture épilaryngée (plus ou moins avec bascule du larynx), on ne peut qu'en approximer la position. De même la vue postérosupérieure que nous détaillerons plus tard ne permet pas la détermination précise du centre de l'aire glottique.

Il existe de potentiels points fixes pouvant servir de référence, pour déterminer la différence entre mouvement de la caméra (et du vélum) et des structures pharyngolaryngées, situés entre la paroi pharyngée postérieure et les parois pharyngées latérales. Un marquage de ces points par application d'un coton imbibé de bleu de méthylène a été testé sur l'auteur, mais est à déconseiller en raison de la mauvaise tolérance du marquage (même avec anesthésie locale à la Xylocaïne spray) et de sa disparition rapide en contexte de production normale de parole (il resterait aussi à déterminer si l'on peut parler de points fixes en production de Beatbox avec des constriction pharyngées importantes).

Ainsi les analyses de vidéo-endoscopie seront principalement basées sur une analyse qualitative visuelle, puisque nous n'avons actuellement pas réussi à généraliser des mesures quantitatives sur les mouvements laryngés dans des contextes et des types de production si divers.

#### ***4.2. Mesures aérodynamiques***

Le système d'Evaluation Vocale Assisté EVA2™ (SQLab, Aix-en-Provence, France) a été utilisé avec l'environnement Phonedit SIGNAIX 4.2.0.8 (développé par le laboratoire Parole et langage, CNRS-Université d'Aix-en-Provence) pour un enregistrement simultané des :

- \_ Débit d' Air Oral (DAO),
- \_ Débit d' Air Nasal (DAN),
- \_ Pression Intra Orale (PIO),
- \_ Signaux acoustiques.

Il a été fait deux séquences en parole normale et en « pseudo-Beatbox » (l'auteur n'étant pas pratiquant de cet Art), avec et sans la mise en place du nasofibroscope.

Le premier enregistrement était réalisé avec l'endoscope dans la fosse nasale droite, fermée hermétiquement avec un petit bouchon, l'enregistrement du débit d'air nasal était donc fait au niveau du seuil narinaire gauche. En l'absence du nasofibroscope, l'enregistrement se faisait des deux côtés avec le système « classique ».

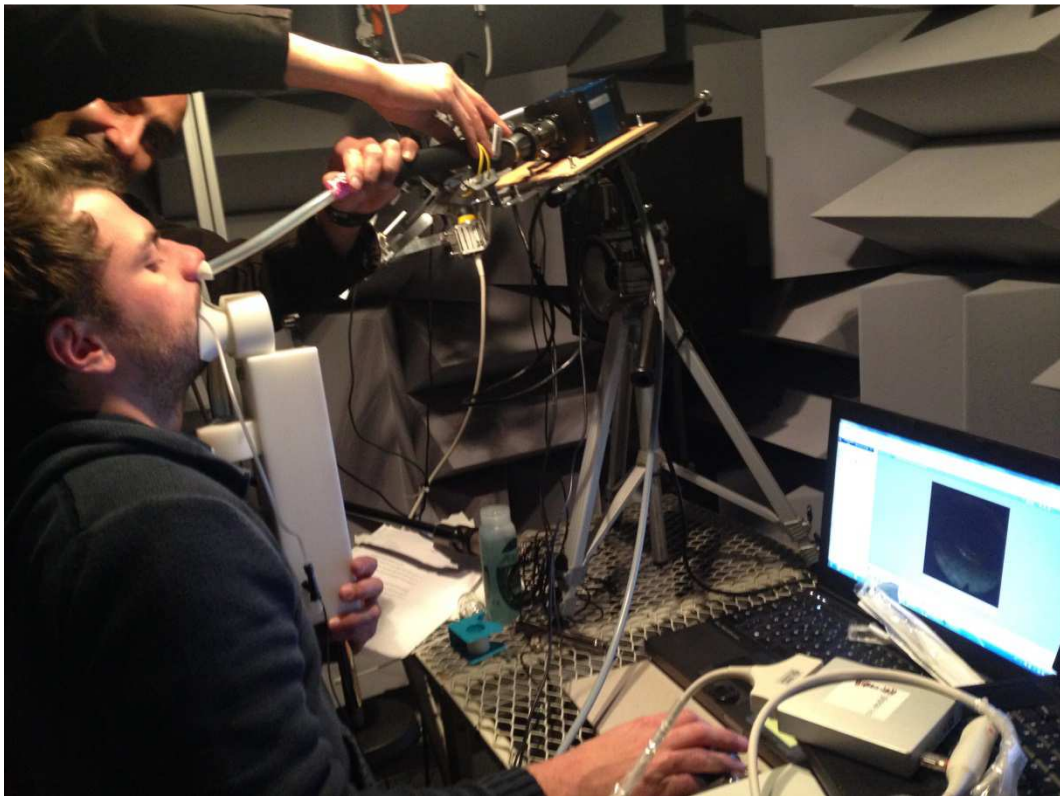


Figure 56 : Enregistrement synchronisé au moyen de la station EVA et de la caméra Mikrotron

Le but de ces enregistrements était de vérifier l'absence de fuite d'air nasale et d'interaction entre les deux systèmes de mesure.

Il n'a pas été réalisé d'analyse statistique en raison de l'absence claire de différence significative de débit d'air nasal en production langagière ou de Beatbox avec ou sans la fibroscopie.



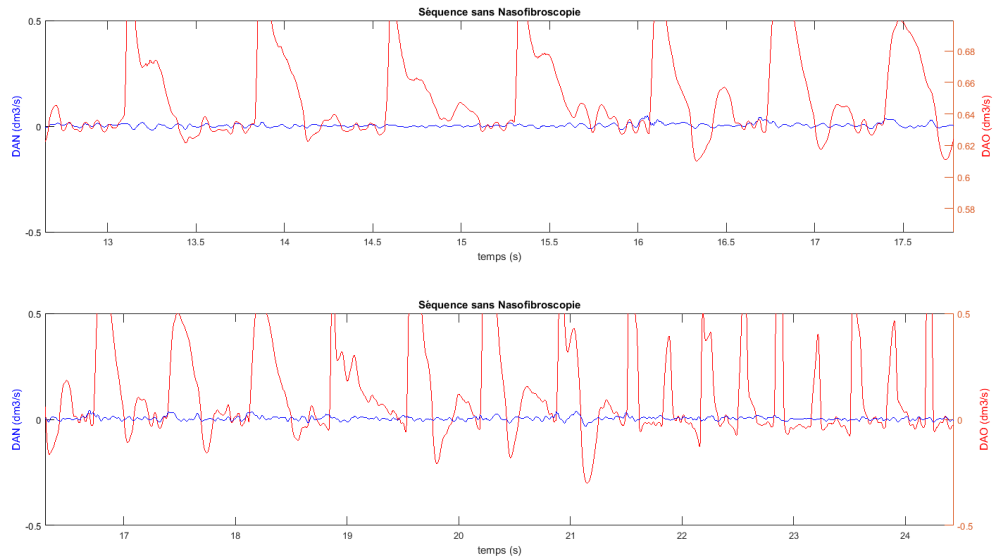


Figure 57 : Enregistrements EVA avec et sans introduction de l'endoscope.

### 4.3. Echographie

Les paramètres de mesure et la position de la sonde permettant la meilleure visualisation des structures laryngées ont été empiriquement déterminés au préalable par des essais sur l'auteur de ce mémoire.

Nous avons trouvé les paramètres suivants optimaux pour l'évaluation laryngée avec le boîtier d'échographie portable Telemed et une sonde :

- \_ fréquence ultrasonore : 8 MHz
- \_ puissance : 84dB
- \_ gain : 79%
- \_ profondeur de champ : entre 3 et 5cm
- \_ 100 images par secondes

Pour l'étude des mouvements verticaux laryngés nous avons trouvé une position de sonde se trouvant approximativement à 1 travers de doigt en arrière et à droite de la proéminence du cartilage thyroïde, dans un plan strictement vertical et légèrement orienté en oblique dans le plan avant-arrière afin de visualiser le processus musculaire du cartilage aryénoïde droit. En effet, il est difficile de distinguer les différentes structures du larynx en raison de la présence d'air nous poussant à rechercher les cartilages car ils sont hyperéchogènes, le processus musculaire du cartilage aryénoïde est facilement repéré en phonation ainsi que les vibrations des plis vocaux. L'ascension et la bascule en avant du

larynx sont donc évaluables à partir des mouvements au niveau des cartilages thyroïde et aryténoïdes.

Enfin, un dispositif permettant une tenue fixe de la sonde au moyen d'un harnais de caméra à l'épaule est en cours de fabrication et d'évaluation afin d'éliminer les changements même minimes de positionnement de la sonde dus à l'opérateur ou au sujet.

## Annexe 2

Consonnes :

/p/				
	Début mouvement		Fin mouvement	
	OM / D/ BA	FM / C /BR	OM / D/ BA	FM / C /BR
CV	-465 (-480 ; -260)	-228 (-298; -151)	-322	-78 (-173 ; -14)
Hypopharynx	-184.85 (-224.4 ; -134.42)		24.85 (16 ; 47)	
Etat glottique	-252.84 (-285.92 ; -192.78)		-47.38 (-117 ; 0.46)	
/b/				
	Début mouvement		Fin mouvement	
	OM / D/ BA	FM / C /BR	OM / D/ BA	FM / C /BR
CV		-319 (-380 ; -272)		-189 (-254.44 ; -140.75)
Hypopharynx	-71.33 (-102 ; -51.8)		168.83 (125.66 ; 237.20)	
Etat glottique	-413 (-430 ; -396)	-179 (-245.6 ; -136.4)	-263 (-279 ; -252.3)	219 (173.6 ; 272.)
/t/				
	Début mouvement		Fin mouvement	
	OM / D/ BA	FM / C /BR	OM / D/ BA	FM / C /BR
CV		-92.5 (-101 ; -92.5)		30 (22 ; 38)
Etat glottique	-95.5 (-187.92 ; -95.5)	-4 (-42 ; -2)	-4 (-95.5 ; 67.5)	67.5 (-4 ; 139)
/k/				
	Début mouvement		Fin mouvement	
	OM / D/ BA	FM / C /BR	OM / D/ BA	FM / C /BR
Pharynx haut		-29 (-51 ; -14.33)		154 (115 ; 174)

Epiglotte		-47 (-69 ; -36)		106 (85 ; 142.5)
Hypopharynx	-104 (-144 ; -104)		18.5 (-1 ; 18.5)	
Etat glottique		-108 (-118 ; -108)		55 (4.99 ; 54)

**Beatbox :**

P				
	Début mouvement		Fin mouvement	
	OM / D/ BA/O	FM / C /BR/A/F	OM / D/ BA/O	FM / C /BR/A/F
Aryténoïdes		-56.58 (-69 ; -46)		13.41 (-0.002 ; 29)
Epiglotte		-30.75 (-51 ; -18)		43.37 (16 ; 62)
Retrocricoïde	-60.47 (-71 ; -51)		12.13 (-0.49 ; 20)	
Pharynx haut		-31.73 (-43 ; -22)		62.78 (49.50 ; 79)
Hypopharynx	-76.43 (-85 ; -69)		57.56 (46 ; 67)	
BV		-83 (-109 ; -62)		-15 (-33 ; 6)
Larynx		-31.83 (-59.3 ; -13.6)		67 (29.16 ; 88.16)
CV	-243 (-357 ; -76)	-182.8 (-199 ; -164)	-177.66 (-296 ; -10)	-65.35 (-79 ; -53)

Etat glottique	-309.27 (-343;-252)	-90.07 (-15 ;-67)	-148.22 (-181 ;-67)	30.51 (9 ;57)
<b>PS</b>				
	Début mouvement		Fin mouvement	
	OM / D/ BA/O	FM / C /BR/A/F	OM / D/ BA/O	FM / C /BR/A/F
Retrocricoïde	-33 (-45 ;-18)		166 (115 ;251)	
Hypopharynx	-73 (-92 ;-58)		373 (275 ;485)	
Larynx	-33 (-53 -7)		52 (16 ;143)	
CV	-23.75 (-30.5 ;-13.5)	-242.5 (-319 ;-193)	103.5 (57 ;137)	-18 (-28 ;-4)
Etat glottique	-34,1 (-41 ;0.12)		105 (88 ;386)	
<b>T</b>				
	Début mouvement		Fin mouvement	
	OM / D/ BA/O	FM / C /BR/A/F	OM / D/ BA/O	FM / C /BR/A/F
Aryténoïdes		-44 (-56 ;-32)		45 (20 ;83)
Epiglotte	-99		44	
Retrocricoïde	-49 (-61 ;-37)		44 (31 ;56)	
Pharynx haut	-13 (-23 ;-13)	-66.27 (-99 ;-44)	242 (202 ;242)	71 (37 ;93)
Hypopharynx	-104 (-133 ;-93)		37 (24.6 ;49)	

Larynx		-9.5 (-21 ; -9)		92 (81 ; 92)
CV	-222 (-282 ; -155)	-181 (-211 ; -148)	-33 (-139.28 ; 33)	-45 (-70 ; -13)
Etat glottique	-198 (-261 ; -15)	-261 (-281 ; -150)	19 (-111 ; 139)	249 (13 ; 334)
K				
	Début mouvement		Fin mouvement	
	OM / D/ BA/O	FM / C /BR/A/F	OM / D/ BA/O	FM / C /BR/A/F
Epiglotte		-84 (-123 ; -43)		58 (47 ; 72)
Retrocricoide	-71.42 (-86 ; -56)		30 (17 ; 42)	
Pharynx haut		-59 (-42 ; -77)		50 (16 ; 64)
CV	-155 (-153 ; -158)	-253 (-267 ; -239)	290 (-12 ; 312)	-121.5 (-158 ; -108)
Etat glottique	-272 (-302 ; -252)	-259.33 (-273 ; -247)	-94.25 (-148 ; -35)	-82.66 (-155 ; -40)

Annexe 1 : moyenne des débuts et fins avec leur intervalle de confiance retrouvés en Bootstrap, des temps correspondants à 1 mouvement (1ere colonne de type augmentation de volume/ouverture des sphincters, 2<sup>e</sup> colonne : fermeture ou réduction de volume) pour les consonnes et les effets de Beatbox.

### **Annexe 3**

Table des abréviations :

a : coefficient de pente entre début de la phase fermée et le temps de pression moyenne d'occlusion

a\_oaf : pente du débit d'air oral entre temps de relâchement et maximum sur un cycle

b : coefficient de pente entre temps de pression moyenne d'occlusion et pression au moment du relachement

BV : plis ventriculaires

C : dans les séquences temporelles (Praat) signifie croissance

CV : plis vocaux

D : dans les séquences temporelles (Praat) signifie croissance

d : a-b permettant de caractériser la forme de la courbe

DAO : débit d'air oral

DAN : débit d'air nasal

EGG : électroglottographie

PIO : pression intraorale

# Table des matières

Sommaire .....	1
Introduction .....	2
<b>PARTIE 1 - CADRE THEORIQUE .....</b>	<b>3</b>
CHAPITRE 1. CONTEXTE DE RECHERCHE .....	4
1. Bref rappels de physiologie de la parole.....	4
1.1. La soufflerie .....	4
1.2. Le vibreur.....	4
1.3. Les articulateurs et résonateurs .....	5
2. Le larynx comme articulateur .....	5
3. Les consonnes.....	6
3.1. Les consonnes occlusives ou plosives.....	6
3.2. Consonnes affriquées .....	9
4. Mesures expérimentales articulaires des consonnes plosives .....	10
4.1. Les signaux acoustiques (Calliope 1989).....	10
4.2. Evaluation de l'articulation bucco-linguale.....	11
4.2.1. La palatographie .....	11
4.2.2. Les techniques radiologiques .....	12
4.2.3. Imagerie par Résonance Magnétique (IRM).....	13
4.2.4. Echographie .....	14
4.2.5. Articulographie Electro Magnétique (EMA).....	15
4.3. L'exploration laryngée.....	15
4.3.1. L'endoscopie.....	15
4.3.2. Autres systèmes optiques .....	18
4.3.3. Mesures électromyographiques.....	18
4.3.4. L'Electroglottographie.....	19
4.3.5. Mesures aérodynamiques indirectes.....	19
4.4. Conclusion sur les méthodes exploratoires .....	20
5. Le Human Beatbox : état de l'art.....	20
5.1. Un art vocal récent .....	20
5.2. Etude scientifique du Beatbox.....	22
5.2.1. Aspect paralinguistique.....	22
5.2.1. Aspect acoustique et phonétique.....	23
5.3. Son potentiel thérapeutique.....	24
CHAPITRE 2. QUESTIONS DE RECHERCHE.....	26
1. Est-il possible de corrélérer les gestes laryngés produits pendant le Human Beatbox à des mesures aérodynamiques et acoustiques ?.....	26
2. Quelles sont les différences entre les consonnes plosives de la parole normale et leurs équivalents en Human Beatbox tant au point de vue des gestes laryngés qu'au niveau de l'aérodynamique et de la production sonore ?.....	26
3. Un même effet peut-il être produit par des stratégies acoustiques différentes ? .....	26
<b>PARTIE 2 - EVALUATION EXPERIMENTALE .....</b>	<b>27</b>
CHAPITRE 4. MATERIEL ET METHODES .....	28
4. Description des sujets .....	28
4. Effets étudiés et consonnes associées .....	28
5. Protocole expérimental .....	29
5.1. Corpus.....	29
5.2. Méthodes de mesure.....	30
5.2.1. Enregistrements endoscopiques .....	30
5.3. Mesures aérodynamiques et EGG .....	31
5.4. Echographie laryngée.....	32
6. Méthodes d'analyse .....	33
6.1. Endoscopie.....	33
6.2. Mesures aérodynamiques .....	37
6.3. Mesures échographiques .....	39



7. Approche statistique .....	39
CHAPITRE 5. RESULTATS .....	41
1. Caractérisation endoscopique des mouvements laryngés concomitants à la production des trois effets..	41
1.1. kick - /p/-/b/.....	41
1.1.1. Sujet 1 : beatboxeur professionnel .....	41
1.1.2. Sujet 2 : beatboxeur amateur.....	48
1.2. hihat - /t/.....	50
1.2.1. Sujet 1 : beatboxeur professionnel .....	50
1.2.2. Sujet 2 : beatboxeur amateur.....	58
1.3. snare - /Ps/.....	59
1.3.1. Sujet 1 : beatboxeur professionnel .....	59
1.3.2. Sujet 2 : beatboxeur amateur.....	65
1.4. rimshot - /k/.....	66
1.4.1. Sujet 1 : beatboxeur professionnel .....	66
1.4.2. Sujet 2 : beatboxeur amateur.....	71
1.5. tableaux résumé des données temporelles.....	76
2. Etude aérodynamique et acoustique.....	76
2.1. Acoustique .....	76
2.2. PIO et débit d'air oral (DAO) .....	78
2.3. Débit d'air nasal (DAN).....	87
2.4. EGG .....	87
3. Corrélations entre les données endoscopiques et EVA2 .....	89
4. Echographie.....	96
<b>PARTIE 3 - DISCUSSION ET MISES EN PERSPECTIVE .....</b>	<b>99</b>
CHAPITRE 6. DISCUSSION .....	100
1. Discussion et perspectives .....	100
2. Limites.....	103
Bibliographie.....	107
Table des illustrations.....	117
Table des annexes.....	119
ANNEXE 1 CHAPITRE ANNEXE. ELABORATION DES TECHNIQUES EXPERIMENTALES.....	120
3.1. Cinématographie haute vitesse et limites .....	120
3.2. Mesures aérodynamiques .....	121
3.3. Echographie .....	123
ANNEXE 2 TABLEAU MOYENNES TEMPORELLES VIDEOENDOSCOPIQUES.....	12022
Table des matières .....	130

**MOTS-CLÉS :** Beatbox, consonnes plosives, endoscopie, articulation laryngée

### **RÉSUMÉ**

Le Human Beatbox est une discipline vocale récente dont l'étude n'en est qu'à ses balbutiements. Après avoir décrit l'articulation laryngée d'effets du Beatbox comparativement aux consonnes plosives du Français. Nous essayons de corréler des données endoscopiques avec des données aérodynamiques recueillies chez un de nos sujets.

Les mouvements pharyngo-laryngés du Human Beatbox montrent une grande unité d'articulation au sein d'une famille d'effet, mais avec des conformations laryngées extrêmes permettant une puissance vocale très importante mise en relation avec des variations aérodynamiques très fortes et rapides.

**KEYWORDS:** Human Beatbox, plosive consonant, endoscopy, laryngeal articulation

### **ABSTRACT**

Human Beatbox is a recent vocal art. Little scientific work is available regarding this topic. We describe the mechanisms of laryngeal articulation on main effects of Human Beatbox, and compare them to the movements of articulation of speech consonants. Then we correlate laryngeal movements with aerodynamic data for one subject. Laryngeal articulation of Human Beatbox is highly reproducible within the same effect. Extreme conformations of laryngeal states in Beatbox, high amplitudes of the aerodynamic parameters and their rapid variation can be related to the high vocal energy of this paralinguistic production.

