

Université Paris III – Sorbonne Nouvelle  
Institut de Linguistique et Phonétique Générales Appliquées



Mémoire de Master 2 Traitement Automatique des Langues  
Parcours Recherche & Développement

Présenté par Clara PONCHARD

**Classification automatique  
des lieux d'articulation des consonnes occlusives  
du français**

Sous la direction  
du Professeur Didier DEMOLIN & de Monsieur Nicolas AUDIBERT

Année Universitaire 2018-2019

## Attestation de non-plagiat

Déclaration sur l'honneur

Je soussignée, Clara Ponchard, déclare avoir rédigé ce travail sans aides extérieures ni sources autres que celles qui sont citées. Toutes les utilisations de textes préexistants, publiés ou non, y compris en version électronique, sont signalées comme telles. Ce travail n'a été soumis à aucun autre jury d'examen sous une forme identique ou similaire, que ce soit en France ou à l'étranger, à l'université ou dans une autre institution, par moi-même ou par autrui.

Date

Signature manuscrite de l'étudiante

# Table des matières

ABSTRACT	5
1. INTRODUCTION	6
2. ÉTAT DE L'ART	7
2.1 PARTIE PHONETIQUE	7
2.1.1 Phase aérodynamique	8
2.1.2 Pression intra-orale	10
2.1.3 Pression sous-glottique	11
2.1.4 Phase acoustique	12
2.2 PARTIE TRAITEMENT AUTOMATIQUE DES LANGUES	13
2.2.1 Segmentation automatique	15
3. CORPUS DE TRAVAIL	19
3.1 SUJETS ET MATERIELS	19
3.2 BASE DE DONNEES	21
3.3 PROCEDURES	22
4. METHODES	23
4.1 SEGMENTATION MANUELLE	23
4.2 SEGMENTATION AUTOMATIQUE	26
4.3. COMPARAISON ENTRE SEGMENTATION MANUELLE ET SEGMENTATION AUTOMATIQUE	27
5. RESULTATS ET DISCUSSION	35
5.1 EXTRACTIONS DES DESCRIPTEURS	35
5.2 ANALYSE DES DONNEES AERODYNAMIQUES	39
5.2.1 Analyse de la pression intra-orale	39
5.2.1.1 L'opposition de voisement	41
5.2.1.2 Le contexte intervocalique	46
5.2.1.3 Les lieux d'articulation	54
5.2.2 Pression sous-glottique	62
5.2.2.1 L'opposition de voisement	63
5.2.2.2 Le contexte intervocalique	67
5.2.1.3 Les lieux d'articulation	74
5.2.3 La durée	80
5.3 CLASSIFICATION AUTOMATIQUE	83

5.3.1 L'opposition de voisement	85
5.1.2 Le contexte intervocalique	91
5.3.3 Les lieux d'articulation	97
6. CONCLUSION	102
A. BIBLIOGRAPHIE	103
B. LISTE DES FIGURES	107
C. LISTE DES TABLEAUX	111
D. ANNEXES	118
D.1 CONVERSION FICHIERS MRK EN TEXTGRID	118
D.2 SCRIPT D'EXTRACTION AUTOMATIQUE DES MESURES AERODYNAMIQUES	120



## Abstract

Ce mémoire décrit un travail de recherche qui vise à étudier les paramètres aérodynamiques impliqués dans la production des occlusives du français. Nous centrerons notre analyse sur les mesures de pression intra-orale et de pression sous-glottique. L'objectif majeur de cette étude est d'automatiser le traitement des données aérodynamiques afin de contribuer à apporter des données de référence qui sont en nombre limité dans la littérature. Nous souhaitons également analyser les variations de pression en fonction de l'opposition de voisement, du contexte intervocalique et des différents lieux d'articulation. Pour ce faire, nous procéderons à une analyse phonétique au moyen de tests statistiques afin d'analyser les descripteurs jugés pertinents pour comprendre les variations des segments consonantiques. Puis, nous procéderons à une tâche de classification supervisée afin d'évaluer la pertinence de ces descripteurs. Le champ de recherche est donc pluridisciplinaire et fait intervenir des méthodes relevant à la fois du domaine de la phonétique, du Traitement Automatique des Langues et de l'apprentissage automatique.

# 1. Introduction

L'analyse des données aérodynamiques est un champ de recherche peu exploré dans le domaine du Traitement Automatique des Langues. En effet, les travaux déjà effectués dans le domaine de la classification automatique de signaux audio se basent généralement sur des indices acoustiques tels que la fréquence fondamentale, les formants, les coefficients spectraux à fréquence Mel (MFCC) etc. Les données aérodynamiques sont peu exploitées car elles sont difficiles à acquérir puisqu'elles nécessitent, pour certaines mesures, d'un encadrement hospitalier. De plus, il n'existe pas de méthodes permettant de les traiter de façon automatique ce qui limite grandement la taille des corpus. Nous souhaitons dans ce travail de recherche exploiter ces mesures afin d'identifier s'il peut s'agir d'un paramètre pertinent pour la classification automatique de segments consonantiques. Nous pensons qu'il s'agit de données importantes à prendre en compte puisque c'est au stade de la phase aérodynamique que les sons de la parole sont générés. Leur analyse s'avère donc indispensable pour comprendre les phénomènes qui régissent la production des occlusives du français et étudier les mouvements des organes articulateurs du conduit vocal. Notre motivation principale est donc d'automatiser le traitement des données aérodynamiques afin de réaliser une analyse approfondie de ces mesures à plus grande échelle et de montrer leurs importances dans la production des occlusives du français et des traits articulatoires qui les caractérisent. La première partie de ce mémoire se focalisera sur la présentation des notions clés du Traitement Automatique des Langues et de la phonétique. La deuxième partie sera dédiée à la description du corpus utilisé. Puis, nous présenterons dans une troisième partie les méthodes utilisées pour traiter les données. Enfin, dans la quatrième partie, nous présenterons les différents descripteurs et nous nous arrêterons sur les résultats obtenus.

## 2. État de l'art

### 2.1 Partie Phonétique

La production de la parole est un processus complexe constitué d'une succession d'étapes (cf. figure 1) qui ont été définies par (Catford, 1977) comme suit :

1. Programmation neurolinguistique : sélection, séquençage et synchronisme des étapes qui vont suivre.
2. Phase neuromusculaire : transmission des impulsions nerveuses sortantes (efférentes), mise à feu des unités motrices et contraction des muscles individuels.
3. Phase articulatoire : postures et mouvements de tous les organes.
4. Phase aérodynamique : dilatation, compression et circulation de l'air dans et à travers le conduit vocal.
5. Phase acoustique : propagation d'ondes sonores à partir du conduit vocal.
6. Phase neuro-réceptive : stimulation auditive périphérique et transmission des impulsions nerveuses entrantes (afférentes).
7. Identification neurolinguistique : identification potentielle ou réelle des signaux entrants en tant que sons de parole spécifiques.

Selon (Catford, 1977), en termes d'accessibilité et d'utilité taxonomique, il n'y a en réalité que trois phases dans le processus de la parole qui peuvent faire l'objet d'une attention sérieuse : La phase articulatoire, la phase aérodynamique et la phase acoustique. Nous étudierons dans ce mémoire ces 3 phases.

## les phases de la production de la parole

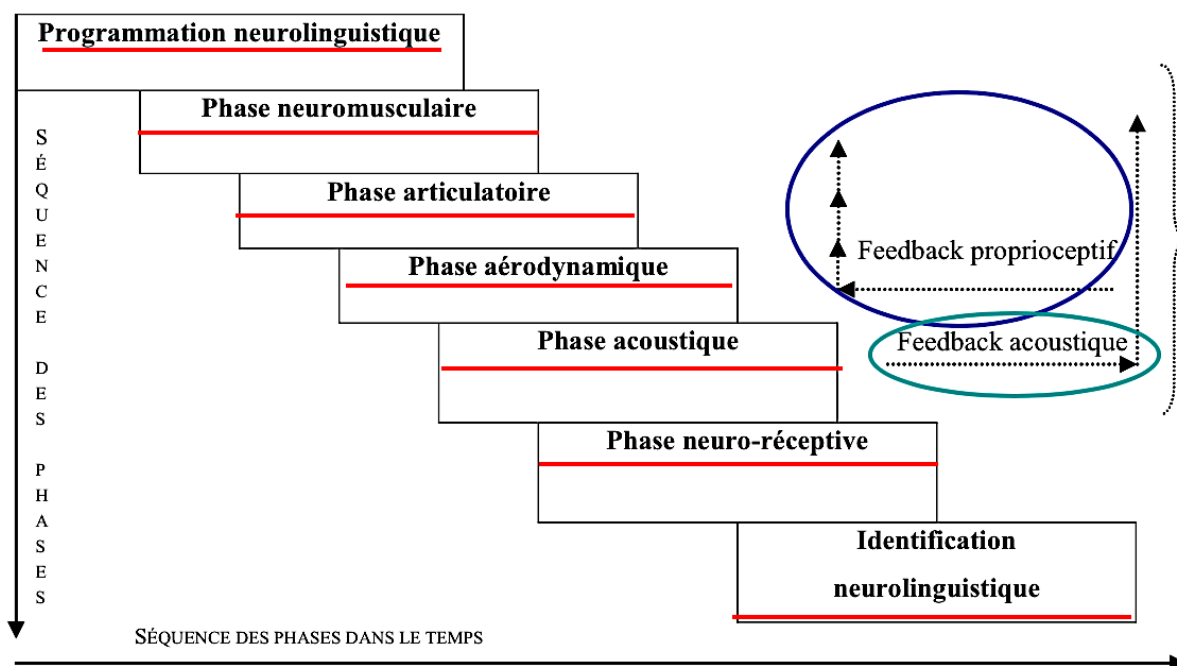


Figure 1 : Schéma représentant les différentes phases de la production de la parole introduites par (Catford, 1977).

### 2.1.1 Phase aérodynamique

La production de la parole implique, comme nous l'avons vu précédemment, une succession de phases complexes et hautement coordonnées qui incluent des phénomènes neuromusculaires, de mouvements et de variations aérodynamiques. Ces événements commencent avec la respiration et se terminent avec la génération d'un signal de parole acoustique. Comme (Warren, 1982) l'a déclaré, « les changements de forme et de taille des voies respiratoires inférieures produisent un flux d'air continuellement modifié par les structures laryngées et supra-laryngée ». La phase aérodynamique constitue le lien entre l'activité corporelle du locuteur permettant ces changements de formes (dans la phase articulaire) et les ondes sonores résultantes (dans la phase acoustique). C'est au stade de

la phase aérodynamique que les sons de la parole sont générés. Les méthodes d'évaluation aérodynamique donnent donc des informations très riches sur les mouvements des organes articulatoires du conduit vocal. Les paramètres aérodynamiques sont le volume d'air (mesure de la quantité d'air), le débit (le taux de changement de volume), la pression (force par unité de surface) et la résistance (une variable qui relie le débit et la pression). Les diverses pressions, volumes, flux et résistances atmosphériques associées à la respiration, à la phonation et à l'articulation peuvent être mesurées et documentées de manière quantitative donnant une description chronologique des gestes articulatoires. Selon (Baken and Orlikoff, 2000), « les sons de la parole sont le produit d'une utilisation précise de la pression atmosphérique générée par le système respiratoire. Il est souvent utile de connaître la pression de l'air dans une région donnée de l'appareil vocal (en particulier par rapport à la pression à un autre endroit) et, plus important encore, d'observer les changements dans les valeurs de pression atmosphérique qui résultent de l'activité vocale. Ces observations, associées à une compréhension de la structure et du fonctionnement du conduit vocal, permettent au clinicien de déduire beaucoup de choses sur la nature et le degré d'anomalie de la parole ».

Ces données quantitatives sont donc indispensables dans l'étude des phénomènes consonantiques complexes. Nous analyserons, dans cette étude, les pressions sous-glottique et intra-orale.

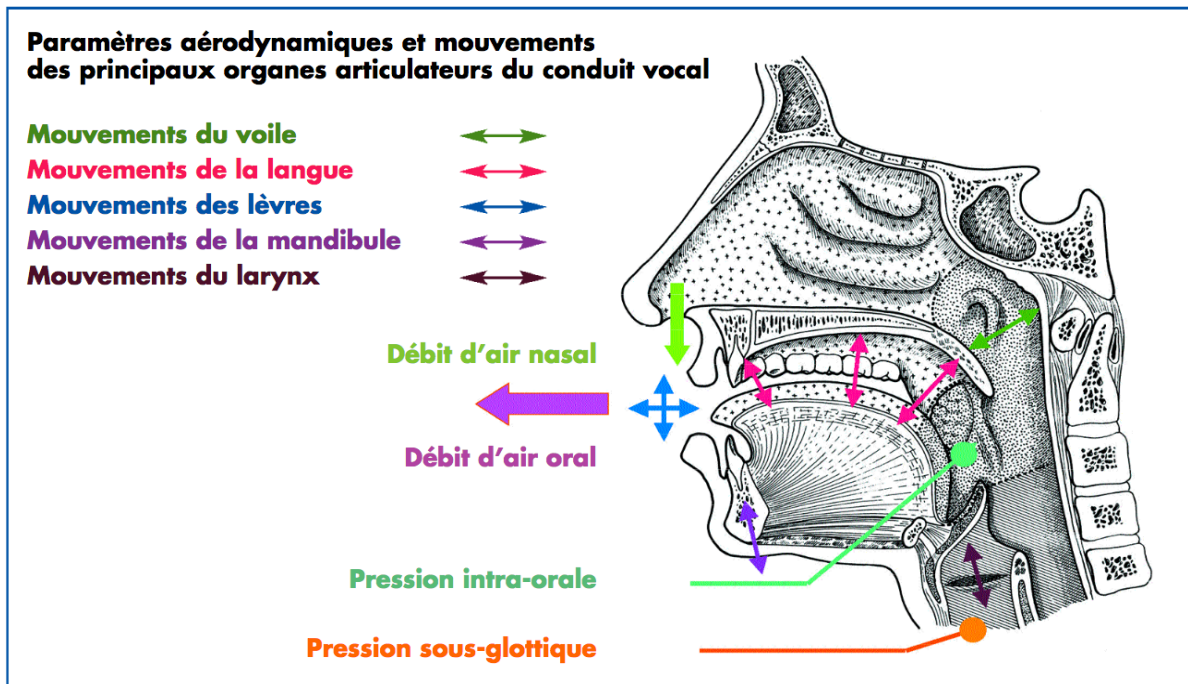


Figure 2 : Dynamique et lieux de constriction des principaux organes articulateur du conduit vocal et situation anatomique des paramètres aérodynamiques (Teston, 2007b).

### 2.1.2 Pression intra-orale

La pression intra-orale (PIO), ou pression supra glottique, correspond à la pression qui s'exerce à l'intérieur de la cavité buccale. Elle s'exprime en hectopascal (hPa). Selon (Baken and Orlikoff, 2000), « les pressions d'air intra-orales qui accompagnent la parole donnent un aperçu de la fonction de l'ensemble du système de la parole. Des valeurs de pressions anormales peuvent refléter une assistance ventilatoire inadéquate ou instable. Mais elles peuvent également indiquer un statut vélopharyngé inapproprié ou une anomalie du degré, de l'emplacement ou du moment de la constriction de l'appareil vocal ». Il s'agit d'un paramètre extrêmement robuste pour comprendre les modèles de sons dans la langue car dès qu'il y a une altération des constriction ou un mouvement d'un articulateur, l'information est encodée dans la pression intra-orale. En effet, selon (Teston, 2007b) « la pression intra-orale est un

indicateur de l'ensemble des états des constrictions du conduit vocal et à ce titre joue un rôle fondamental dans leur contrôle ». C'est pour cette raison que nous avons jugé pertinent d'exploiter les mesures de pression intra-orale pour réaliser la classification automatique des lieux d'articulation.

### **2.1.3 Pression sous-glottique**

Lors de la production de la parole, l'air circule dans le conduit vocal. En effet, au cours de l'inspiration, les poumons se dilatent, entraînant un écoulement d'air de la bouche aux poumons avec la glotte relativement ouverte. Pendant l'expiration, les poumons se contractent, poussant l'air des poumons vers la bouche ce qui exerce une pression sur les plis vocaux qui vont avoir tendance à se fermer, s'opposant ainsi au passage de l'air. La pression sous-glottique correspond à cette pression créée par le flux d'air expiré contre les plis vocaux. Lorsque cette pression est supérieure à la résistance laryngée, les plis vocaux s'ouvrent et laissent passer l'air ce qui entraîne une chute de la pression sous-glottique. Plus précisément, la pression sous-glottique représente « l'énergie immédiatement disponible pour la création des signaux acoustiques de la parole » (Baken and Orlikoff, 2000). Une pression sous-glottique plus ou moins constante doit être maintenue pour la puissance de parole car « des niveaux inappropriés de pression sous-glottique ou une régulation de pression inadéquate peuvent causer des niveaux anormaux d'intensité de la parole ou des changements soudains de la fréquence fondamentale » (Baken and Orlikoff, 2000).

### 2.1.4 Phase acoustique

La phase acoustique a pour objet d'étude la propagation d'ondes sonores telles que celles produites par les organes de phonation. En effet, lors de la production de la parole, nous produisons un signal sonore : une onde acoustique continue et complexe, composée de nombreuses fréquences qui changent au cours du temps. La phonétique acoustique étudie ce signal en le transformant, dans un premier temps, en signal électrique grâce au transducteur approprié : le microphone (lui-même associé à un préamplificateur). Le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements statistiques qui visent à mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perceptuelle : hauteur, intensité, durée et timbre.

L'analyse de la phase acoustique est essentielle car elle aide à estimer le produit de la phase aérodynamique et, plus important encore, elle constitue l'une des principales sources d'informations ou d'inférences sur les phases réceptives et perceptuelles suivantes. De plus, l'étude de la phonétique acoustique a d'importantes applications techniques en ce qui concerne le développement d'équipements de transmission, de reconnaissance et de synthèse de la parole. En effet, les progrès techniques ont fait de la phase acoustique l'une des plus accessibles pour l'observation (instrumentale).



## 2.2 Partie Traitement Automatique des Langues

« Une somme de connaissances en constante accumulation est aujourd'hui éparpillée de par le monde. Elle suffirait probablement à résoudre toutes les grandes difficultés de notre temps, si elle n'était dispersée et inorganisée. » (Wells, 1940)

Dans les années 1990, le World Wide Web fait son apparition dans les foyers permettant au grand public de consulter un ensemble de pages mélangeant du texte, des liens et des images. On assiste également à la prolifération des ordinateurs portables, au développement des premiers logiciels de navigation et à la création d'équipements numériques permettant l'acquisition, le traitement, le stockage et la diffusion d'un grand nombre de données. Ces avancées technologiques vont provoquer un changement de paradigme dans le traitement des données. Selon (Sébillot, 2014), les données à traiter « se déclinent à l'aune des 3 V (variété, volume, vitesse) » suscitant l'émergence de nouveaux champs de recherche afin d'exploiter et d'utiliser cet ensemble important de données réparties sur le web. On assiste à « un passage du rationalisme vers l'empirisme, avec le 'remplacement' des approches symboliques, à base de règles fondées sur l'expertise humaine, par des approches empiriques fondées sur les données, où la connaissance est extraite par des techniques d'apprentissage automatique, en particulier statistique » (Sébillot, 2014). Ce domaine, consistant à extraire des connaissances à partir de données brutes, est connu sous le terme « Extraction de Connaissances à partir de Données (ECD) » ou « Knowledge Discovery in Databases (KDD) » en anglais. Il se situe à l'intersection de disciplines diverses telles que les statistiques et les technologies de l'information (bases de données, apprentissage automatique, intelligence artificielle).

C'est un processus complexe décomposé en une suite d'opérations distinctes que nous appliquerons au cours de cette étude (Lefébure et Ventari, 2001)(Preux, 2011) : (1) définir la problématique en cernant les objectifs, le contexte et les besoins ; (2) collecter des informations et les organiser dans une base de données ; (3) sélectionner les données les plus pertinentes pour répondre à la problématique ; (4) nettoyer les données (bruit, normalisation, inexactitudes, erreurs) ; (5) transformer les données dans un format qui les prépare à la fouille de données ; (6) fouille de données, application d'une ou plusieurs techniques (réseaux de neurones, classification, réseaux bayésiens, arbres de décision, etc.) pour extraire des patterns ; (7) visualisation des données : histogramme, camembert, arbre, visualisation 3D ; (8) évaluer les résultats : mesurer l'intérêt des patterns générés de manière qualitative (facilité de compréhension du résultat) et quantitative (tel que l'intervalle de confiance, etc.).

La fouille de données est l'étape principale de ce processus qui est définie par (Seidman, 2001) comme étant "the process of discovering meaningful patterns and relationships that lie hidden within very large databases". Elle regroupe « l'ensemble des techniques, soit descriptives (qui visent à mettre en évidence des informations présentes mais cachées par le volume des données), soit prédictives (cherchant à extrapoler de nouvelles connaissances à partir des informations présentes dans les données) » (Elghazel, 2007a). Les données traitées lors de cette étape peuvent être de natures diverses : corpus en langage naturel, images, son ou vidéo. Nous analyserons dans ce mémoire des données audio, il s'agira donc de réaliser de la fouille audio définie par (Chandra and Senthildevi , 2015) comme étant « une branche de l'exploration de données utilisée pour rechercher et analyser automatiquement le contenu du signal audio ». Ces données ont connu le même

accroissement que les données textuelles grâce aux avancées technologiques, ce qui a permis d'augmenter la taille des corpus et de contribuer à améliorer l'étude de la parole en observant des phénomènes à plus grande échelle. Cependant, cet accroissement a eu la conséquence directe de rendre les méthodes utilisées jusqu'alors pour constituer, segmenter, annoter et classifier les corpus trop coûteuses et, dans de nombreux cas, tout à fait impossible. De nombreuses méthodes automatiques ou semi-automatiques ont donc vu le jour afin d'être en mesure d'exploiter cette masse de données. Nous en exploiterons deux dans ce mémoire : la segmentation et la classification automatique.

### **2.2.1 Segmentation automatique**

Afin d'accéder aux informations contenues dans les corpus vocaux, il est nécessaire de procéder à une segmentation et un à étiquetage phonémique. L'opération d'étiquetage « vise à projeter sur le signal de parole une séquence d'étiquettes ou de labels se rattachant à un ou plusieurs niveaux d'analyse » (Nguyen et Adda-Decker, 2013). Dans le cas de notre étude, le niveau d'analyse sera celui du phonème ; nous synchroniserons une séquence de phonèmes avec le son. Cet étiquetage s'accompagne généralement d'une segmentation du signal de parole qui a pour objectif « d'associer à chaque étiquette un intervalle temporel dans ce signal » (Nguyen et Adda-Decker, 2013). Ces deux procédés sont nécessaires tant dans la recherche linguistique que dans les technologies de la parole car une des caractéristiques de ces corpus est que le son ne peut être étudié seul. En effet, une transcription orthographique et phonétique des fichiers audio est nécessaire car l'analyse prosodique, articulatoire et phonatoire, nécessite de connaître la position précise des limites temporelles phonétiques.

Ces deux tâches peuvent être effectuées manuellement mais c'est une procédure fastidieuse qui ne peut pas être une solution envisageable pour un grand nombre de données au regard du coût important du processus. Il a été calculé par (Bazillon et al., 2008a) que neuf heures sont requises pour effectuer l'annotation manuelle d'une heure de parole spontanée. Nous privilégierons donc, pour notre étude, les outils existants permettant de procéder à ces deux opérations de manière automatique ou semi-automatique.

Nous utiliserons pour cela le système automatique MAUS (Munich AUtomatic Segmentation) qui est un logiciel d'alignement forcé de transcriptions orthographiques ou phonétiques avec un signal de parole développé par Florian Schiel à l'institut de phonétique et de traitement de la parole de la LMU à Munich (Schiel, 1999). L'alignement forcé fonctionne de la manière suivante : l'utilisateur doit donner au système deux fichiers d'entrée : un fichier multimédia contenant un signal de parole enregistré et un fichier texte contenant les mots prononcés dans l'enregistrement. Ensuite, une transcription canonique est créée en SAMPA (Speech Assessment Methods Phonetic Alphabet) pour chaque mot avec l'outil Balloon (Reichel, 2012). Sur la base de cette forme phonologique, un graphique est pondéré à l'aide d'outils statistiques analysant toutes les réalisations possibles (variantes de prononciation). Enfin, un système de reconnaissance de la parole basé sur un « modèle de Markov caché » choisit la variante de prononciation la plus probable pour chaque mot et place les limites de segment (Schiel, 1999). Le résultat de ce processus est un alignement orthographique et phonétique (segmentation et étiquetage) de la parole enregistrée, qui est ensuite restitué dans le format cible souhaité (BPF, Emu, TextGrid) et renvoyé à l'utilisateur via le navigateur Web.

## 2.2.2 Classification automatique

La classification automatique fait partie des méthodes statistiques abondamment utilisées dans la fouille de données. Son objectif principal est de regrouper des objets en leur associant une classe afin de structurer un ensemble de données et d'inférer des connaissances. Dans notre cas, les objets sont des signaux ou des segments de signaux audio qu'il s'agit d'assigner à des classes. Les classes doivent être le plus homogènes possibles de sorte que des objets d'une même classe se ressemblent autant que possible et, qu'au contraire, les objets appartenant à des classes différentes en soient dissemblables. En d'autres termes, « on cherche des classes qui, d'une part, forment chacune un tout cohérent (compacité) et, d'autre part, sont distinctes les unes des autres (séparabilité) » (Le Roux, 2014). L'apprentissage automatique permet de construire automatiquement cette procédure de classification à l'aide de deux étapes :

- Une étape d'apprentissage qui se base sur des exemples, c'est-à-dire sur un échantillon dit « d'apprentissage » constitué d'un ensemble limité de données disponibles.
- Une étape de test par laquelle les performances du système de classification sont évaluées en utilisant de nouvelles données.

Si les classes possibles ont été définies à l'avance par l'expert humain, on parle d'apprentissage supervisé. La procédure est la suivante : on fournit au programme d'apprentissage des données d'entrées étiquetées avec les sorties souhaitées. Ainsi, l'algorithme va pouvoir « apprendre » en comparant les sorties qu'il produit avec les sorties d'entraînement pour trouver les erreurs et les points d'amélioration et ainsi modifier le modèle

en conséquence. A contrario, si les classes n'ont pas été prédéfinies par l'expert humain, il s'agit d'un apprentissage non supervisé, c'est-à-dire que l'apprentissage par la machine se fait indépendamment de l'expert humain qui ne lui fournit pas d'exemple de résultats attendus. Nous utiliserons, dans ce mémoire, la méthode d'apprentissage supervisé. Dans un premier temps, nous allons extraire des signaux audio les descripteurs acoustiques sous forme d'attributs numériques. Puis, nous allons sélectionner, parmi ces descripteurs, ceux qui sont pertinents pour notre étude et qui garantissent les meilleurs résultats de classification. Pour finir, nous réaliserons l'apprentissage supervisé sur les attributs précédemment sélectionnés pour ensuite réaliser la classification automatique des lieux d'articulation sur de nouveaux exemples (de test).

Afin de procéder à la classification automatique, nous utiliserons un logiciel libre d'apprentissage et d'Extraction de Connaissances à partir de Données nommé WEKA (Waikato Environment for Knowledge Analysis ). Il a été développé à l'Université de Waikato en Nouvelle Zélande et il est écrit en Java et distribué sous les termes de la GNU General PublicLicense (Eibe et al., 2016). WEKA fournit des implémentations d'algorithmes d'apprentissage ainsi qu'une variété d'outils pour transformer les ensembles de données, tels que les algorithmes de discrétisation. Il permet de prétraiter un ensemble de données, de les intégrer dans un schéma d'apprentissage et d'analyser les résultats de l'algorithme de classification et ses performances (Eibe et al., 2016). On utilisera ici l'interface utilisateur graphique appelée TheExplorer.

## 3. Corpus de travail

### 3.1 Sujets et matériels

Les données acoustiques ont été collectées auprès de deux femmes et deux hommes ayant respectivement un larynx normal et aucun problème de voix. Les enregistrements ont eu lieu en 1996 à l'hôpital Erasme de l'Université libre de Bruxelles. La procédure a préservé les droits et le bien-être des sujets de recherche humains, conformément aux normes du comité d'éthique (Comité, 2017a). La pression sous-glottique (Psg), la pression intra-orale (PIO), le débit d'air buccal (dab), l'intensité (Int) et le signal de parole ont été enregistrés simultanément au moyen de la procédure suivante :

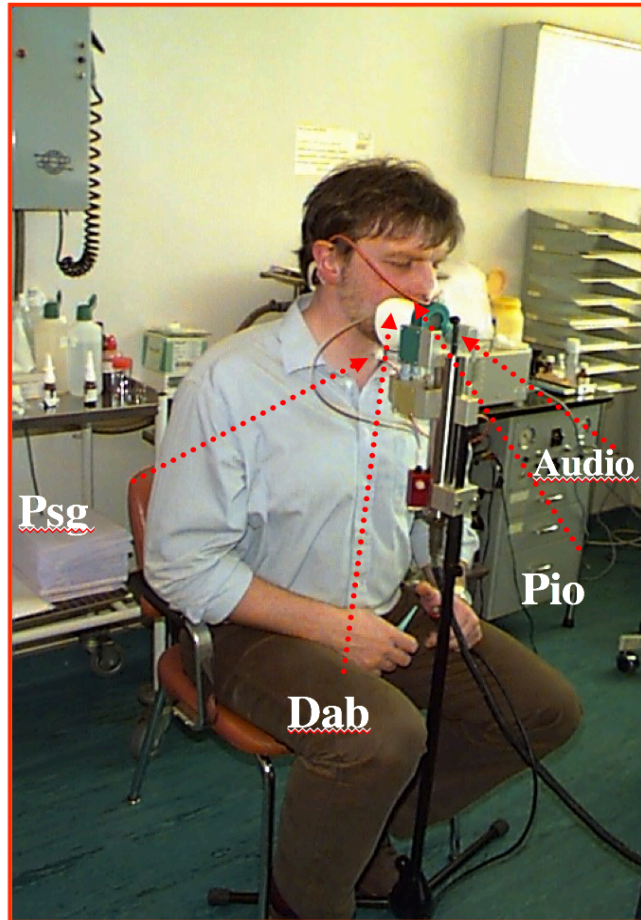
- La pression sous-glottique a été mesurée par ponction trachéale à l'aide d'une aiguille insérée sous le cartilage cricoïde. La mise en place de l'aiguille a été effectuée après l'administration d'une injection de 2% de xylocaïne pour l'anesthésie locale. Un tube en plastique attaché à un transducteur de pression a été connecté à l'aiguille. Le principal avantage de la technique de ponction trachéale selon (Baken and Orlikoff, 2000) est son caractère direct. Contrairement à d'autres méthodes qui reposent sur l'observation d'un corrélat de pression sous-glottique, cette procédure ne nécessite aucun étalonnage particulier, à l'exception du capteur de pression lui-même. Tous les autres facteurs étant égaux, cela donne une mesure intrinsèquement exacte.
- Les mesures de pression intra-orale ont été obtenues à l'aide d'un petit tube inséré dans la cavité nasale et placé derrière le vélum dans l'oropharynx. Les données ont été recueillies avec un masque de silicone souple placé sur la bouche de l'orateur. Le

signal acoustique a été enregistré avec un microphone directionnel AKG C 419 placé à une distance stable (4 cm) du masque de silicone. L'avantage de la mise en place du tube nasal est qu'il n'y a aucune interférence avec le mouvement des articulatoires, ou un risque que les mouvements de la langue ou des lèvres ne déplacent le tube. Selon (Baken and Orlikoff, 2000), l'avantage de cette prise de mesure est que « le mouvement vélaire ne semble pas compromettre la mesure de manière significative, pas plus que la fermeture du vélopharynx, à condition que le tube soit suffisamment fin ».

- Le débit d'air buccal a été mesuré avec un masque flexible en silicone.
- L'intensité a été mesurée à partir du signal enregistré par un micro haute-fidélité placé sur le bloc des transducteurs. La bouche est restée à une distance constante de 4 cm.

Les tubes de pressions sous-glottique et intra-orale, le masque de silicone et le microphone ont été connectés au poste de travail Physiologia (cf. figure 3), un système d'acquisition de données multiparamétriques permettant l'enregistrement simultané du signal de parole et de divers paramètres aérodynamiques (Galindo and Teston, 1989). Les appareils de mesures utilisés dans cette étude ont une précision au millibar près pour les mesures de pression. Les valeurs de pression sont données en hPa (hectopascal). Le Pascal (Pa) est l'unité internationale pour les mesures de pression en physique. L'utilisation hPa donne également une correspondance facile au cm H<sub>2</sub>O qui était traditionnellement utilisé dans les sciences de la parole, 1 cm H<sub>2</sub>O = 1,2 hPa.





*Figure 3 : Dispositif EVA 2 et position des sujets pour l'enregistrement des données acoustiques (fréquence fondamentale et intensité) et aérodynamiques (pressions sous-glottique et intra-orale, débit oral d'air buccal). Station de travail « Physiologia ».*

### **3.2 Base de données**

Dans le cadre de cette étude, nous avons contribué au développement d'une base de données pour la mise à disposition des données aérodynamiques à la communauté scientifique. Cette base de données est un apport important car les données aérodynamiques sont extrêmement difficiles à acquérir notamment les mesures de pression sous-glottique prise par ponction trachéale qui nécessite un encadrement hospitalier et qui pose des problèmes d'éthique de par sa nature invasive. La base de données a été développée au Laboratoire de

Phonétique et Phonologie de l'Université Sorbonne Nouvelle, Paris 3 (Demolin et al., 2019). Elle a été conçue avec le langage SQL servant à exploiter des bases de données relationnelles et le site de mise à disposition a été conçu en PHP. Toutes les données présentées dans ce mémoire sont accessibles et téléchargeables via le site de la base de données.

The screenshot shows a web interface with a navigation bar (HOME, BASKET, ABOUT, ADMIN) and filter sections for 'table record', 'table production', and 'table speaker'. Below the filters is a table with the following data:

Filename	Type of production	description	Date of the recording	Aerodynamics files	transcription of the production	gender of the speaker	code of the speaker	Age of the speaker	Language of the speaker
baba.wav	Logatome		1999-06-24 00:00:00	baba.dab baba.dan baba.int baba.par baba.pio baba.psg	1 baba 2 baba 3 baba 4 baba 5 baba	F F F F F	F1 F1 F1 F1 F1	22 22 22 22 22	French French French French French

Figure 4 : Page d'accueil du site de la base données aérodynamiques développé au Laboratoire de Phonétique et Phonologie à l'Université Sorbonne Nouvelle, Paris 3.

### 3.3 Procédures

Les sujets ont été enregistrés en prononçant une série de logatomes (enchaînements syllabiques vides de sens) combinant chacune des consonnes occlusives du français (cf. tableau 1) avec les voyelles /a, i, u/ (ex: papa, kiki, nuu). Ces logatomes ont été insérés dans la phrase « [C1VC2V] dit [C1VC2V] encore », C1 et C2 correspondant aux occurrences de la consonne et V à celles de la voyelle. Nous avons choisi de porter notre étude uniquement sur les occlusives en position C2 dans chaque logatomes car les frontières sont plus facilement détectables. La phrase devait être répétée cinq fois consécutives par les sujets pour chacune des occlusives dans les trois contextes intervocaliques. Plusieurs répétitions de chaque

phonème sont nécessaires selon (Baken and Orlikoff, 2000) car « la pression moyenne et sa variabilité dans la forme du coefficient de déviation standard ou de ses variations sont toutes deux importantes ».

*Tableau 1 : Consonnes occlusives du français.*

Point d'articulation	Bilabiales	Alvéolaires	Vélares
Mode d'articulation			
Occlusives sourdes	p	t	k
Occlusives sonores	b	d	g
Nasales	m	n	

## 4. Méthodes

### 4.1 Segmentation manuelle

Les données ont d'abord été traitées manuellement à l'aide du logiciel Phonedit Signaux, un logiciel d'analyse de signaux sonore, aérodynamique, articulatoire et électrophysiologique développé par le laboratoire Parole et Langage d'Aix-en-Provence (Ghio, 2002). Ce logiciel a été choisi car il fournit un environnement complet pour l'enregistrement, la lecture, l'affichage, l'analyse et l'annotation de données multiparamétriques. Nous avons, dans un premier temps, procédé à une segmentation et à un étiquetage manuel des occlusives en position C2 sur les données de deux sujets (un homme et une femme). Nous avons mesuré

les paramètres suivants : (1) durée totale de la consonne, (2) frontières de début, milieu et fin de la consonne pour les mesures de pression sous-glottique (Psg), pression intra-orale (PIO), débit d'air buccal (dab) et d'intensité (Int).

Afin de marquer précisément les frontières de début et de fin, nous nous sommes basés sur les critères suivants pour les consonnes sonores : (1) La mise en place des constriction occasionne un blocage complet de l'écoulement d'air au niveau de la bouche, du pharynx et de la glotte ce qui provoque une accumulation de l'air derrière ces parois fermées et peut occasionner une augmentation graduelle de la pression sous-glottique ainsi que de la pression intra-orale et simultanément une variation abrupte au niveau de la courbe d'intensité qui chute. (2) A l'inverse, au moment de la phase d'explosion qui correspond au moment de la prononciation réelle de la consonne (la sortie acoustique), on observe une chute de la pression intra-orale et sous-glottique due à l'air qui s'échappe, ce qui provoque une augmentation simultanée de l'intensité. La durée de l'occlusion a donc été définie dans l'intervalle temporel entre le début de l'occlusion et le début du relâchement, tous deux repérés sur la courbe d'intensité (cf. figure 5). Pour les consonnes sourdes, nous avons réalisé deux types de segmentation : une première se basant sur les mêmes critères que pour les consonnes sonores, excluant le délai d'établissement du voisement (VOT), et une deuxième incluant le VOT afin d'obtenir des mesures au moment de l'explosion mais également à la fin effective de la consonne.

À la fin de la segmentation manuelle, nous avons obtenu 30 segments pour chaque occlusive (trois contextes combinés) représentant 180 segments au total pour chacun des

sujets. Les valeurs correspondant aux frontières de début, milieu et de fin ont été extraites pour les mesures de pression sous-glottique (Psg) et de pression intra-orale (PIO).

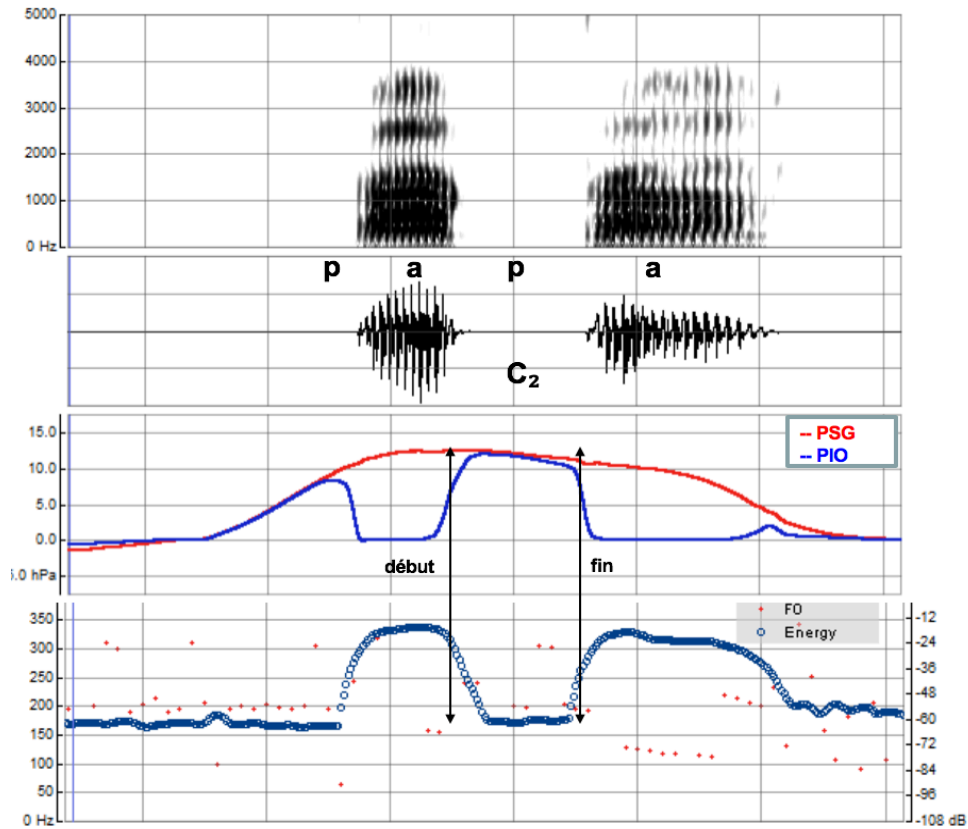


Figure 5 : Spectrogramme, signal acoustique, pression intra-orale (PIO), pression sous-glottique (Psg) et intensité pour le sujet masculin disant 'papa'. Les flèches indiquent le moment où les valeurs ont été prises.

L'étiquetage ainsi que la segmentation manuelle de ces 360 segments ont constitué une tâche extrêmement fastidieuse et chronophage mais essentielle afin d'obtenir un corpus de référence pour évaluer les performances du système de segmentation automatique.

## 4.2 Segmentation automatique

Afin de réaliser la segmentation automatique avec le système MAUS, nous avons utilisé le module pipeline sur l'interface en ligne en choisissant la procédure 'CHUNKPREP-G2P-MAUS'. Afin d'utiliser ce module, il faut avoir réalisé au préalable une segmentation et un étiquetage en phrases, paragraphes ou groupes de mots sur un logiciel tel que Praat. Cette procédure n'est pas la plus ergonomique car il existe des modules qui prennent en entrée uniquement un fichier texte avec la transcription orthographique des fichiers audio mais les résultats sont nettement moins précis. Une fois ces fichiers constitués, nous les fournissons au programme accompagnés des fichiers audio correspondant afin qu'ils soient convertis en fichiers BPF adaptés à une entrée MAUS. Le fichier BPF résultant contient trois niveaux : (1) ORT correspondant à la transcription orthographique, (2) KAN correspondant à la transcription phonologique et (3) TRN correspondant aux groupes de mots fournis dans le fichier d'entrée de base. Les fichiers BPF sont ensuite exploités par le module MAUS qui procède à l'alignement orthographique et phonétique (segmentation et étiquetage) de la parole enregistrée, qui est ensuite restituée dans le format cible souhaité qui est le TextGrid dans le cas de notre étude (cf. figure 6). La totalité de ce processus, conversion et alignement, a pris moins de deux minutes pour 360 segments.

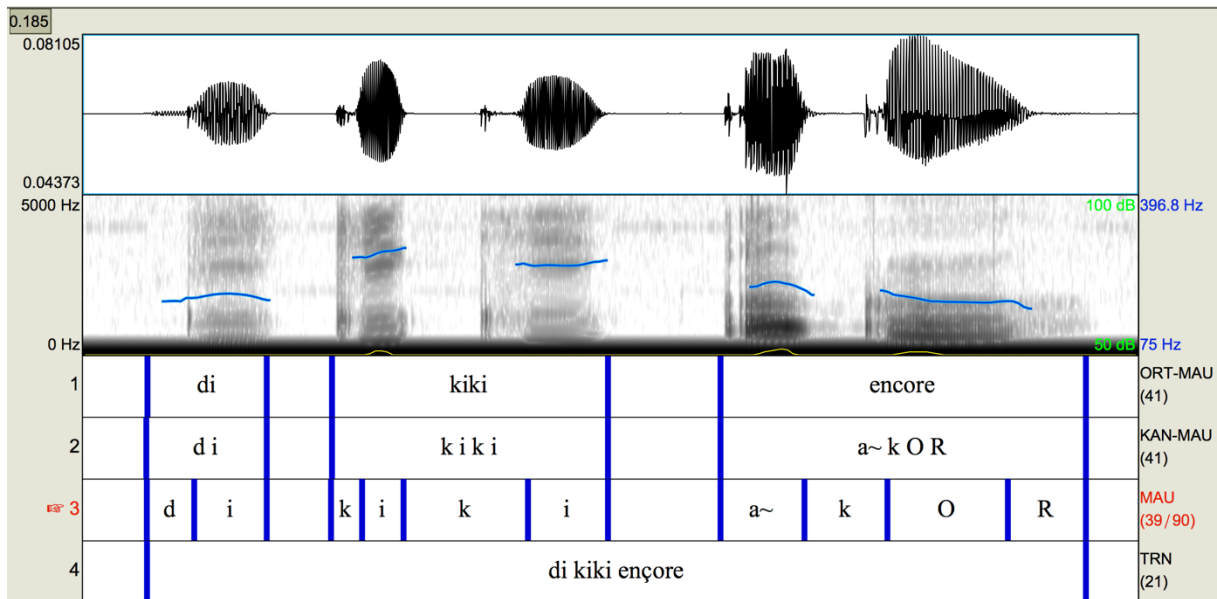


Figure 6 : Résultat de l'alignement automatique effectué avec MAUS.

### 4.3. Comparaison entre segmentation manuelle et segmentation automatique

Avant de procéder à la segmentation automatique sur l'ensemble des données, nous avons voulu évaluer les performances du système d'alignement automatique afin de savoir si il était pertinent de l'utiliser dans le cadre de notre étude ou s'il générerait trop d'erreurs. Afin de réaliser cette tâche, il n'existe pas de méthode spécifique. Dans le cadre de notre étude, nous avons choisi de calculer le pourcentage de décalage dans l'attribution des frontières de début et de fin entre une segmentation manuelle de référence et la segmentation automatique obtenue par MAUS. Nous avons conscience que les segmentations manuelles sont soumises aux jugements individuels des annotateurs humains et diffèrent donc toujours dans une certaine mesure, sans toutefois que l'une soit plus légitime qu'une autre d'être la segmentation de référence. Assurément, les décalages par rapport à une segmentation manuelle de référence unique ne peuvent donc pas nécessairement être considérés comme des erreurs dans la segmentation évaluée. Si nous voulions précisément évaluer le système MAUS, il

faudrait comparer plusieurs segmentations manuelles et combiner plusieurs critères de comparaison mais ce n'est pas le propos de notre étude. Nous voulons dans cette partie, quantifier le décalage par rapport aux critères que nous avons fixés lors de la segmentation manuelle car ils s'appuient sur des paramètres aérodynamiques robustes afin de détecter les phases d'occlusion et d'explosion des segments consonantiques.

Afin de réaliser cette comparaison, nous avons commencé par transformer les fichiers issus de la segmentation manuelle sur Phonedit au format TextGrid car ils sont à l'origine au format Mrk. Or, ces deux formats ne sont pas encodés de la même manière (cf. figure 7), nous avons donc écrit un script python (cf. annexe D.1) permettant de récupérer les étiquettes des segments consonantiques dans les fichiers Mrk et de générer un fichier en sortie ayant exactement la même structure qu'un fichier TextGrid.

<pre> item [3]:   class = "IntervalTier"   name = "MAU"   xmin = 0   xmax = 14.41792   intervals: size = 83   intervals [1]:     xmin = 0     xmax = 1.53544     text = "&lt;p:&gt;"   intervals [2]:     xmin = 1.53544     xmax = 1.62544     text = "b"   intervals [3]:     xmin = 1.62544     xmax = 1.7591431104343325     text = "a"   intervals [4]:     xmin = 1.7591431104343325     xmax = 1.8244878481374251     text = "b" </pre>	<pre> 1 [DSC_LEVEL_AA] 2 DSC_LEVEL_NAME=LEVEL_LBL_LEVEL_AA 3 DSC_LEVEL_CREATION_DATE=2018/02/23 17:20:47 4 DSC_LEVEL_LASTMODIF_DATE=2018/02/23 17:20:47 5 DSC_LEVEL_SOFTWARE=Phonedit Application 4.2.0.8 6 [LBL_LEVEL_AA] 7 LBL_LEVEL_AA_000000="calistri-aba-a02" 1759,986261 1825,951030 8 LBL_LEVEL_AA_000001="calistri-aba-a04" 3079,984453 3136,566341 9 LBL_LEVEL_AA_000002="calistri-aba-b02" 4260,106264 4320,084111 10 LBL_LEVEL_AA_000003="calistri-aba-b04" 5629,915579 5679,971041 11 LBL_LEVEL_AA_000004="calistri-aba-c02" 6820,142141 6880,141879 12 LBL_LEVEL_AA_000005="calistri-aba-c04" 8190,165396 8259,954353 13 LBL_LEVEL_AA_000006="calistri-aba-d02" 9349,967110 9408,074782 14 LBL_LEVEL_AA_000007="calistri-aba-d04" 10666,015314 10727,542424 15 LBL_LEVEL_AA_000008="calistri-aba-e02" 11869,894540 11928,776903 16 LBL_LEVEL_AA_000009="calistri-aba-e04" 13150,009844 13209,793075 17 [LINK_LEVEL_AA] 18 LBL_LEVEL_AA_000000="calistri-aba-a02" TS_000000 TS_000001 19 LBL_LEVEL_AA_000001="calistri-aba-a04" TS_000002 TS_000003 20 LBL_LEVEL_AA_000002="calistri-aba-b02" TS_000004 TS_000005 21 LBL_LEVEL_AA_000003="calistri-aba-b04" TS_000006 TS_000007 22 LBL_LEVEL_AA_000004="calistri-aba-c02" TS_000008 TS_000009 </pre>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 7 : Formats des fichiers d'étiquettes avec à gauche le fichier TextGrid et à droite le fichier Mrk.

Nous avons pu, grâce à cette transformation, fusionner les fichiers TextGrid obtenus automatiquement avec ceux de la segmentation de référence pour obtenir un unique fichier TextGrid (cf. figure 8) nous permettant de comparer les frontières. Nous avons ensuite



parcouru ce fichier et récupéré les intervalles correspondant aux segments consonantiques étudiés afin de calculer le décalage moyen des emplacements des frontières de début et de fin.

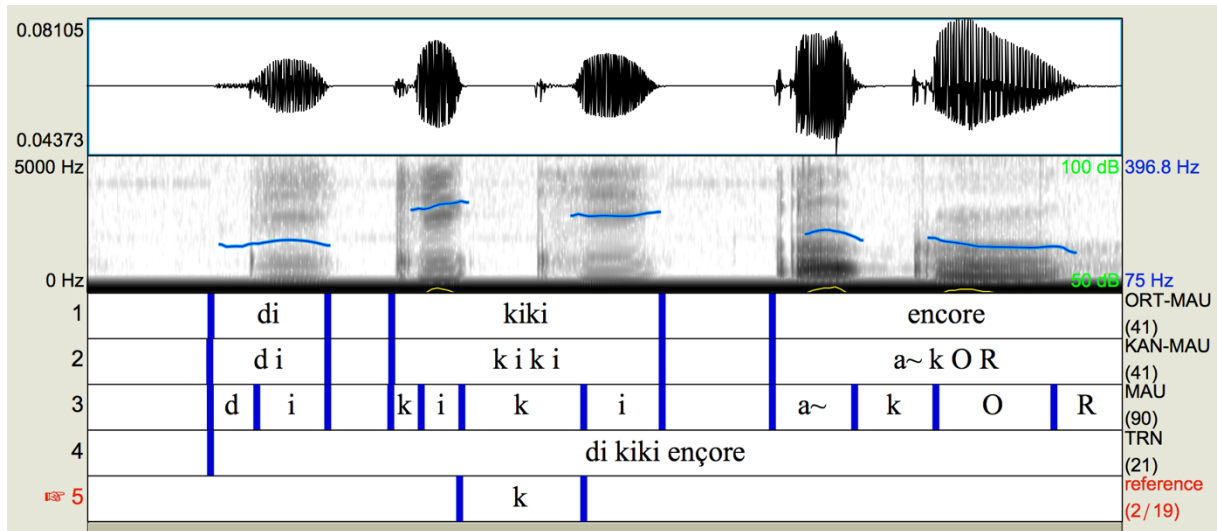


Figure 8 : TextGrid fusionné contenant la référence (reference) et le résultat de la segmentation automatique (MAU).

Afin de calculer le décalage moyen pour chaque consonne, nous avons calculé les paramètres suivants (cf. figure 9) :

1. Décalage début = intervalle de début du segment de référence - intervalle de début du segment à évaluer.
2. Décalage fin = intervalle de fin du segment de référence – intervalle de fin du segment à évaluer.
3. Décalage total = décalage début (valeurs absolues) + décalage fin (valeurs absolues).
4. Décalage moyen = moyenne du décalage total pour chaque consonne.

	debut_man	debut_auto	fin_man	fin_auto	decalage_total	consonne
0	1.240413	1.240272	1.349773	1.362419	0.012505	p
1	2.168733	2.168051	2.270022	2.284760	0.014056	p
2	3.850122	3.850458	3.960021	3.975402	0.015717	p
3	4.739693	4.740757	4.850033	4.861505	0.012536	p
4	6.269535	6.267209	6.360030	6.369776	0.007420	p
5	7.160097	7.161240	7.270081	7.281240	0.012302	p
6	8.620207	8.620511	8.730001	8.737063	0.007366	p
7	9.645466	9.645574	9.740153	9.760280	0.020235	p
8	11.099325	11.095189	11.199809	11.211742	0.007797	p
9	11.938707	11.937273	12.029789	12.041760	0.010536	p
10	13.740406	13.746300	13.830779	13.841976	0.017091	p
11	14.639848	14.638520	14.739383	14.741980	0.001269	p
12	16.122436	16.120690	16.209664	16.221033	0.009623	p
13	17.130662	17.133360	17.230239	17.237137	0.009596	p
14	18.680233	18.690120	18.790124	18.799521	0.019284	p
15	19.599547	19.606508	19.709592	19.716186	0.013555	p
16	21.209709	21.212840	21.320643	21.324890	0.007378	p
17	22.039990	22.041120	22.132763	22.143218	0.011585	p
18	23.720767	23.726504	23.820223	23.839292	0.024806	p
19	24.709951	24.709537	24.800268	24.810658	0.009975	p
20	26.581185	26.587379	26.650792	26.703279	0.058681	p
21	28.159937	28.156643	28.272104	28.292581	0.017183	p
22	29.649900	29.652040	29.750069	29.775189	0.027260	p
23	30.730009	30.723295	30.859593	30.863185	0.003121	p
24	32.340597	32.341872	32.469262	32.471771	0.003784	p
25	33.420402	33.420145	33.549789	33.566697	0.016651	p
26	34.840228	34.844948	34.940073	34.963189	0.027837	p
27	35.910130	35.911161	36.023248	36.040520	0.018302	p
28	37.340537	37.343056	37.480534	37.482946	0.004931	p
29	38.250036	38.255680	38.386607	38.386758	0.005795	p

Le décalage moyen pour le segment p est 0.0142726

*Figure 9 : Résultat du programme python calculant le décalage moyen pour chaque occurrence de la consonne « p » entre les intervalles de la segmentation de référence (debut\_ref et fin\_ref) par rapport à ceux obtenus à l'issue de la segmentation automatique.*

Pour être en mesure d'évaluer les résultats obtenus, nous avons cherché dans les études existantes le décalage moyen obtenu lors des tâches d'évaluation afin de fixer un seuil au-delà duquel nous estimons qu'il faut corriger manuellement les résultats de la segmentation automatique. (Pitt et al., 2005a) ont obtenu un décalage moyen dans l'attribution des frontières entre plusieurs annotateurs humains de 16ms. Il a également été observé dans (Hosom, 2008b) que l'accord entre deux experts humains était en moyenne de 93,78% lorsque le décalage moyen se situait à hauteur de 20ms. (Auran et al., 2003a) rapporte que l'alignement obtenu de manière automatique est fiable à près de 70 % pour un seuil d'acceptabilité de 20 ms et (Sjölander, 2003b) a obtenu une performance du système d'alignement automatique de 85% de frontières correctement détectées dans un intervalle inférieur à 20 ms au regard de

l'alignement manuel. Selon (Bigi et Meunier, 2018), un décalage moyen de 20 ms pour une segmentation automatique occasionne des variations très raisonnables. Nous avons donc choisi de fixer le seuil de décalage moyen à 20ms en considérant acceptables les segmentations ayant un décalage moyen inférieur à ce seuil et à l'inverse celles ayant un décalage supérieur seront corrigées manuellement.

Nous avons souhaité évaluer dans quelle mesure ces décalages étaient influencés par le sexe du locuteur (cf. figure 10), la consonne prononcée (cf. figure 11,12) ainsi que les différences de voisement (cf. figure 13).

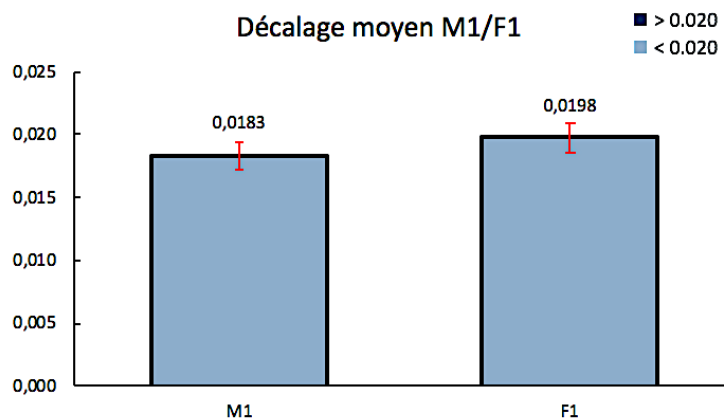


Figure 10 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence pour les données du sujet masculin à gauche et féminin à droite. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 180 (M1), N = 180 (F1)).

Nous pouvons constater que le décalage moyen est inférieur au seuil de 20ms pour le sujet féminin et que le décalage moyen du sujet masculin est encore moins élevé. Néanmoins, l'écart entre ces deux décalages est trop faible pour tirer des conclusions sur l'influence du sexe sur les résultats de la segmentation automatique. Nous avons cependant remarqué lors

de la segmentation manuelle que le sujet féminin avait tendance à réaliser certaines occlusives comme des approximantes ce qui peut expliquer un décalage légèrement plus important.

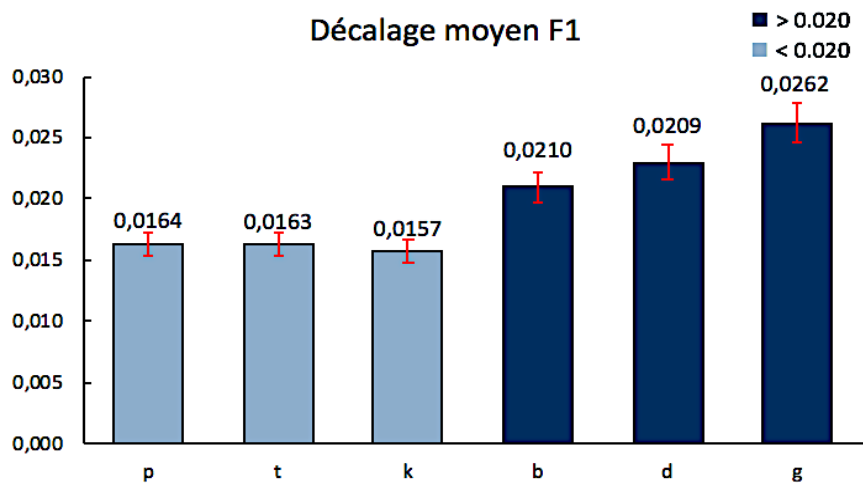


Figure 11 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence par consonnes pour le sujet féminin. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 180, N = 30 (consonnes)).

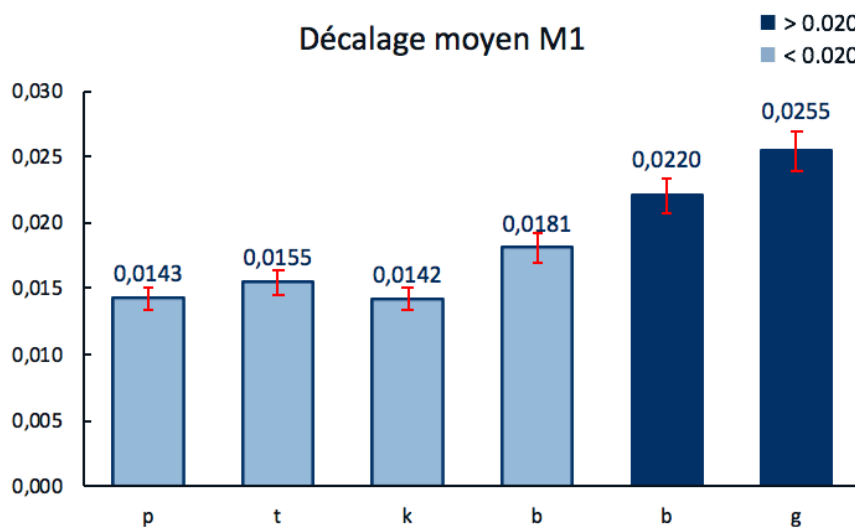


Figure 12 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence par consonnes pour le sujet masculin. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N=180, N = 30 (consonnes)).

On constate que le décalage moyen entre la segmentation manuelle et la segmentation automatique est plus important pour les consonnes sonores (b,d,g) que pour les consonnes sourdes (p,t,k). Cette tendance est observable chez les deux sujets avec une nouvelle fois un décalage un peu moins élevé pour le sujet masculin. Nous obtenons pour les consonnes sourdes un décalage moyen comparable aux résultats obtenus dans les tâches d'évaluation entre des annotateurs humains (16ms). En revanche, le décalage moyen pour les consonnes sonores est globalement supérieur au seuil de référence fixé à 20ms. La consonne qui semble être la plus éloignée de la référence pour les deux sujets est le /g/ qui a un décalage moyen supérieur à 25 ms. On observe une certaine constance dans les résultats entre les deux sujets qui obtiennent des seuils comparables pour toutes les consonnes. La différence la plus notable entre les deux sujets concerne le /b/ qui a un décalage inférieur à 20ms pour le sujet masculin alors qu'il est supérieur pour le sujet féminin. Comme indiqué précédemment nous avons remarqué que le sujet féminin avait tendance à prononcer certaines occlusives comme des approximantes, cela concerne plus particulièrement la consonne /b/ ce qui peut expliquer cette différence de résultat.

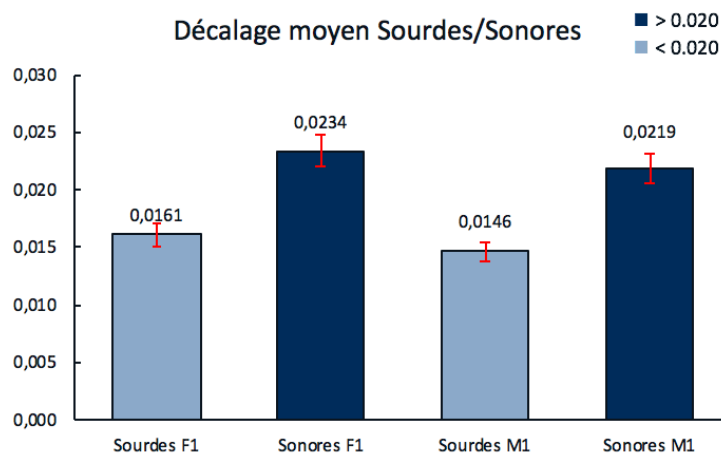


Figure 13 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence pour l'opposition de voisement. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 90).

Pour finir, nous avons comparé le décalage moyen entre les consonnes sourdes et sonores pour les deux sujets. Ces résultats nous confirment que le décalage est nettement plus important pour les consonnes sonores que pour les consonnes sourdes, avec un seuil supérieur à 20ms pour les deux sujets. Nous pensons que cette différence peut être expliquée par le fait que les systèmes de reconnaissance de la parole se basent sur des indices acoustiques afin de discriminer les sons. Or, plus ces indices sont nombreux et redondants, plus ils facilitent la discrimination. Les indices acoustiques pouvant être exploités pour discriminer les consonnes en français sont l'intensité de l'explosion, les valeurs de fréquence fondamentale, les transitions formantiques, la durée des segments ainsi que le voisement. Plus précisément, les indices acoustiques associés au voisement selon (Landron, 2017b) sont : le VOT (Voice Onset Time, le délai d'établissement du voisement), le taux de voisement durant la consonne (ou v-ratio), le pourcentage de voisement à différents moments de la consonne (ou v-pattern), la présence de périodicité du signal lié aux vibrations des plis vocaux au moment du relâchement, la durée de l'occlusion, la durée de la voyelle précédente, la valeur de la F0 au début de la voyelle suivante ou en fin de voyelle précédente. Sachant que les consonnes sourdes ont un délai de voisement plus long que les consonnes sonores nous pensons que les frontières sont plus facilement détectables ce qui pourrait expliquer cette différence.

Au regard des résultats obtenus, nous pensons qu'il est pertinent d'utiliser le système de segmentation automatique pour les deux sujets restants car nous obtenons globalement des résultats qui se rapprochent des valeurs de référence. De plus, lorsque nous avons regardé de plus près les décalages les plus importants, nous avons pu constater que le

système prenait en compte des frontières trop larges mais il ne détectait pas un segment à la place d'un autre. Il faudra cependant porter une attention toute particulière aux consonnes sonores qui sont susceptibles de générer le plus d'erreurs et qu'il faudra certainement corriger manuellement. Néanmoins, il sera dans tous les cas plus ergonomique de corriger les erreurs de fichiers déjà étiquetés et segmentés que de réaliser entièrement manuellement la segmentation et l'étiquetage.

## **5. Résultats et discussion**

Dans cette partie, nous souhaitons analyser les paramètres acoustiques et aérodynamiques impliqués dans la production de segments consonantiques afin de comprendre pourquoi et comment ces paramètres varient selon les catégories de sons. Notre objectif principal est d'apporter des valeurs de référence afin de mieux comprendre les mécanismes de phonation et de trouver des descripteurs pertinents permettant de comprendre comment les segments consonantiques étudiés se distinguent les uns des autres et au contraire quelles sont leurs caractéristiques communes. Cette analyse nous permettra par la suite de procéder à une classification automatique supervisée à partir des descripteurs extraits afin de classer les occlusives du français en fonction de leurs lieux d'articulation.

### **5.1 Extractions des descripteurs**

Afin d'extraire les descripteurs, nous avons choisi de focaliser notre étude sur l'analyse des mesures de pression intra-orale et de pression sous-glottique. L'extraction de ces mesures se fait généralement sur le logiciel Phonedit à travers une procédure manuelle qui consiste à

positionner un curseur à l'endroit souhaité sur les courbes de PIO et de Psg afin de connaître la valeur exacte de ces mesures et de pouvoir ensuite coller celles-ci dans un tableau Excel (cf. figure 14). Cette procédure est extrêmement fastidieuse et chronophage car il faut traiter les paramètres un à un ce qui implique le traitement d'une faible quantité de données dans un temps assez important. Cependant, nous n'avons pas trouvé de méthodes automatiques disponibles permettant de traiter automatiquement les paramètres aérodynamiques. Afin de résoudre ce problème, nous avons mis en place un programme d'extraction automatique de mesures aérodynamiques afin de pouvoir traiter un plus grand nombre de données et éviter les erreurs dues à la prise de mesure manuelle. Le programme d'extraction a été conçu avec le langage de programmation Python (cf. annexe D.2).

Le programme prend en entrée les fichiers TextGrid étiquetés et segmentés ainsi que les fichiers aérodynamiques. Il commence par parcourir les fichiers Textgrid afin d'accéder à la Tier « MAU » contenant le résultat de la segmentation automatique. Il procède ensuite à un filtrage afin de récupérer uniquement les étiquettes et les intervalles temporels des segments consonantiques étudiés en position C2. La deuxième étape consiste à parcourir les fichiers aérodynamiques afin de récupérer les valeurs du signal à un instant donné, celui des intervalles temporels attribués lors la segmentation. Si on se reporte à la figure 14, on constate que le segment /b/ a comme intervalles temporels 4 260 et 4 320 ms. Lorsqu'on positionne le curseur sur le signal à 4 260 ms, on obtient les valeurs de pression et d'intensité à cet instant donné, ce sont ces valeurs que notre programme d'extraction va récupérer automatiquement dans l'ensemble des fichiers.



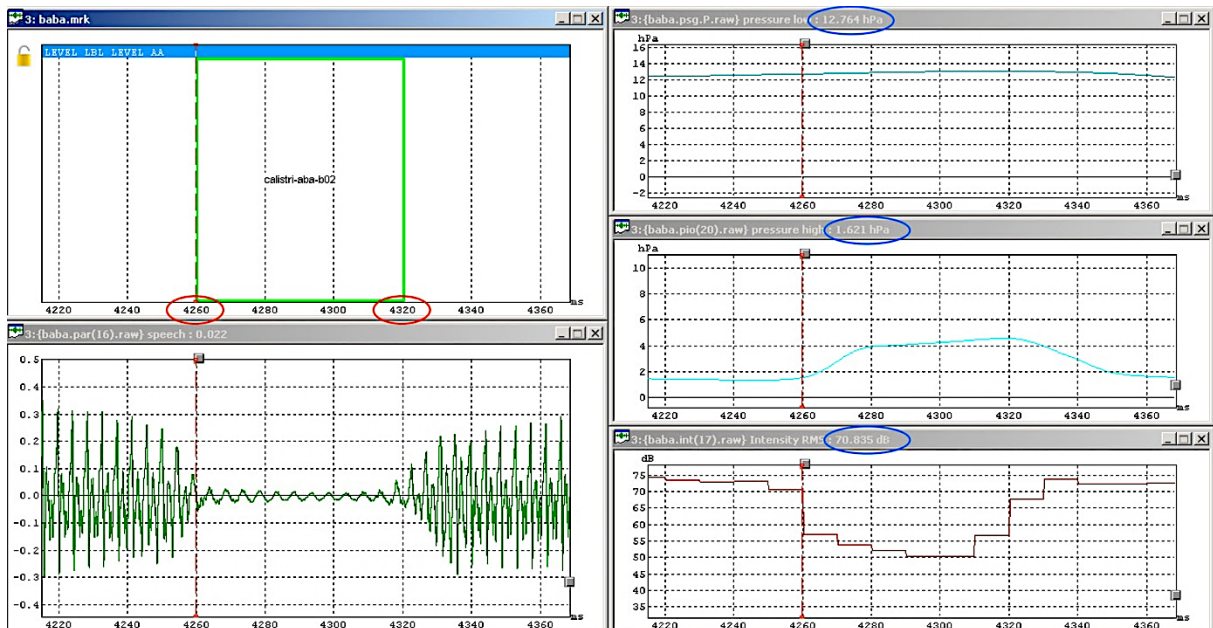


Figure 14 : Visualisation de la procédure d'extraction manuelle sur le logiciel Phonedit. Le curseur correspond à la ligne verticale rouge en pointillé. Les mesures entourées en rouge correspondent aux intervalles temporels et celles en bleu aux mesures de PIO, PSG et d'INT lorsque nous positionnons le curseur sur l'intervalle de début de l'occlusive /b/.

L'exécution de ce programme prend moins d'une minute pour traiter la totalité des données et structurer les informations dans un tableau, ce qui représente un gain de temps non négligeable. De plus, cela nous a permis de pouvoir augmenter considérablement le nombre de mesures car nous avons pu extraire entre les intervalles de début et de fin, huit points de mesures supplémentaires ce qui représente au total 10 points d'analyse par segment consonantique. L'extraction de ces 10 points de mesure nous permettra d'analyser les variations de pression au cours des différentes phases impliquées dans la production d'une occlusive (initiation, occlusion, burst et transition). Nous pourrions ainsi comparer la pertinence de chaque point afin de choisir ceux qui fournissent le plus d'information pour la classification automatique et, au contraire, de déterminer ceux qui pourraient générer le plus d'erreurs. On pourra également évaluer la synchronisation ou, au contraire, le manque de coordination entre les différents paramètres aérodynamiques.

Le résultat du programme d' extraction est un tableau (cf. figure 15) nous indiquant le nombre de valeurs extraites (index), le nom du fichier TextGrid traité (tgfile), le paramètre étudié (Psg, PIO), le sujet (F1, F2, M1, M2), le type de consonnes (sourdes, sonores), le segment consonantique concerné (p, t, k, b, d, g), l'index du point extraits (0 à 9), la valeur temporelle de ce point (ms), la valeur du paramètre correspondant (mesures de Psg, PIO) et enfin le contexte intervocalique (a, i, u). Nous obtenons un total de 1 800 mesures par sujet, 7 200 par paramètre soit un total de 14 400 mesures pour les deux paramètres. Les valeurs relevées constituent les données brutes qui serviront de base pour tous les calculs et traitements statistiques nécessaires à notre étude.

index	tgfile	datas	subject	type	consonnes	point_index	points	valeurs	contexte
0	DIDI.TextGrid	psg	F2	sonores	d	0	0.507044	9.013672	i
1	DIDI.TextGrid	psg	F2	sonores	d	1	0.519004	8.945312	i
2	DIDI.TextGrid	psg	F2	sonores	d	2	0.530965	8.886719	i
3	DIDI.TextGrid	psg	F2	sonores	d	3	0.542925	8.837891	i
4	DIDI.TextGrid	psg	F2	sonores	d	4	0.554886	8.808594	i
5	DIDI.TextGrid	psg	F2	sonores	d	5	0.566846	8.779297	i
6	DIDI.TextGrid	psg	F2	sonores	d	6	0.578807	8.710938	i
7	DIDI.TextGrid	psg	F2	sonores	d	7	0.590767	8.642578	i
8	DIDI.TextGrid	psg	F2	sonores	d	8	0.602728	8.544922	i
9	DIDI.TextGrid	psg	F2	sonores	d	9	0.614688	8.505859	i
10	DIDI.TextGrid	psg	F2	sonores	d	0	1.577310	7.958984	i

*Figure 15 : Résultat du programme d'extraction automatique.*

## 5.2 Analyse des données aérodynamiques

### 5.2.1 Analyse de la pression intra-orale

Les valeurs de la PIO ont constitué le point de départ de notre analyse car il s'agit d'un paramètre extrêmement robuste pour comprendre les caractéristiques des consonnes occlusives du français. En effet, dès qu'il y a une altération des constriction ou un mouvement d'un articulatoire, l'information est encodée dans les mesures de pression intra-orale. Notre objectif sera de comprendre comment ce paramètre varie à différents points de mesures en fonction des différents traits articulatoires impliqués dans la production des consonnes occlusives du français. Nous traiterons les sujets séparément afin d'analyser la variabilité intra-locuteur des mesures ou au contraire la constance de celles-ci et savoir si elles sont influencées par des facteurs extralinguistiques comme le sexe. Nous procéderons à un ensemble de tests paramétriques afin de trouver des descripteurs communs aux différents sujets et sélectionner uniquement ceux qui sont pertinents pour la classification automatique des lieux d'articulation.

Pour la réalisation des analyses statistiques nous avons procédé de la façon suivante : nous avons commencé par calculer les moyennes des mesures de PIO en fonction des traits articulatoires analysés. Puis, nous avons observé la distribution des données afin de savoir si il était possible d'appliquer un test-t de Student. Le test-t de Student est un test paramétrique permettant de comparer les moyennes de deux groupes d'échantillons afin d'observer si elles sont significativement différentes. Il existe plusieurs variantes de ce test, nous avons choisi d'appliquer le test-t de Student comparant deux groupes d'échantillons indépendants (non appariés). Avant d'appliquer ce test, il faut vérifier que les données respectent deux

conditions : (1) la distribution de l'échantillonnage suit une loi normale, (2) les variances des deux groupes sont égales. Dans les cas où la taille des échantillons est suffisamment grande ( $n > 30$ ), nous avons appliqué le théorème central limite disant que la distribution de l'échantillonnage tend à suivre la loi normale lorsque la taille est grande ( $n > 30$ ). Dans le cas contraire, nous avons appliqué le test de normalité Shapiro-Wilk. Ensuite, après avoir vérifié le respect de cette condition, nous avons comparé les variances des deux groupes. Nous avons constaté que les variances étaient significativement différentes. Or, si les deux groupes d'échantillons suivent une loi normale, mais ont des variances inégales, le test de Welch doit être utilisé. Nous avons donc appliqué ce test qui est une adaptation du test-t de Student afin de comparer les deux groupes. Nous avons appliqué cette méthodologie tout au long de notre étude. Nous présenterons à chaque fois les résultats des différents tests de Welch effectués dans des tableaux en commençant par les résultats obtenus pour les sujets féminins suivis de ceux obtenus pour les sujets masculins. Afin d'obtenir une visualisation graphique de ces résultats, nous avons réalisé, avec le langage de programmation R, les courbes moyennes des valeurs de PIO. Ces courbes représentent l'évolution des valeurs moyennes de PIO au cours de l'occlusive. Ainsi, nous pouvons observer comment ces valeurs évoluent dans le temps en fonction des différents points de mesures extraits mais également en fonction des différents traits étudiés. Elles nous permettront de modéliser les résultats obtenus lors des différents tests statistiques et de pouvoir les comparer entre les sujets d'étude et les différents paramètres analysés.

### 5.2.1.1 L'opposition de voisement

L'opposition de voisement est un trait qui permet de distinguer les sons dont la production s'accompagne d'une vibration des plis vocaux par rapport aux sons non voisés produits avec une absence de vibrations. En français, la corrélation de voisement (opposition entre consonnes voisées et non voisées) permet d'opposer les séries symétriques /p/, /t/, /k/, et /b/, /d/, /g/. Les mesures de PIO constituent un avantage particulier pour étudier l'opposition de voisement, car elle reflète directement la combinaison des nombreuses actions articulatoires aux niveaux laryngé et supra-laryngé. Nous étudierons dans cette partie, comment ces mesures varient en fonction de la corrélation de voisement. Les résultats obtenus pour les sujets féminins sont les suivants :

*Tableau 2 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet F1. La p-value (P) est indiquée avec son niveau de significativité (\* p < 0,05; \*\* p < 0,01; \*\*\* p < 0,001) (N=900 (NV), N=900(V)).*

	Moyennes des mesures de PIO pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>NV</b>	3.11	5.41	8.09	10.31	11.04	11.13	11.04	10.50	9.28	7.29
<b>V</b>	2.72	3.17	3.61	4.00	4.35	4.67	4.98	5.26	5.48	5.54
<b>P</b>	0.57	0.009 **	1.255 <sup>e</sup> -05***	4.406 <sup>e</sup> -09***	2.265 <sup>e</sup> -09***	3.462 <sup>e</sup> -09***	6.523 <sup>e</sup> -09***	1.249 <sup>e</sup> -08***	1.215 <sup>e</sup> -06***	0.005 **

Tableau 3 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet F2. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PIO pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>NV</b>	1.51	1.81	3.03	3.67	3.92	4.07	4.18	4.27	4.32	4.10
<b>V</b>	1.04	1.26	1.62	1.85	2.02	2.17	2.36	2.60	2.90	3.19
<b>P</b>	0.04*	0.11	0.019 *	0.017 *	0.017 *	0.014 *	0.012 *	0.016 *	0.04 *	0.16

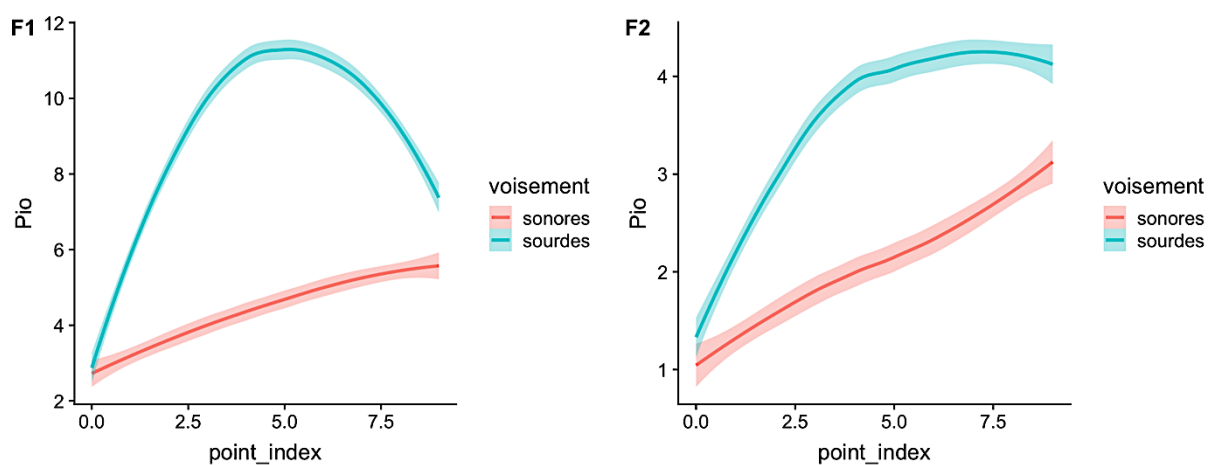


Figure 16 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction.

Les résultats obtenus pour les sujets masculins sont les suivants :

Tableau 4 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet M1. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PIO pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
NV	3.45	6.19	9.52	11.96	13.27	13.82	13.89	13.55	12.43	9.206
V	2.69	3.00	3.53	3.97	4.25	4.57	4.97	5.35	5.64	5.73
P	0.004 **	2.2e16 ***	2.2e16 ***	2.2e16 ***	2.2e-16 ***	2.2e-16 ***	2.2e16 ***	2.2e16 ***	2.2e-16 ***	5.228e-16 ***

Tableau 5 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet M2. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PIO pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
NV	1.58	2.93	6.16	8.20	8.87	8.99	8.94	8.68	7.89	5.89
V	1.22	1.54	2.37	3.27	3.90	4.44	4.88	5.24	5.51	5.53
P	0.04 *	1.42e-08 ***	2.2e-16 ***	2.2e-16 ***	2.2e-16 ***	2.2e-16 ***	2.2e-16 ***	2.2e-16 ***	2.2e-16 ***	0.14

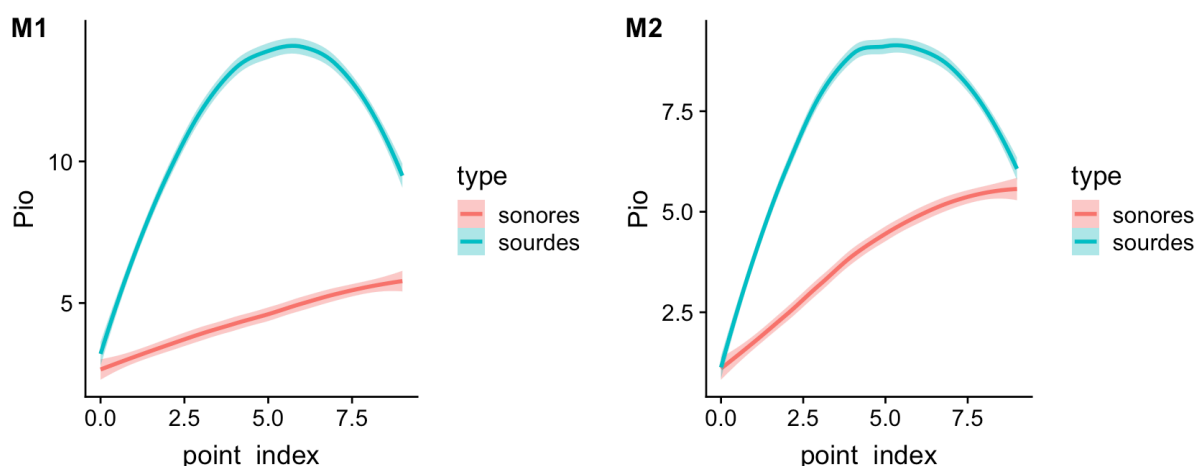


Figure 17 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleu (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction.

Nous pouvons constater que les mesures de PIO sont plus élevées lors de la production de consonnes non voisées que lors de la production de consonnes voisées. Ce *pattern* est identifiable sur les courbes moyennes (cf. figures 16, 17), on constate que la PIO monte rapidement après la fermeture des lèvres pour les consonnes non voisées, puis se maintient avec un plateau (très légèrement ascendant) jusqu'au moment où les lèvres se relâchent, ce qui provoque une chute rapide de la PIO. Pour les consonnes voisées, la PIO monte immédiatement puis elle s'infléchit rapidement et elle suit une courbe ascendante jusqu'à un pic, qui est nettement moins élevé que celui des consonnes non voisées. Cela s'explique par le fait que la glotte étant en adduction (les plis vocaux sont en contact), le flux d'air est ralenti et la PIO monte moins vite. Cette tendance est observable chez les quatre sujets. Les résultats du test de Welch, nous confirment cette hypothèse car il existe une différence hautement significative entre les deux groupes pour les sujets F1, M1 et M2. En revanche, cette différence est nettement moins marquée chez le sujet F2 dont les valeurs de PIO sont assez faibles, contrairement aux autres sujets mais l'opposition de voisement est tout de même présente (cf. tableau 3). Afin de comprendre pourquoi l'opposition de voisement est moins marquée pour ce sujet nous avons voulu regarder si cela pouvait provenir de l'influence de l'environnement vocalique car celui-ci peut notamment affecter le caractère voisé ou non voisé des consonnes (cf. figure 18). On constate que l'environnement vocalique semble effectivement avoir une influence sur l'opposition de voisement pour ce sujet. La différence entre consonnes voisées et non voisées est faiblement marquée dans le contexte intervocalique /i/ comparé à /u/ et /a/. Plus spécifiquement, le contexte intervocalique /i/ semble affecter les consonnes voisées uniquement car la courbe moyenne des consonnes non voisées conserve le même *pattern* contrairement à celle des consonnes voisées. En effet,



nous avons constaté pour tous les autres sujets une différence hautement significative entre consonnes voisées et non voisées. Or, dans le contexte /i/ les mesures de PIO des consonnes voisées augmentent considérablement, ce qui explique que cette différence soit moins marquée pour ce sujet. Au vu de ces résultats, nous procéderons à une analyse plus détaillée du contexte intervocalique afin de mieux comprendre ces répercussions sur les différences de voisement et sur les mesures de PIO.

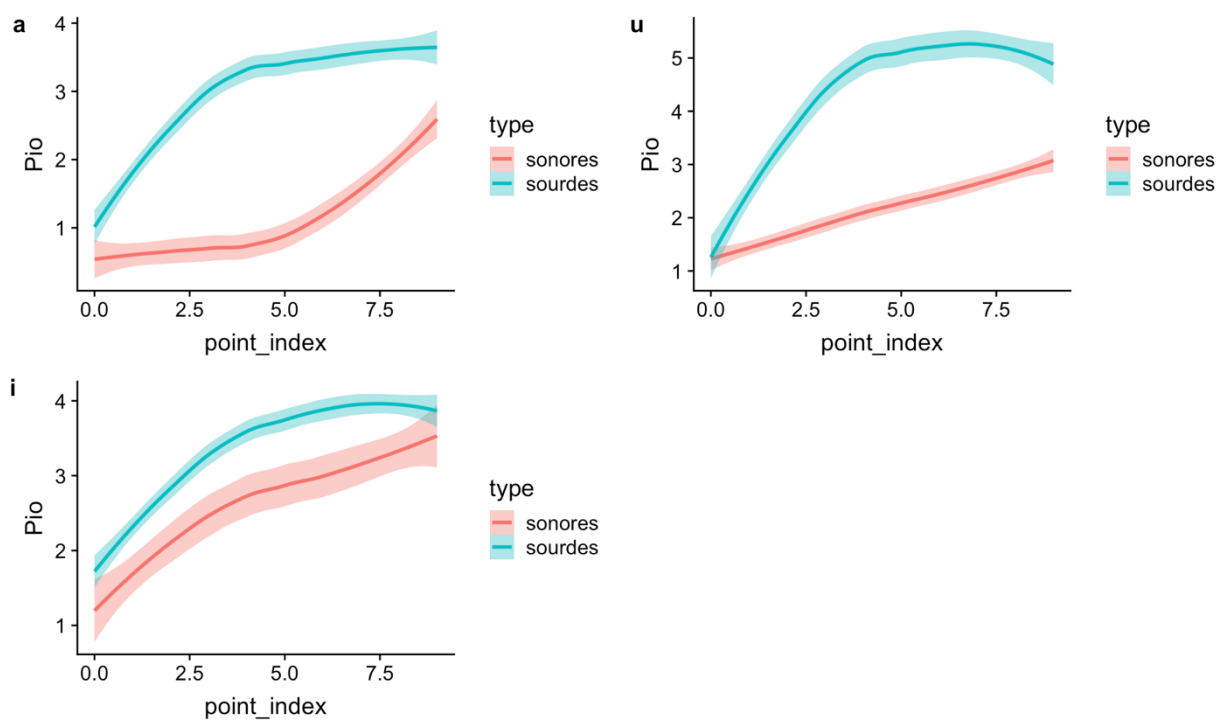


Figure 18 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour le sujet F2.

On remarque également que les hommes ont des valeurs de PIO plus élevées que les femmes pour les consonnes non voisées. Cela est probablement dû à diverses caractéristiques anatomiques : les différences de longueur des plis vocaux qui peuvent affecter la résistance glottique, les différences de longueur du conduit vocal et les différents volumes pulmonaires entre les hommes et femmes. En ce qui concerne les points de mesure, il ressort

que les points les moins pertinents pour l'analyse sont les points [0] et [9] et au contraire les mesures prises au milieu semblent être les descripteurs les plus robustes pour détecter le voisement.

### 5.2.1.2 Le contexte intervocalique

« Le contexte phonétique qui entoure le phonème influence les gestes de la langue et des lèvres, du voile du palais et de la glotte par lequel il est réalisé. Comparez les positions de vos lèvres durant la réalisation de la première consonne des mots toute et tête ; elles sont plus arrondies durant la production des consonnes /t/ dans toute que dans tête » (Vaissière, 2006).

En fonction du contexte phonétique entourant le phonème, l'ensemble de l'appareil vocal va adapter sa configuration et sa position. Il est donc essentiel d'étudier l'impact du contexte intervocalique car la configuration du conduit vocal durant la production des occlusives sera dépendante du contexte dans lequel la consonne est réalisée. Comme nous l'avons vu précédemment, le contexte intervocalique semble avoir une influence sur l'opposition de voisement pour le sujet F2 mais également sur les valeurs de pression intra-orale. Nous allons donc dans cette partie, étudier l'influence du contexte intervocalique sur les consonnes voisées et non voisées en comparant les mesures de PIO dans les trois contextes intervocaliques /a,i,u/. La voyelle la plus ouverte est le /a/ qui est une voyelle antérieure fermée et étirée, /i/ est une voyelle antérieure fermée et étirée et /u/ une voyelle fermée, arrondie et postérieure. Nous chercherons à savoir s'il existe une différence significative entre les mesures de PIO dans le contexte /a/ comparé à /i/, puis dans le contexte /a/ comparé à /u/ et

pour finir dans le contexte /u/ comparé à /i/. Les résultats obtenus pour les sujets féminins sont les suivants :

Tableau 6 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\* p < 0,05; \*\* p < 0,01; \*\*\* p < 0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	1.66	4.36	7.25	9.46	10.20	10.43	10.53	10.45	9.86	7.66
<b>i</b>	4.58	6.67	9.01	11.19	11.85	11.75	11.50	10.89	9.72	8.22
<b>u</b>	3.09	5.21	8.00	10.27	11.08	11.23	11.11	10.17	8.26	5.98
<b>a vs. i</b>	3.091 <sup>e</sup> -10 ***	0.0003 ***	0.01 *	0.005 **	0.004 **	0.02 *	0.085	0.41	0.78	0.35
<b>a vs. u</b>	6.965 <sup>e</sup> -06 ***	0.10	0.17	0.08	0.04 *	0.06	0.18	0.55	0.0009 ***	0.0007 ****
<b>u vs. i</b>	0.0006 ***	0.01 *	0.13	0.08	0.11	0.28	0.44	0.19	0.018 *	0.0003 ***

Tableau 7 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\* p < 0,05; \*\* p < 0,01; \*\*\* p < 0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	1.38	1.76	2.27	2.63	2.95	3.30	3.63	4.07	4.66	5.08
<b>i</b>	2.86	3.43	3.88	4.36	4.79	5.17	5.59	5.91	5.99	5.93
<b>u</b>	3.93	4.32	4.70	5.02	5.32	5.56	5.71	5.79	5.81	5.61
<b>a vs. i</b>	8.37 <sup>e</sup> -09***	1.22 <sup>e</sup> - 09***	7.26 <sup>e</sup> -07***	9.68 <sup>e</sup> -06***	3.997 <sup>e</sup> -05***	0.0001 ***	0.0001 ***	0.0005 ***	0.01 *	0.09
<b>a vs. u</b>	1.85 <sup>e</sup> -07***	3.63 <sup>e</sup> -08***	3.57 <sup>e</sup> -08***	2.56 <sup>e</sup> -08 ***	3.51 <sup>e</sup> -08 ***	6.09 <sup>e</sup> -07 ***	1.40 <sup>e</sup> -05 ***	0.0004 ***	0.02 *	0.31
<b>u vs. i</b>	0.01 * ***	0.02 * ***	0.05	0.14	0.27	0.46	0.83	0.83	0.75	0.55

Tableau 8 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	1.19	1.40	2.60	3.09	3.27	3.38	3.48	3.56	2.06	3.63
<b>i</b>	1.83	2.06	2.89	3.34	3.56	3.73	3.86	3.97	4.00	3.83
<b>u</b>	1.51	1.96	3.61	4.59	4.93	5.09	5.20	5.28	3.07	4.82
<b>a vs. i</b>	1.396 <sup>e</sup> -06***	1.099 <sup>e</sup> -07***	0.106	0.23	0.15	0.10	0.07	0.05	0.08	0.32
<b>a vs. u</b>	7.114 <sup>e</sup> -05***	4.11 <sup>e</sup> -08***	4.67 <sup>e</sup> -05***	5.18 <sup>e</sup> -06***	2.68 <sup>e</sup> -06***	2.06 <sup>e</sup> -06 ***	1.87 <sup>e</sup> -06 ***	1.88 <sup>e</sup> -06 ***	3.06 <sup>e</sup> -06 ***	6.66 <sup>e</sup> -05***
<b>u vs. i</b>	0.007 **	0.309	0.001 **	4.19 <sup>e</sup> -05 ***	3.814 <sup>e</sup> -05***	5.05 <sup>e</sup> -05 ***	7.27 <sup>e</sup> -05 ***	0.00 1**	0.000 1***	0.000 6***

Tableau 9 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	0.51	0.60	0.76	0.68	0.70	0.87	1.18	1.58	2.06	2.57
<b>i</b>	1.25	1.58	2.10	2.50	2.74	2.86	2.97	3.12	3.36	3.52
<b>u</b>	1.26	1.38	1.63	2.06	2.10	2.27	2.44	2.62	2.84	3.07
<b>a vs. i</b>	3.389 <sup>e</sup> -09***	3.57 <sup>e</sup> -07***	1.10 <sup>e</sup> -05***	6.12 <sup>e</sup> -07 ***	4.87 <sup>e</sup> -08***	1.91 <sup>e</sup> -08***	7.35 <sup>e</sup> -08***	2.68 <sup>e</sup> -06 ***	0.000 5***	0.03 *
<b>a vs. u</b>	2.622 <sup>e</sup> -10***	1.307 <sup>e</sup> -09***	4.192 <sup>e</sup> -06***	1.72 <sup>e</sup> -07 ***	3.204 <sup>e</sup> -09***	1.066 <sup>e</sup> -10***	1.758 <sup>e</sup> -10***	5.65 <sup>e</sup> -07 ***	0.005 **	0.18
<b>u vs. i</b>	0.90	0.23	0.08	0.057	0.046 *	0.050	0.068	0.084	0.091	0.12

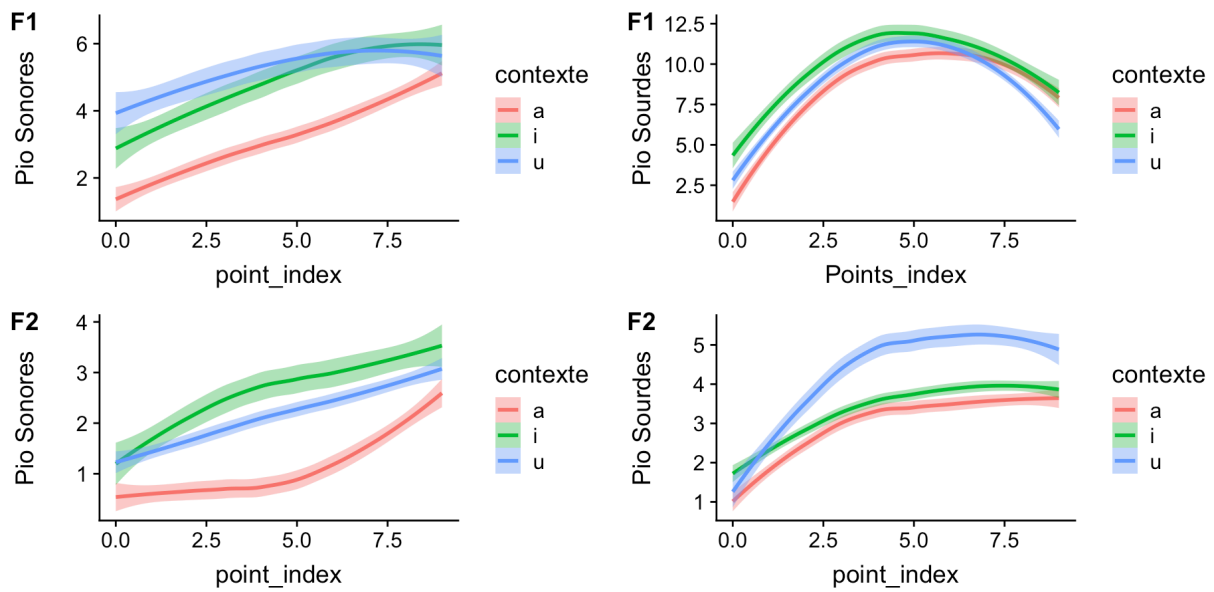


Figure 19 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets féminins.

Les résultats obtenus pour les sujets masculins sont les suivants :

Tableau 10 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	1.71	5.26	9.69	11.96	12.59	12.66	12.58	12.37	11.87	7.78
<b>i</b>	4.79	7.29	10.05	12.83	14.51	15.27	15.34	14.80	13.39	10.71
<b>u</b>	3.86	6.02	8.81	11.07	12.72	13.56	13.77	13.48	12.04	9.11
<b>a vs. i</b>	1.528 <sup>e</sup> -11***	9.672 <sup>e</sup> -05***	0.54	0.16	0.0004 ***	2.081 <sup>e</sup> -07***	1.789 <sup>e</sup> -09***	7.467 <sup>e</sup> -09***	0.000 8***	0.000 5***
<b>a vs. u</b>	1.327 <sup>e</sup> -08***	0.12	0.14	0.14	0.80	0.07	0.01 *	0.01 *	0.729	0.09
<b>u vs. i</b>	0.016 *	0.009 **	0.03 *	0.012 *	0.0097 **	0.005 *	0.004 **	0.008 **	0.015 *	0.008 **

Tableau 11 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

Moyennes des mesures de PIO pour les consonnes V pour le sujet M1										
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	1.52	1.77	2.40	2.88	2.95	3.17	3.59	3.95	4.23	4.50
<b>i</b>	3.59	3.91	4.29	4.61	4.91	5.26	5.62	6.01	6.41	6.53
<b>u</b>	2.98	3.33	3.88	4.41	4.88	5.29	5.70	6.08	6.28	6.16
<b>a vs. i</b>	1.854e -06***	2.79e -06 ***	0.00 01 ***	0.000 9***	0.0002 ***	0.000 1***	0.000 1***	5.37e -05 ***	6.55e -06 ***	3.77e -05 ***
<b>a vs. u</b>	5.693e -07***	3.06e -06 ***	7.12 8e- 05 ***	0.000 2***	1.563e -05***	5.781e -06 ***	1.772e -05 ***	2.182e -05 ***	3.957 e-05 ***	0.000 9***
<b>u vs. i</b>	0.12	0.18	0.38	0.69	0.95	0.95	0.88	0.88	0.80	0.45

Tableau 12 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

Moyennes des mesures de PIO pour les consonnes NV pour le sujet M2										
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	0.17	1.01	5.05	6.85	7.31	7.49	7.59	7.57	7.09	5.02
<b>i</b>	1.89	3.60	6.85	9.31	10.01	9.98	9.84	9.46	8.51	6.63
<b>u</b>	2.68	4.19	6.59	8.42	9.30	9.51	9.39	9.00	8.07	6.01
<b>a vs. i</b>	5.365e -10 ***	1.295 e-09 ***	9.984 e-05 ***	4.786e -11 ***	1.181e -15***	4.269e -15 ***	1.397e -14 ***	1.889e -13 ***	5.892 e-07 ***	0.000 1***
<b>a vs. u</b>	2.2e -16***	2.2e- 16***	7.022 e-05 ***	8.185e -08 ***	5.617e -13***	2.363e -14 ***	2.785e -12 ***	2.46e -07***	0.007 **	0.017 *
<b>u vs. i</b>	0.0009 ***	0.10	0.44	0.005 **	0.0070 **	0.052	0.060	0.078	0.19	0.062

Tableau 13 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

Moyennes des mesures de PIO pour les consonnes V pour le sujet M2										
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	0.05	0.14	0.85	1.67	2.14	2.57	2.97	3.40	3.82	3.82
<b>i</b>	1.80	2.22	3.21	4.21	4.95	5.52	5.91	6.17	6.30	6.28
<b>u</b>	1.82	2.25	3.05	3.93	4.61	5.22	5.77	6.16	6.42	6.50
<b>a vs. i</b>	8.937 <sup>e</sup>	6.805	4.137	1.931 <sup>e</sup>	3.29 <sup>e</sup>	4.913 <sup>e</sup>	2.2 <sup>e</sup>	4.388 <sup>e</sup>	3.557	4.276 <sup>e</sup>
	-10	<sup>e</sup> -10	<sup>e</sup> -11	-12	-14	-16	-16	-15	<sup>e</sup> -12	-13
	***	***	***	***	***	***	***	***	***	***
<b>a vs. u</b>	5.65 <sup>e</sup>	2.2 <sup>e</sup>	2.2 <sup>e</sup>	2.2 <sup>e</sup>	2.2 <sup>e</sup>	2.2 <sup>e</sup>	2.2 <sup>e</sup>	7.902 <sup>e</sup>	4.5 <sup>e</sup>	1.097 <sup>e</sup>
	-15 <sup>***</sup>	-16 <sup>***</sup>	-16 <sup>***</sup>	-16 <sup>***</sup>	-16 <sup>***</sup>	-16 <sup>***</sup>	-16 <sup>***</sup>	-15 <sup>***</sup>	-12 <sup>***</sup>	-12 <sup>***</sup>
<b>u vs. i</b>	0.93	0.92	0.55	0.27	0.17	0.23	0.56	0.99	0.63	0.36

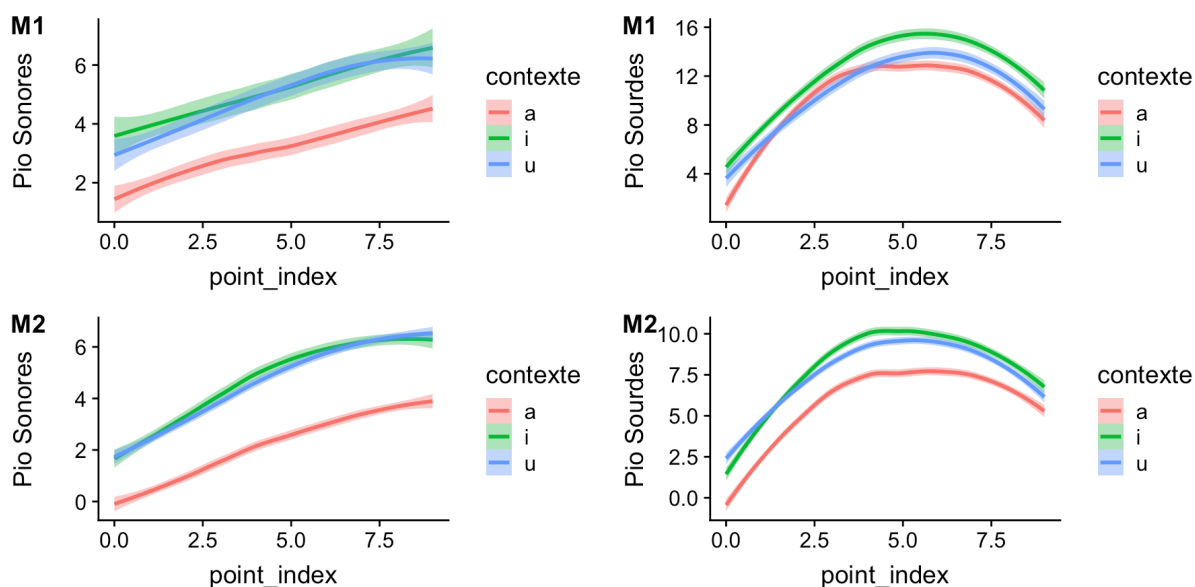


Figure 20 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets masculins.

Tableau 14 : Résumé des différences significatives des mesures de PIO entre les différents contextes intervocaliques

	Non Voisées	Voisées
F1	[a vs. i]	[a vs. i] & [a vs. u]
F2	[a vs. u] & [i vs. u]	[a vs. i] & [a vs. u]
M1	[a vs. i] & [i vs. u]	[a vs. i] & [a vs. u]
M2	[a vs. i] & [a vs. u]	[a vs. i] & [a vs. u]

Les différents tests nous permettent de confirmer que le contexte intervocalique a effectivement une influence sur les mesures de PIO et sur le voisement. Pour les consonnes voisées, on constate un invariant stable pour les 4 sujets qui est la différence significative des mesures de PIO dans le contexte /a/ comparé à /i/ et /u/ (cf. tableau 13). Les valeurs de la PIO dans le contexte /a/ sont toujours inférieures à celles dans les contextes /i/ et /u/ ce qui explique cette différence significative. Ce résultat n'est pas étonnant, il peut être expliqué par les différences de degré d'ouverture du conduit vocal lors de la production des consonnes occlusives dans les différents contextes intervocaliques (cf. figure 21). En effet, dans le contexte /a/, la masse de la langue est reculée vers l'arrière et la langue se trouve en bas de la bouche. Cette configuration du conduit vocal permet de laisser le passage de l'air ouvert ce qui permet à celui-ci de s'échapper rapidement et a pour conséquence de faire diminuer considérablement la pression intra-orale. Lors de la production d'un /a/, la résistance à l'écoulement d'air est donc relativement faible. A l'inverse, lorsqu'on regarde la configuration du conduit vocal lors de la production d'un /i/ et d'un /u/ (cf. figure 21), on remarque qu'elles ont un degré d'ouverture du conduit vocal réduit à cause de la masse de la langue qui se situe près du palais pour les voyelles fermées. Le passage de l'air est donc presque fermé ce qui augmente la résistance à l'écoulement d'air et empêche l'air de s'évacuer rapidement. Ainsi,



il est attendu d'obtenir des mesures de PIO plus élevées dans les contextes /i,u/ que dans le contexte /a/. Ce *pattern* est également identifiable pour les consonnes non voisées mais il est plus variable. Dans le contexte /a/, les valeurs de PIO restent plus faibles mais en fonction des sujets elles se rapprochent fortement des valeurs de /i/ et /u/. Pour le sujet F1, il y a une différence significative entre /a/ et /i/ mais les valeurs de /a/ et /u/ sont très proches (cf. tableau 13). Pour les sujets F2 et M1, il y a une différence significative entre /a/ et /u/ mais cette fois-ci ce sont les valeurs de /a/ et de /i/ qui sont très proches. Pour les consonnes non voisées, on constate pour les sujets F2 et M1 une différence significative entre [i vs. u], cela vient peut-être du fait que la voyelle /i/ a un allongement minimal du conduit vocal contrairement à /u/ qui a un allongement maximal du conduit vocal.

On ne remarque pas de différences particulières entre les sujets féminins et masculins concernant ce descripteur.

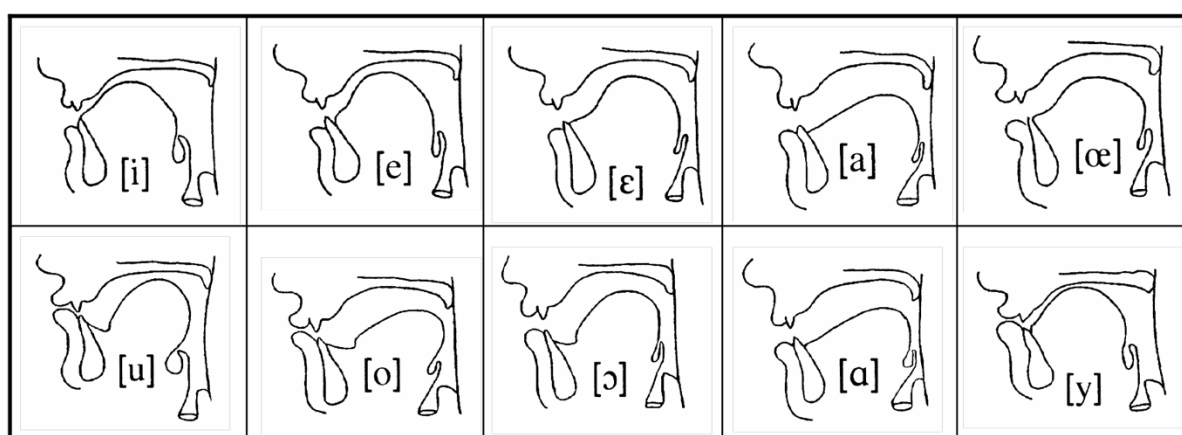


Figure 21 : Coupes sagittales du conduit vocal d'un locuteur standard dans la réalisation des voyelles orales du français.

### 5.2.1.3 Les lieux d'articulation

Les lieux d'articulation sont les endroits précis dans la cavité buccale où l'air rencontre un obstacle total ou partiel. Plus précisément, il s'agit du point de rapprochement ou de contact entre la partie mobile (langue ou lèvre intérieure) et la partie fixe (lèvre supérieure, dents ou palais) de l'appareil phonateur au cours de l'émission sonore. On distingue, en haut, sur le plafond de la bouche, les zones labiale, dentale, alvéolaire, palatale, vélaire, uvulaire et pharyngale ; et en bas, sur la langue, les zones apicale, dorsale et radicale. Ainsi, en combinant la région de la partie fixe et la région de la partie mobile qui se rapprochent ou entre en contact au cours de l'émission d'un son, on peut déterminer avec précision le lieu d'articulation de ce son. Nous travaillons dans le cadre de notre étude sur les occlusives du français qui peuvent se distinguer selon trois lieux d'articulation (cf. figure 22) : bilabiale (/p/,/b/), alvéolaire (/t/,/d/) et vélaire (/k/,/g/). Pour les occlusives bilabiales, l'occlusion est réalisée par les deux lèvres qui prennent fermement contact l'une contre l'autre. Pour les alvéolaires, l'occlusion est réalisée par la pointe de la langue qui s'accolle contre la crête alvéolaire. Pour finir, pour les consonnes vélares, l'occlusion est réalisée par la partie postérieure de la langue qui se rapproche du palais mou.

L'objectif principal de notre travail est de procéder à une classification automatique des mesures de PIO selon ces trois lieux d'articulation. Nous avons vu précédemment que l'opposition de voisement était significativement marquée dans les mesures de PIO nous allons donc continuer de séparer ces deux groupes pour l'analyse. Nous allons dans cette partie, comparer les mesures de PIO entre les occlusives bilabiales par rapport aux alvéolaires, entre les bilabiales par rapport aux vélares et pour finir entre les vélares par

rapport aux alvéolaires afin de savoir si elles ont des mesures de PIO significativement différentes.

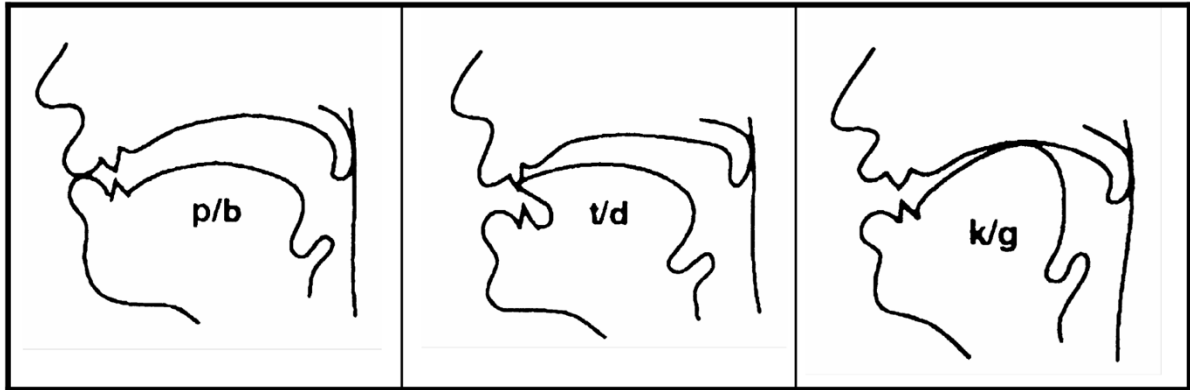


Figure 22 : Coupes sagittales du conduit vocal d'un locuteur standard dans la réalisation des lieux d'articulation du français.

Les résultats obtenus pour les sujets féminins sont les suivants :

Tableau 15 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	2.43	3.94	6.62	9.55	10.67	10.82	10.79	10.16	8.96	7.38
<b>Alvéolaire</b>	3.38	6.79	9.31	10.62	11.08	11.30	11.23	10.81	9.51	7.11
<b>Vélaire</b>	3.53	5.52	8.33	10.76	11.36	11.29	11.11	10.55	9.37	7.39
<b>Bilabiale vs. Alvéolaire</b>	0.03*	8.115 <sup>e</sup> -07***	3.212 <sup>e</sup> -05***	0.06	0.47	0.40	0.44	0.26	0.37	0.68
<b>Bilabiale vs. Vélaire</b>	0.01*	0.004**	0.005**	0.01*	0.11	0.28	0.47	0.39	0.45	0.99
<b>Vélaire vs. Alvéolaire</b>	0.75	0.03*	0.13	0.80	0.59	0.98	0.81	0.62	0.77	0.54

Tableau 16 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	2.51	2.81	3.20	3.52	3.84	4.10	4.44	4.82	5.02	4.88
<b>Alvéolaire</b>	2.99	3.44	3.80	4.13	4.46	4.84	5.21	5.55	5.84	6.07
<b>Vélaire</b>	2.67	3.26	3.85	4.37	4.76	5.09	5.30	5.42	5.59	5.68
<b>Bilabiale vs. Alvéolaire</b>	0.36	0.21	0.22	0.22	0.23	0.17	0.17	0.21	0.17	0.03*
<b>Bilabiale vs. Vélaire</b>	0.52	0.10	0.03*	0.01*	0.02*	0.02*	0.06	0.19	0.19	0.06
<b>Vélaire Vs. Alvéolaire</b>	0.52	0.70	0.91	0.63	0.57	0.66	0.87	0.83	0.68	0.49

Tableau 17 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	1.44	1.77	3.14	4.02	4.34	4.51	4.64	4.75	4.82	4.50
<b>Alvéolaire</b>	1.69	1.88	2.90	3.45	3.67	3.81	3.93	4.03	4.09	3.93
<b>Vélaire</b>	1.40	1.78	3.00	3.49	3.69	3.81	3.90	3.97	3.97	3.81
<b>Bilabiale vs. Alvéolaire</b>	0.08	0.44	0.35	0.10	0.08	0.07	0.07	0.06	0.06	0.08
<b>Bilabiale vs. Vélaire</b>	0.61	0.91	0.41	0.02	0.01	0.008	0.005	0.003	0.001	0.0006
<b>Vélaire Vs. Alvéolaire</b>	0.02	0.39	0.67	0.88	0.95	0.99	0.93	0.86	0.72	0.66

Tableau 18 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	1.11	1.11	1.22	1.41	1.59	1.73	1.85	1.97	2.10	2.22
<b>Alvéolaire</b>	0.96	0.92	1.04	1.25	1.47	1.68	1.87	2.07	2.30	2.52
<b>Vélaire</b>	1.09	1.63	2.33	2.60	2.73	2.86	3.09	3.45	3.95	4.40
<b>Bilabiale vs. Alvéolaire</b>	0.04	0.01	0.01	0.04	0.23	0.68	0.87	0.50	0.24	0.10
	*	*	*	*						
<b>Bilabiale vs. Vélaire</b>	0.91	0.007	0.0002	0.0013	0.003	0.002	0.0002	2.016	1.417	4.685
		**	***	**	**	**	***	e-06	e-09	e-14
								***	***	***
<b>Vélaire Vs. Alvéolaire</b>	0.28	0.0004	e-05	0.0003	0.001	0.001	0.0002	4.047	9.646	4.392
		***	***	***	**	**	***	e-06	e-09	e-13
								***	***	***

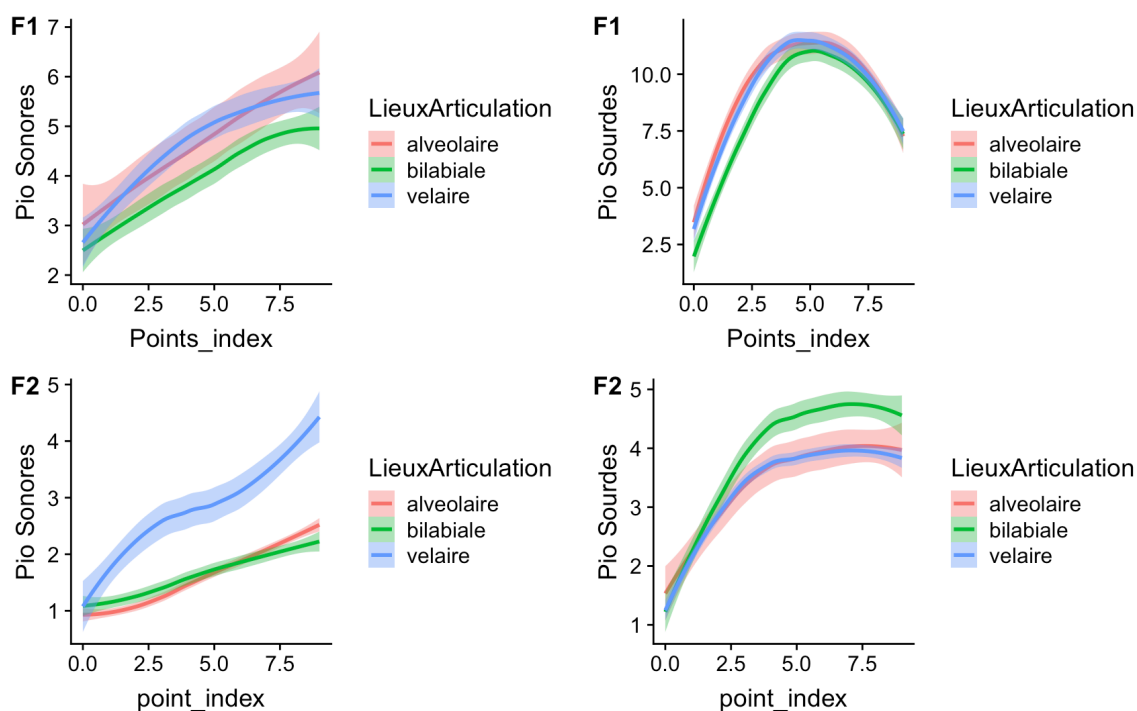


Figure 23 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les trois lieux d'articulation pour les sujets féminins.

Les résultats obtenus pour les sujets masculins sont les suivants :

Tableau 19 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	2.14	2.29	2.52	2.78	3.08	3.43	3.81	4.21	4.54	4.56
<b>Alvéolaire</b>	2.55	2.85	3.64	4.41	4.72	4.99	5.38	5.75	6.05	6.21
<b>Vélaire</b>	3.39	3.87	4.41	4.71	4.95	5.31	5.72	6.08	6.33	6.42
<b>Bilabiale vs. Alvéolaire</b>	0.26	0.14	0.006 **	0.0002 ***	0.0003 ***	0.0008 ***	0.0008 ***	0.000 g***	0.001 **	0.000 g****
<b>Bilabiale vs. Vélaire</b>	0.003 **	0.000 5***	0.0001 ***	0.0001 ***	0.0005 ***	0.0008 ***	0.0008 ***	0.001 **	0.001 **	0.000 g***
<b>Vélaire Vs. Alvéolaire</b>	0.04 *	0.02 *	0.08	0.51	0.65	0.55	0.52	0.52	0.56	0.64

Tableau 20 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	2.14	2.29	2.52	2.78	3.08	3.43	3.81	4.21	4.54	4.56
<b>Alvéolaire</b>	2.55	2.85	3.64	4.41	4.72	4.99	5.38	5.75	6.05	6.21
<b>Vélaire</b>	3.39	3.87	4.41	4.71	4.95	5.31	5.72	6.08	6.33	6.41
<b>Bilabiale vs. Alvéolaire</b>	0.26	0.14	0.006 **	0.0002 ***	0.0003 ***	0.000 g***	0.000 g***	0.000 0***	0.001 **	0.0009 ***
<b>Bilabiale vs. Vélaire</b>	0.003 **	0.0005 ***	0.0001 ***	0.0001 ***	0.0005 ***	0.000 g***	0.000 g***	0.001 **	0.001 **	0.0009 ***
<b>Vélaire Vs. Alvéolaire</b>	0.04 *	0.02 *	0.08	0.51	0.65	0.55	0.52	0.52	0.56	0.64

Tableau 21 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes NV pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	1.37	2.47	5.33	7.45	8.09	8.24	8.28	8.18	7.65	6.07
<b>Alvéolaire</b>	1.29	2.50	6.43	8.39	9.15	9.29	9.27	9.12	8.61	6.37
<b>Vélaire</b>	2.09	3.83	6.72	8.75	9.38	9.45	9.27	8.73	7.41	5.22
<b>Bilabiale vs. Alvéolaire</b>	0.76	0.94	0.001 **	0.002 **	0.001 **	0.000 8***	0.000 8***	0.001 **	0.002 **	0.34
<b>Bilabiale vs. Vélaire</b>	0.04 *	0.009 **	0.004 **	0.002 **	0.001 **	0.001 **	0.004 **	0.093	0.49	0.06
<b>Vélaire Vs. Alvéolaire</b>	0.01 *	0.003 **	0.46	0.34	0.53	0.64	0.99	0.20	0.000 6***	0.007 **

Tableau 22 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PIO pour les consonnes V pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	0.89	1.21	2.01	2.84	3.39	3.79	4.06	4.26	4.42	4.51
<b>Alvéolaire</b>	0.94	1.10	2.06	3.28	3.98	4.62	5.16	5.59	5.97	6.17
<b>Vélaire</b>	1.84	2.30	3.02	3.69	4.32	4.89	5.43	5.88	6.14	5.93
<b>Bilabiale vs. Alvéolaire</b>	0.79	0.58	0.85	0.09	0.06	0.02 *	0.005 **	0.001 **	0.000 1***	5.418 e-05 ***
<b>Bilabiale vs. Vélaire</b>	0.002 **	0.003 **	0.01 *	0.05	0.03 *	0.01 *	0.001 **	0.000 1***	2.125 e-05 ***	0.0003 ***
<b>Vélaire Vs. Alvéolaire</b>	0.005 **	0.001 **	0.01 *	0.32	0.41	0.52	0.49	0.44	0.62	0.4956

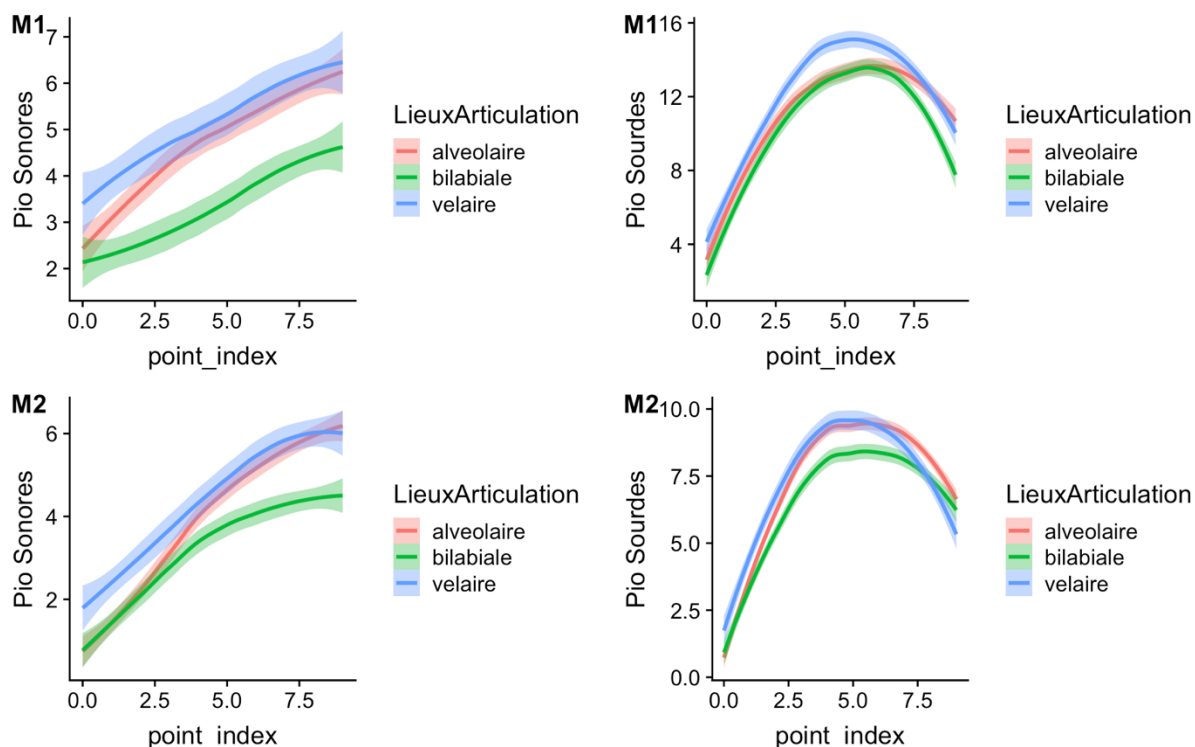


Figure 24 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les trois lieux d'articulation pour les sujets masculins.

Pour le sujet F1, il n'y a pas de différence majeure attestée entre les mesures de PIO des différents lieux d'articulation pour les consonnes non voisées (cf. tableau 14). Les seules différences observables sont entre les bilabiales par rapport aux alvéolaires pour les points d'extraction 0 à 2 avec une différence hautement significative uniquement pour les points 1 et 2. Il y a également une différence entre les bilabiales et les vélares pour les points 0 à 3 avec un niveau de significativité assez faible. Pour les consonnes voisées, les différences sont trop minimes pour être prises en compte (cf. tableau 15). Pour le sujet F2, on constate qu'il y a une différence significative entre les bilabiales et les vélares pour les consonnes non voisées des points 3 à 9 mais ces différences restent encore assez limitées (cf. tableau 16). En revanche, pour les consonnes voisées (cf. tableau 17), il existe une différence significative entre les bilabiales et les vélares et également entre les vélares et les alvéolaires. Cette fois-ci les différences sont attestées à tous les points d'extraction sauf le point 0. Les différences les plus



marquées sont aux points 2, 6, 7, 8 et 9 avec une différence hautement significative (p-value < 0,001). On constate pour les deux sujets féminins, que les mesures de PIO des consonnes voisées sont plus faibles pour les bilabiales par rapport aux alvéolaires et vélares et que les vélares ont les mesures de PIO les plus élevées. En revanche, cette différence n'est pas attestée pour les occlusives non voisées pour le sujet F2 dont les valeurs de PIO des bilabiales sont plus élevées que les celles des vélares et des alvéolaires (cf. figure 22).

Pour le sujet M1, une différence est attestée entre les occlusives bilabiales et alvéolaires ainsi qu'entre les occlusives bilabiales et vélares pour les consonnes voisées et non voisées avec de meilleurs résultats pour les consonnes voisées (cf. tableau 18, 19). Pour le sujet M2, les différences sont un peu moins marquées mais on retrouve le même pattern que pour le sujet M1, il y a une différence significative entre les mesures de PIO des occlusives bilabiales et alvéolaires et entre les bilabiales et les vélares pour les consonnes voisées et non voisées avec de meilleurs résultats pour les consonnes voisées. On retrouve un autre *pattern* stable pour les mesures de PIO pour les sujets masculins. En effet, on peut observer que les mesures de PIO sont inférieures pour les bilabiales par rapport aux alvéolaires et vélares pour les consonnes voisées et non voisées. De plus, les vélares ont les valeurs de PIO les plus hautes.

Pour résumer les résultats obtenus, on constate que les lieux d'articulation qui sont distinguables par les mesures de PIO sont les bilabiales par rapport aux vélares ainsi que les bilabiales par rapport aux alvéolaires. Ces différences sont plus accentuées lorsqu'il s'agit de consonnes voisées sauf pour le sujet F1 qui obtient de meilleurs résultats pour les consonnes non voisées. La différence entre les occlusives vélares par rapport aux alvéolaires n'est pas

attestée dans les mesures de PIO. Les consonnes bilabiales ont les mesures de PIO les moins élevées et les vélares les plus élevées. Ces patterns sont plus stables pour les sujets masculins que féminins. Les lieux d'articulation semblent donc effectivement avoir une incidence sur les mesures de PIO mais il faudra ajouter des descripteurs supplémentaires pour distinguer les différents lieux d'articulation lors de l'étape de classification automatique car aux regards des résultats nous n'avons pas suffisamment d'invariant stable.

### 5.2.2 Pression sous-glottique

Le principal paramètre physiologique de la production de la parole est la pression sous-glottique (PSG) qui correspond à la pression créée par le flux d'air expiré par les poumons contre les plis vocaux. Nous souhaitons dans cette partie analyser ces variations afin de comprendre l'implication de la pression sous-glottique dans la production des consonnes occlusives du français. Pour cela, nous suivrons la même méthodologie que celle appliquée pour l'analyse de la pression intra-orale afin d'analyser les mêmes descripteurs pour être en mesure de confronter les résultats obtenus et de comparer les deux paramètres. De plus, cela nous permettra d'analyser si la pression sous-glottique permet de renforcer la détection de *patterns* déjà trouvés ou au contraire si elle permet de détecter de nouveaux *patterns* qui n'étaient pas encodés dans les mesures de pression intra-orale pour la classification automatique des lieux d'articulation. Nous pensons qu'il est essentiel d'étudier ces deux paramètres successivement car ils sont impliqués simultanément dans la production de la parole.

### 5.2.2.1 L'opposition de voisement

Tableau 23 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet F1. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PSG pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>NV</b> 8	11.5	11.50	11.60	11.80	11.89	11.93	11.80	11.54	11.54	11.30
<b>V</b> 1	11.6	11.58	11.52	11.50	11.49	11.48	11.47	11.50	11.50	11.54
<b>P</b>	0.91	0.76	0.77	0.27	0.14	0.09	0.19	0.87	0.87	0.22

Tableau 24 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet F2. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PSG pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>NV</b>	7.56	7.42	7.50	7.50	7.47	7.47	7.50	7.54	7.52	7.35
<b>V</b>	7.81	7.76	7.73	7.69	7.65	7.62	7.59	7.57	7.54	7.51
<b>P</b>	0.16	0.05	0.19	0.27	0.30	0.38	0.57	0.85	0.86	0.32

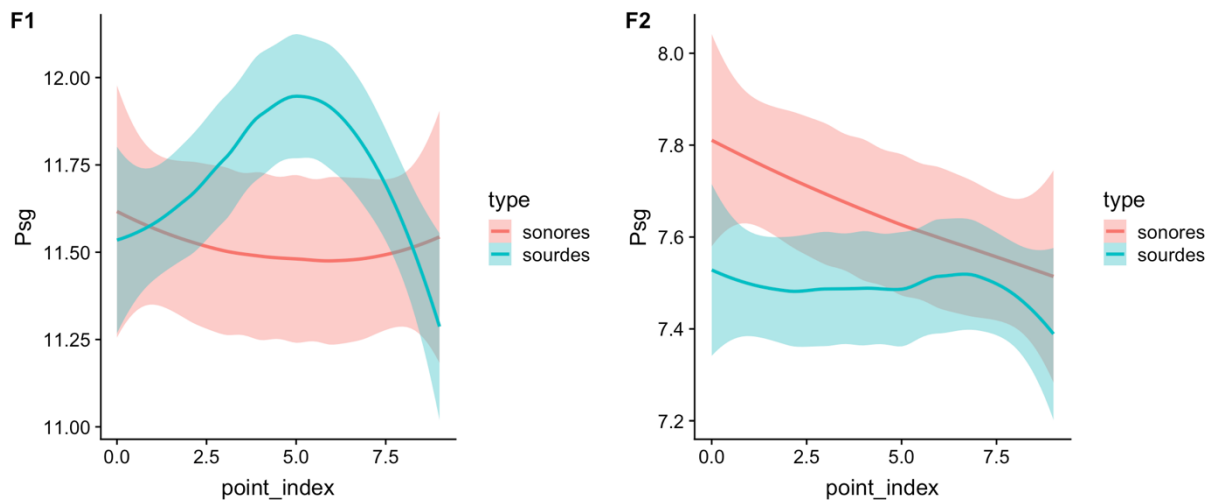


Figure 25 : Courbes des valeurs moyennes des mesures de Psg (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction.

Tableau 25 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet M1. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

	Moyennes des mesures de PSG pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
NV	15.3	15.47	15.52	15.53	15.52	15.45	15.33	15.17	14.92	14.63
V	15.3	15.52	15.72	15.90	16.04	16.14	16.22	16.27	16.29	16.29
P	0.95	0.91	0.68	0.44	0.25	0.10	0.02	0.004	0.000	3.842 <sup>e</sup> -06 <sup>***</sup>

Tableau 26 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet M2. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=900).

		Moyennes des mesures de PSG pour le sujet M2									
		0	1	2	3	4	5	6	7	8	9
NV		9.36	9.15	9.20	9.31	9.46	9.53	9.57	9.55	9.44	9.13
V	1	11.2	11.15	11.12	11.10	11.06	11.02	10.96	10.89	10.82	10.72
P		3.08 e-08 ***	3.39 <sup>e</sup> -09 ***	2.205 <sup>e</sup> -09 ***	4.457 <sup>e</sup> -09 ***	3.478 <sup>e</sup> -08 ***	8.376 <sup>e</sup> -08 ***	1.859 <sup>e</sup> -07 ***	2.056 <sup>e</sup> -07 ***	7.185 <sup>e</sup> -08 ***	1.304 <sup>e</sup> -09 ***

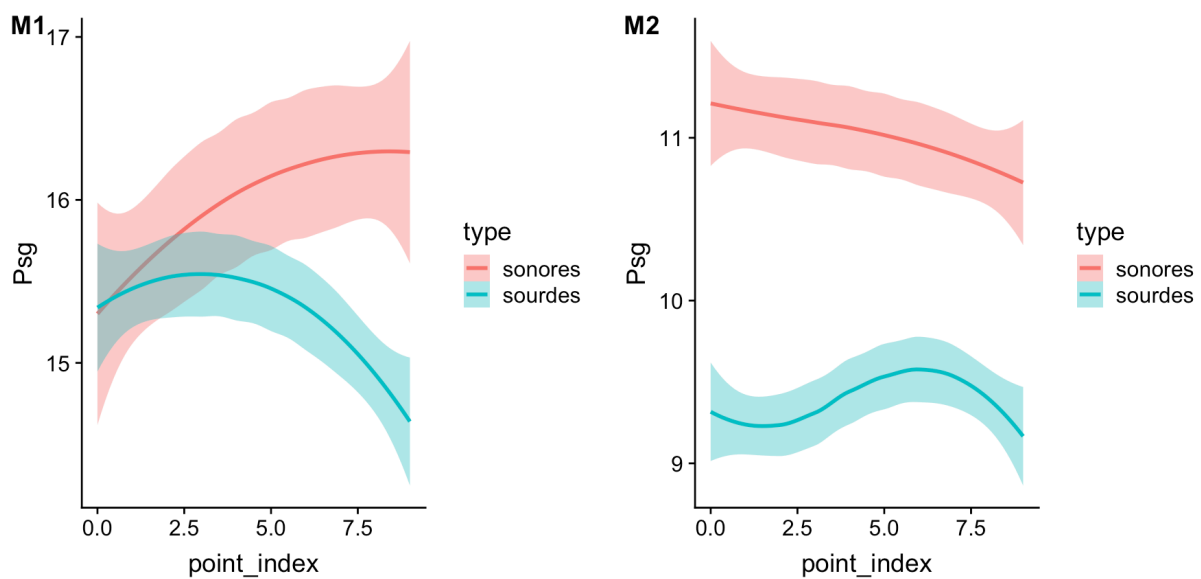


Figure 26 : Courbes des valeurs moyennes des mesures de P<sub>sg</sub> (hpa) pour les consonnes non voisées en bleu (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction.

Les sujets féminins ont des mesures de PSG quasiment similaires entre consonnes voisées et non voisées. En effet, on constate qu'il n'y a aucune différence significative attestée pouvant marquer l'opposition de voisement. En revanche, les résultats sont différents pour les sujets masculins puisqu'on observe une différence significative entre consonnes voisées et non voisées des points 6 à 9 pour le sujet M1 et à tous les points d'extraction pour le sujet M2. On observe également que les sujets masculins ont des mesures de PSG supérieures à celles des sujets féminins. Il semblerait donc que le sexe a une influence sur les mesures de PSG. On peut néanmoins observer un *pattern* commun entre les sujets féminins et masculins : les consonnes voisées ont des valeurs de PSG légèrement supérieures à celles des non voisées. Il s'agit d'un résultat attendu car pour les consonnes voisées, l'écoulement d'air est ralenti au niveau de la glotte ce qui provoque une résistance plus grande à l'écoulement d'air et a pour conséquence d'augmenter légèrement la PSG sous la glotte. En revanche, le résultat inverse est observé pour le sujet F1 (cf. figure 24). Il est assez étonnant de constater que la courbe moyenne des consonnes non voisées est assimilable à celle obtenue lors de l'analyse de la pression intra-orale avec une montée progressive après la fermeture des lèvres, puis la présence d'un plateau très légèrement ascendant suivi d'une chute rapide. On s'attend pour la PSG à une courbe plus linéaire car une pression sous-glottique plus ou moins constante doit être maintenue pour la puissance de parole. Cette courbe suggère que nous sommes face à un paramètre actif et non passif.

Contrairement à la pression intra-orale, les mesures de PSG sont beaucoup plus variables pour l'opposition de voisement. Nous pensons, aux regards des résultats, qu'il peut s'agir d'un descripteur pertinent pour la classification automatique uniquement pour les sujets masculins.

### 5.2.2.2 Le contexte intervocalique

Tableau 27 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	10.65	10.64	10.95	11.32	11.42	11.50	11.55	11.47	11.07	10.67
<b>i</b>	12.80	12.66	12.69	12.79	12.79	12.73	12.63	12.48	12.31	12.16
<b>u</b>	11.30	11.19	11.17	11.30	11.45	11.56	11.59	11.45	11.23	11.06
<b>a vs. i</b>	1.767 e-08 ***	1.473 e-07 ***	9.593 e-06 ***	0.000 1***	0.0003* **	0.004 **	0.006 **	0.006 **	0.0005 ***	3.024 e-05 ***
<b>a vs. u</b>	0.03 *	0.07	0.49	0.95	0.92	0.87	0.90	0.96	0.59	0.15
<b>u vs. i</b>	8.553 e-06 ***	2.29 e-05 ***	2.147 e-05 ***	4.073 e-05 ***	0.0002 ***	0.001 **	0.004 **	0.004 **	0.002 **	0.001 **

Tableau 28 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	11.47	11.48	11.50	11.51	11.52	11.53	11.53	11.54	11.53	11.50
<b>i</b>	11.95	11.86	11.68	11.61	11.58	11.56	11.54	11.54	11.65	11.84
<b>u</b>	11.41	11.39	11.39	11.37	11.37	11.36	11.35	11.34	11.32	11.30
<b>a vs. i</b>	0.29	0.40	0.73	0.85	0.92	0.95	0.98	0.99	0.78	0.28
<b>a vs. u</b>	0.89	0.85	0.80	0.76	0.72	0.64	0.64	0.60	0.57	0.57
<b>u vs. i</b>	0.38	0.44	0.64	0.71	0.74	0.75	0.75	0.73	0.50	0.15

Tableau 29 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	6.48	6.31	6.48	6.45	6.46	6.47	6.51	6.56	6.61	6.51
<b>i</b>	8.01	7.88	7.93	7.91	7.86	7.84	7.86	7.89	7.79	7.65
<b>u</b>	8.21	8.08	8.12	8.15	8.13	8.12	8.14	8.18	8.17	7.92
<b>a vs. i</b>	6.469e-11***	6.257e-12***	1.872e-10***	4.132e-11***	7.867e-11***	8.464e-11***	5.565e-11***	5.413e-11***	5.757e-10***	1.275e-09***
<b>a vs. u</b>	1.336e-12***	3.713e-13***	1.266e-11***	5.268e-12***	5.596e-12***	4.968e-12***	3.089e-12***	1.805e-12***	1.53e-12***	3.171e-12***
<b>u vs. i</b>	0.27	0.27	0.34	0.23	0.19	0.16	0.16	0.14	0.04*	0.13

Tableau 30 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	6.18	6.11	6.09	6.07	6.05	6.05	6.06	6.06	6.07	6.08
<b>i</b>	8.16	8.14	8.12	8.10	8.07	8.05	8.03	8.01	7.99	7.95
<b>u</b>	8.5	8.48	8.42	8.36	8.30	8.24	8.18	8.13	8.08	8.01
<b>a vs. i</b>	2.453e-13***	3.571e-14***	2.623e-14***	1.486e-14***	1.532e-14***	1.31e-14***	1.333e-14***	1.08e-14***	8.992e-15***	1.204e-14***
<b>a vs. u</b>	2.5e-14***	9.192e-15***	1.275e-14***	1.131e-14***	1.743e-14***	2.302e-14***	4.357e-14***	6.027e-14***	7.932e-14***	6.185e-14***
<b>u vs. i</b>	0.07	0.10	0.15	0.20	0.28	0.36	0.46	0.57	0.68	0.77



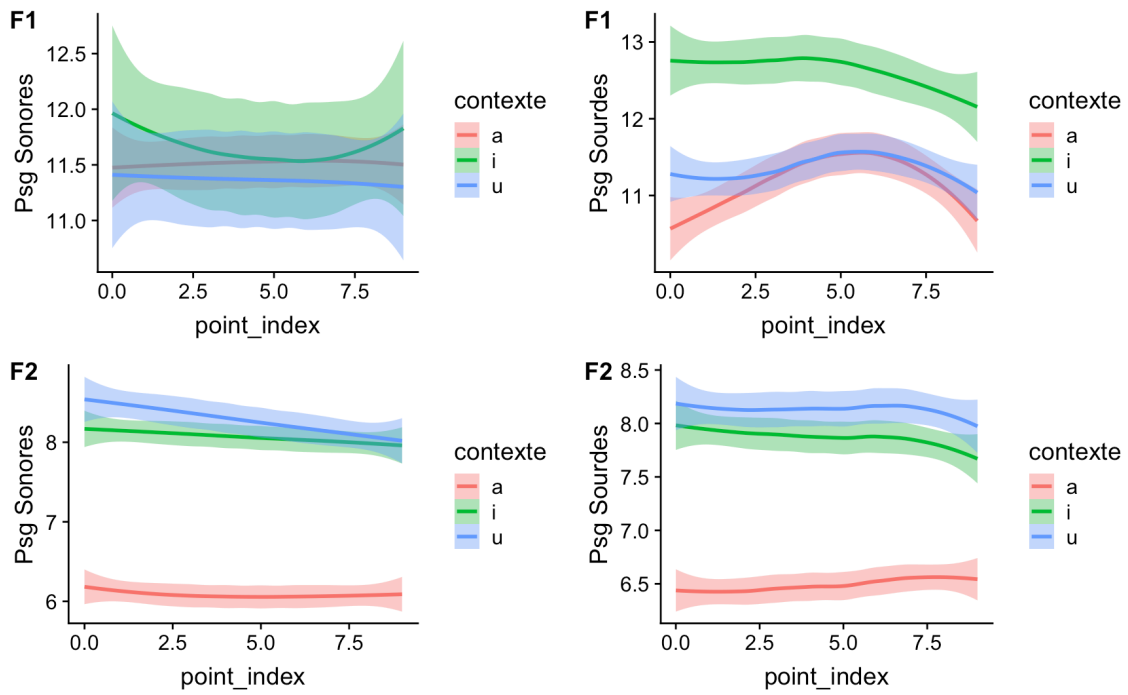


Figure 27 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets féminins.

Les résultats obtenus pour les sujets masculins sont les suivants :

Tableau 31 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	13.51	13.65	13.69	13.70	13.66	13.56	13.44	13.30	13.02	12.67
<b>i</b>	16.91	17.07	17.14	17.15	17.12	17.04	16.89	16.68	16.45	16.16
<b>u</b>	15.57	15.68	15.72	15.76	15.78	15.76	15.67	15.53	15.31	15.08
<b>a vs. i</b>	4.602 e-12 ***	1.675 e-11 ***	2.414 e-11 ***	2.869 e-11 ***	1.938 e-11 ***	6.798 e-12 ***	2.324 e-12 ***	5.128 e-13 ***	2.305 e-14 ***	2.394 e-15 ***
<b>a vs. u</b>	1.171 e-05 ****	1.668 e-05 ***	1.734 e-05 ***	1.145 e-05 ***	4.011 e-06 ***	1.045 e-06 ***	3.116 e-07 ***	1.091 e-07 ***	2.052e-08 ***	2.887e-09 ***
<b>u vs. i</b>	0.006 **	0.006 **	0.005 **	0.006 **	0.007 **	0.008 **	0.008 **	0.008 **	0.005 **	0.006 **

Tableau 32 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	13.82	13.95	14.06	14.16	14.23	14.28	14.31	14.32	14.32	14.30
<b>i</b>	15.55	15.86	16.14	16.37	16.57	16.71	16.82	16.89	16.95	16.96
<b>u</b>	16.52	16.77	16.98	17.17	17.32	17.43	17.52	17.58	17.61	17.61
<b>a vs. i</b>	0.16	0.09	0.05	0.02	0.01	0.004	0.001	0.000	0.0001	6.12
				*	*	**	**	5 ***	***	e-05***
<b>a vs. u</b>	0.02	0.009	0.003	0.000	0.000	3.592	4.095	3.697	2.86	2.467
	*	**	**	9 ***	2 ***	e-05	e-06	e-07	e-08	e-09
						***	***	***	***	***
<b>u vs. i</b>	0.49	0.47	0.47	0.46	0.443	0.42	0.39	0.351	0.33	0.30

Tableau 33 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	6.86	6.56	6.93	7.35	7.67	7.86	7.99	8.02	7.87	7.38
<b>i</b>	10.93	10.73	10.58	10.51	10.55	10.52	10.50	10.46	10.38	10.20
<b>u</b>	10.30	10.16	10.08	10.08	10.16	10.20	10.21	10.16	10.07	9.84
<b>a vs. i</b>	2.2e-16	2.2e-16	2.2e-16	2.2	3.622	3.193	7.08	1.547	2.2e-16	2.2e-16
	***	***	***	e-16	e-16	e-15	e-15	e-15	***	***
				***	***	***	***	***		
<b>a vs. u</b>	2.2e-16	2.2e-16	2.2e-16	2.2	2.2	2.2	2.2	4.917	3.422	2.2e-16
	***	***	***	e-16	e-16	e-16	e-16	e-16	e-16	***
				***	***	***	***	***	***	
<b>u vs. i</b>	0.006	0.01	0.04	0.083	0.124	0.19	0.22	0.19	0.15	0.08
	**	*	*							

Tableau 34 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>a</b>	8.54	8.50	8.54	8.61	8.65	8.70	8.73	8.74	8.71	8.64
<b>i</b>	13.22	13.13	13.05	12.97	12.88	12.77	12.65	12.53	12.42	12.31
<b>u</b>	11.89	11.82	11.77	11.71	11.64	11.57	11.49	11.41	11.32	11.21
<b>a vs. i</b>	2.2e-16 ***	2.2e-16 ***	2.2e-16 16 ***	2.2e-16 16	2.2e-16 16	2.2e-16 16	2.2e-16 16	2.2e-16 16	2.2e-16 16	2.2e-16 16
<b>a vs. u</b>	2.2e-16 ***	2.2e-16 ***	2.2e-16 16***	2.2e-16 16***	2.2e-16 16***	2.2e-16 16***	2.2e-16 16***	2.2e-16 16***	2.2e-16 ***	2.2e-16 ***
<b>u vs. i</b>	0.0005 ***	0.0008 ***	0.000 5***	0.000 5***	0.000 5***	0.000 5***	0.000 6***	0.000 7***	0.0005 ***	0.0003 ***

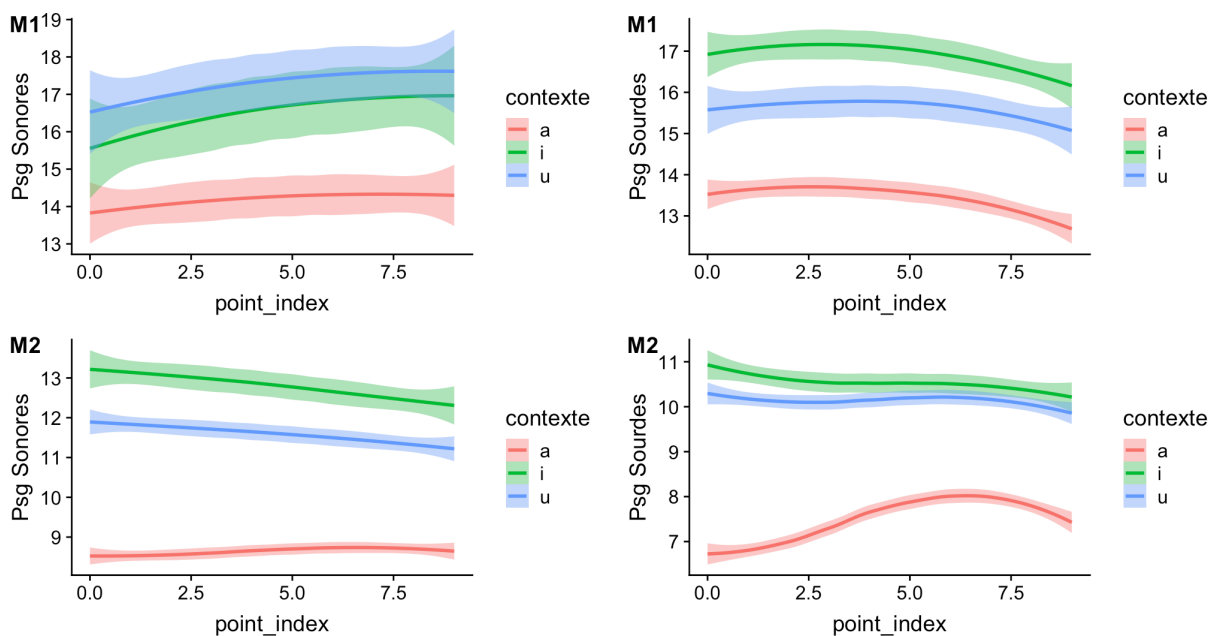


Figure 28 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets masculins.

Tableau 35 : Résumé des différences significatives des mesures de PSG entre les différents contextes intervocaliques.

	Non Voisées	Voisées
F1	[a vs. i] & [u vs. i]	
F2	[a vs. i] & [a vs. u]	[a vs. i] & [a vs. u]
M1	[a vs. i] & [a vs. u] & [i vs. u]	[a vs. i] & [a vs. u]
M2	[a vs. i] & [a vs. u]	[a vs. i] & [a vs. u] & [i vs. u]

Les différents tests effectués nous permettent de constater que le contexte intervocalique a une influence sur les mesures de PSG des quatre sujets d'étude. En effet, les résultats mettent en évidence une différence de pression sous-glottique entre /a/, /i/ et /u/. Dans le contexte intervocalique /a/, les mesures de PSG sont les plus basses. Ce *pattern* est observable pour les sujets F2, M1 et M2 (cf. figure 26, 27) pour les consonnes voisées et non voisées. Pour le sujet F1, ce *pattern* est attesté uniquement pour les consonnes non voisées (cf. figure 26) car les mesures de PSG des consonnes voisées sont extrêmement proches ce qui explique l'absence de différence significative. De plus, les mesures sont plus basses dans le contexte /u/ par rapport à /a/. Cela pourrait être dû au fait que /u/ et /a/ ont une position plus basse du larynx, ce qui réduit leur résistance glottique. On observe la présence d'un invariant stable entre les sujets F2, M1 et M2 : la différence significative entre /a/ et /i/ et /a/ et /u/ pour les consonnes voisées et non voisées (cf. tableau 29). La différence entre /a/ et /u/ n'est pas attestée pour le sujet F1 car comme nous l'avons mentionné les mesures de PSG sont très basses dans le contexte /u/, elles se rapprochent donc des mesures dans le contexte /a/. On constate qu'il y a également une différence entre /i/ et /u/ pour les sujets F1, M1 et M2 (cf. figure 29). On observe une légère différence entre les sujets féminins et masculins, les sujets masculins présentent une différence entre les trois contextes intervocaliques [a vs. i], [a vs. u]

et [i vs. u] pour les consonnes non voisées pour le sujet M1 et pour les consonnes voisées pour le sujet M2. Une fois encore, le sexe semble avoir une influence sur les mesures de PSG.

Ces différences entre les contextes intervocaliques ne sont pas aisées à expliquer, il peut s'agir comme pour la pression intra-orale d'une différence due à l'écoulement d'air qui s'échappe plus ou moins rapidement en fonction de la forme du conduit vocal. Sachant que la pression sous-glottique correspond à la pression créée par le flux d'air expiré par les poumons contre les plis vocaux, ces différences peuvent également être liées à un contrôle respiratoire différent ou à une tension plus ou moins élevée des plis vocaux entre les sujets.

Lors de l'analyse de la pression intra-orale, nous avons constaté que les *patterns* observés étaient plus stables pour les consonnes voisées. Dans le cas de la PSG, il n'y pas de différence dans les *patterns* observés entre les mesures de PSG des consonnes voisées et non voisées. Cette observation est donc dans la continuité des résultats observés dans la partie précédente où il n'y avait pas de différence majeure entre les deux types de consonnes. Contrairement aux résultats obtenus pour l'opposition de voisement, la PSG est un descripteur robuste pour la détection du contexte intervocalique.

### 5.2.1.3 Les lieux d'articulation

Tableau 36 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	11.21	11.10	11.14	11.33	11.49	11.58	11.63	11.57	11.45	11.30
<b>Alvéolaire</b>	11.83	11.86	12.05	12.23	12.31	12.36	12.36	12.23	11.88	11.57
<b>Vélaire</b>	11.70	11.52	11.61	11.85	11.86	11.84	11.78	11.60	11.28	11.02
<b>Bilabiale vs. Alvéolaire</b>	0.12	0.05	0.01 *	0.01 *	0.027	0.035	0.05	0.06	0.23	0.47
<b>Bilabiale vs. Vélaire</b>	0.23	0.30	0.25	0.20	0.36	0.52	0.71	0.93	0.66	0.45
<b>Vélaire Vs. Alvéolaire</b>	0.70	0.33	0.22	0.28	0.18	0.12	0.08	0.05	0.05	0.06

Tableau 37 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet F1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	12.03	12.01	12.00	11.99	11.98	11.98	11.98	11.97	11.95	11.91
<b>Alvéolaire</b>	11.25	11.26	11.19	11.21	11.27	11.33	11.38	11.45	11.61	11.86
<b>Vélaire</b>	11.54	11.46	11.38	11.29	11.21	11.13	11.06	11.00	10.93	10.85
<b>Bilabiale vs. Alvéolaire</b>	0.20	0.20	0.20	0.22	0.25	0.28	0.31	0.35	0.45	0.87
<b>Bilabiale vs. Vélaire</b>	0.10	0.06	0.03 *	0.02 *	0.01 *	0.006 **	0.003 **	0.002 **	0.001 **	0.0009 ***
<b>Vélaire Vs. Alvéolaire</b>	0.63	0.74	0.77	0.90	0.92	0.75	0.59	0.44	0.15	0.006 **

Tableau 38 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	7.36	7.25	7.29	7.33	7.34	7.38	7.44	7.51	7.54	7.39
<b>Alvéolaire</b>	7.61	7.43	7.54	7.53	7.49	7.48	7.50	7.53	7.49	7.30
<b>Vélaire</b>	7.70	7.57	7.67	7.64	7.60	7.56	7.56	7.57	7.52	7.37
<b>Bilabiale vs. Alvéolaire</b>	0.40	0.54	0.39	0.49	0.60	0.73	0.83	0.93	0.83	0.69
<b>Bilabiale vs. Vélaire</b>	0.19	0.21	0.14	0.24	0.32	0.47	0.63	0.79	0.91	0.90
<b>Vélaire Vs. Alvéolaire</b>	0.76	0.59	0.61	0.67	0.67	0.72	0.79	0.86	0.90	0.73

Tableau 39 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet F2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	7.59	7.56	7.55	7.54	7.53	7.52	7.52	7.53	7.52	7.50
<b>Alvéolaire</b>	8.24	8.19	8.13	8.08	8.03	7.97	7.91	7.84	7.78	7.72
<b>Vélaire</b>	7.73	7.68	7.64	7.58	7.53	7.49	7.45	7.42	7.40	7.37
<b>Bilabiale vs. Alvéolaire</b>	0.05	0.06	0.08	0.09	0.11	0.15	0.21	0.30	0.38	0.44
<b>Bilabiale vs. Vélaire</b>	0.66	0.72	0.79	0.90	0.98	0.92	0.82	0.74	0.70	0.65
<b>Vélaire Vs. Alvéolaire</b>	0.08	0.08	0.09	0.08	0.08	0.09	0.10	0.13	0.16	0.17

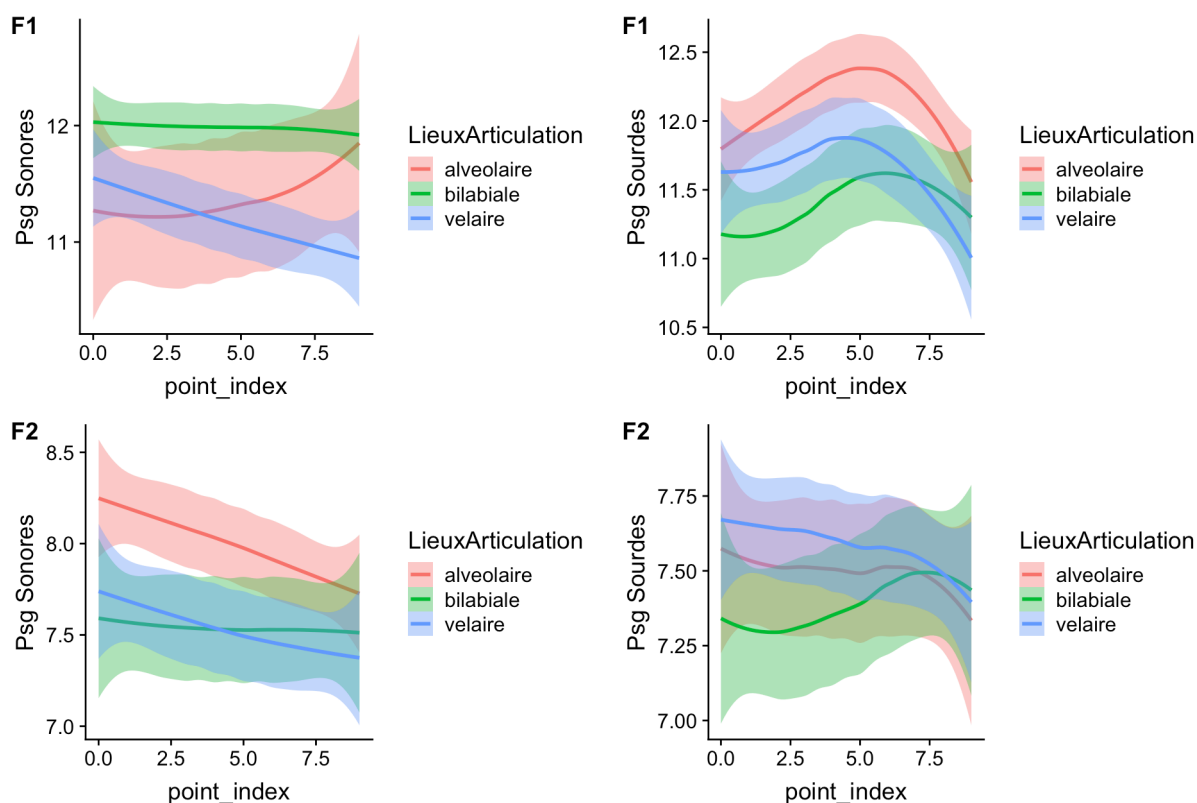


Figure 29 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les trois lieux d'articulation pour les sujets féminins.

Tableau 40 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	14.54	14.61	14.61	14.59	14.54	14.48	14.39	14.29	14.17	13.90
<b>Alvéolaire</b>	15.39	15.58	15.64	15.69	15.68	15.63	15.53	15.36	15.12	14.80
<b>Vélaire</b>	16.05	16.21	16.30	16.32	16.33	16.25	16.07	15.85	15.47	15.21
<b>Bilabiale vs. Alvéolaire</b>	0.08	0.05	0.04	0.02	0.02	0.01	0.01	0.01	0.03	0.04
			*	*	*	*	*	*	*	*
<b>Bilabiale vs. Vélaire</b>	0.01	0.008	0.005	0.004	0.002	0.002	0.002	0.003	0.01	0.01
	*	**	**	**	**	**	**	**	*	*
<b>Vélaire vs. Alvéolaire</b>	0.23	0.26	0.25	0.27	0.26	0.28	0.32	0.36	0.51	0.43



Tableau 41 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet M1									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	10.68	11.35	11.98	12.59	13.16	13.68	14.17	14.60	15.02	15.40
<b>Alvéolaire</b>	17.70	17.70	17.66	17.61	17.52	17.40	17.25	17.08	16.90	16.67
<b>Vélaire</b>	17.51	17.53	17.53	17.49	17.43	17.34	17.24	17.11	16.97	16.79
<b>Bilabiale vs. Alvéolaire</b>	1.326e-08 ***	2.163e-08 ***	5.449e-08 ***	2.313e-07 ***	1.52e-06 ***	1.179e-05 ***	0.0001122 ***	0.001016 ***	0.00949 **	0.07
<b>Bilabiale vs. Vélaire</b>	5.844e-09 ***	5.545e-09 ***	8.255e-09 ***	2.724e-08 ***	1.648e-07 ***	1.505e-06 ***	1.875e-05 ***	0.0002523 ***	0.003305 **	0.03307 *
<b>Vélaire Vs. Alvéolaire</b>	0.82	0.84	0.87	0.88	0.91	0.94	0.99	0.97	0.92	0.86

Tableau 42 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes NV pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	10.38	10.31	10.28	10.26	10.21	10.16	10.11	10.06	10.00	9.94
<b>Alvéolaire</b>	11.51	11.45	11.43	11.41	11.37	11.33	11.26	11.18	11.09	10.97
<b>Vélaire</b>	11.75	11.69	11.65	11.62	11.59	11.55	11.50	11.43	11.36	11.24
<b>Bilabiale vs. Alvéolaire</b>	0.04 *	0.04 *	0.03 *	0.02 *	0.02 *	0.01 *	0.01 *	0.01 *	0.01 *	0.01 *
<b>Bilabiale vs. Vélaire</b>	0.02 *	0.01 *	0.01 *	0.01 *	0.01 *	0.007 **	0.005 **	0.004 **	0.004 **	0.005 **
<b>Vélaire Vs. Alvéolaire</b>	0.69	0.69	0.70	0.72	0.70	0.69	0.64	0.61	0.58	0.57

Tableau 43 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ) (N=30).

	Moyennes des mesures de PSG pour les consonnes V pour le sujet M2									
	0	1	2	3	4	5	6	7	8	9
<b>Bilabiale</b>	8.68	8.39	8.38	8.49	8.64	8.74	8.83	8.90	8.91	8.73
<b>Alvéolaire</b>	9.58	9.38	9.52	9.64	9.77	9.83	9.90	9.91	9.85	9.45
<b>Vélaire</b>	9.83	9.67	9.69	9.81	9.96	9.995	9.96	9.83	9.56	9.23
<b>Bilabiale vs. Alvéolaire</b>	0.06	0.04 *	0.008 **	0.002 **	0.001 **	0.001 **	0.001 **	0.001 **	0.003 **	0.03 *
<b>Bilabiale vs. Vélaire</b>	0.03 *	0.02 *	0.009 **	0.003 **	0.001 **	0.001 **	0.002 **	0.01 *	0.08	0.21
<b>Vélaire Vs. Alvéolaire</b>	0.61	0.57	0.71	0.67	0.62	0.67	0.85	0.82	0.41	0.57

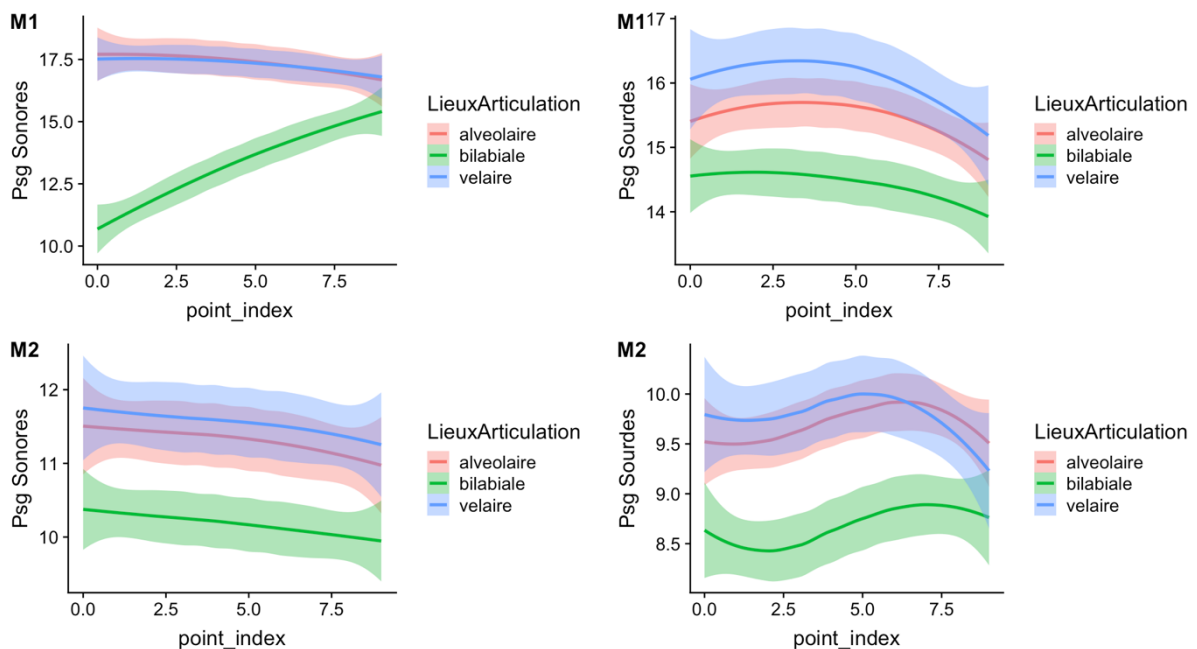


Figure 30 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les trois lieux d'articulation pour les sujets masculins.

La différence entre les lieux d'articulation n'est pas démontrée dans les mesures de PSG pour le sujet F1 pour les consonnes non voisées et pour le sujet F2 pour les deux types de consonnes. Pour le sujet F1, une différence significative est attestée entre les bilabiales et les vélares des points 2 à 9. Pour les sujets masculins, les différences entre les lieux d'articulation sont plus marquées avec des meilleurs résultats pour les consonnes sonores. Pour le sujet M1, il y a une différence entre les bilabiales et les alvéolaires et entre les bilabiales et les vélares pour les consonnes non voisées et voisées avec une p-value  $< 0,001$  pour les consonnes voisées. Pour le sujet M2, les mêmes différences sont démontrées mais elles sont moins marquées que pour le sujet M1. Une nouvelle fois nous pouvons constater que nous obtenons de meilleurs résultats pour les sujets masculins que féminins. Pour les consonnes non voisées, les valeurs de PSG sont inférieures pour les bilabiales comparées aux vélares et aux alvéolaires, cela est valable pour les 4 sujets et les vélares ont les mesures les plus élevées sauf pour le sujet F1 qui obtient des mesures supérieures pour les consonnes alvéolaires. Pour les consonnes voisées, les sujets masculins ont des valeurs de PSG inférieures pour les bilabiales. Cependant, on observe le *pattern* inverse pour les sujets féminins dont les mesures de PSG sont plus basses pour les vélares. Ces résultats sont similaires avec ceux obtenus lors de l'analyse de PIO avec un *pattern* plus stable pour les sujets masculins qui obtiennent des valeurs plus faibles pour les bilabiales. Une nouvelle fois, comme pour les mesures de PIO, la différence entre les consonnes vélares et alvéolaires n'est pas démontrée.

### 5.2.3 La durée

Nous présentons ici les résultats relatifs à l'effet du voisement sur les durées consonantiques des occlusives puis nous analyserons si la durée peut s'avérer pertinente pour discriminer les différents lieux d'articulation. Nous avons calculé la durée à partir des intervalles temporels issus de l'alignement automatique, les résultats présentés sont en ms.

*Tableau 44 : Résultat du test de Welch comparant les durées en ms des consonnes voisées (V) et non voisées (NV). La p-value (P) est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001) (N=900).*

	V	NV	P
<b>F1</b>	0.06	0.11	2.2e-16 ***
<b>F2</b>	0.12	0.16	2.2e-16 ***
<b>M1</b>	0.08	0.11	2.2e-16 ***
<b>M2</b>	0.10	0.14	2.2e-16 ****

*Tableau 45 : Résultat du test de Welch comparant les durées en ms des consonnes voisées (V) en fonction des lieux d'articulation. La p-value (P) est indiquée avec son niveau de significativité (\* p <0,05; \*\* p <0,01; \*\*\* p <0,001).*

	F1	F2	M1	M2
<b>Bilabiale</b>	0.06	0.12	0.08	0.10
<b>Alvéolaire</b>	0.05	0.11	0.07	0.10
<b>Vélaire</b>	0.07	0.13	0.08	0.09
<b>Bilabiale vs. Alvéolaire</b>	0.003 **	0.28	0.009**	0.16
<b>Bilabiale vs. Vélaire</b>	0.52	0.03 *	0.30	0.72
<b>Vélaire vs. Alvéolaire</b>	0.002 **	0.001 **	0.14	0.14

Tableau 46 : Résultat du test de Welch comparant les durées en ms des consonnes non voisées (NV) en fonction des lieux d'articulation. La p-value (P) est indiquée avec son niveau de significativité (\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ).

	F1	F2	M1	M2
<b>bilabiale</b>	0.12	0.16	0.12	0.14
<b>alvéolaire</b>	0.11	0.17	0.10	0.14
<b>vélaire</b>	0.12	0.17	0.12	0.13
<b>Bilabiale vs. Alvéolaire</b>	0.19	0.39	6.452e-05 ***	0.84
<b>Bilabiale vs. Vélaire</b>	0.48	0.06	0.1705	0.02 *
<b>Vélaire vs. Alvéolaire</b>	0.002 **	0.34	1.402e-06 ***	0.004 **

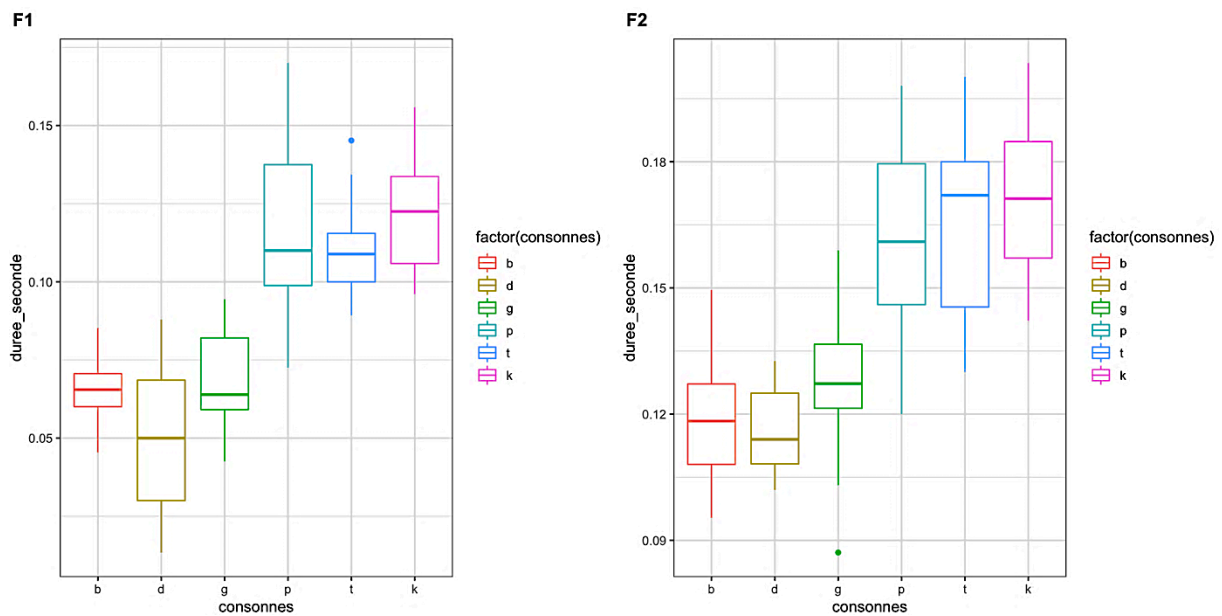


Figure 31 : Diagrammes en boîtes à moustache représentant les durées en ms des consonnes voisées (b,d,g) et non voisées (p,t,k) du français pour le sujet F1 à gauche et F2 à droite.

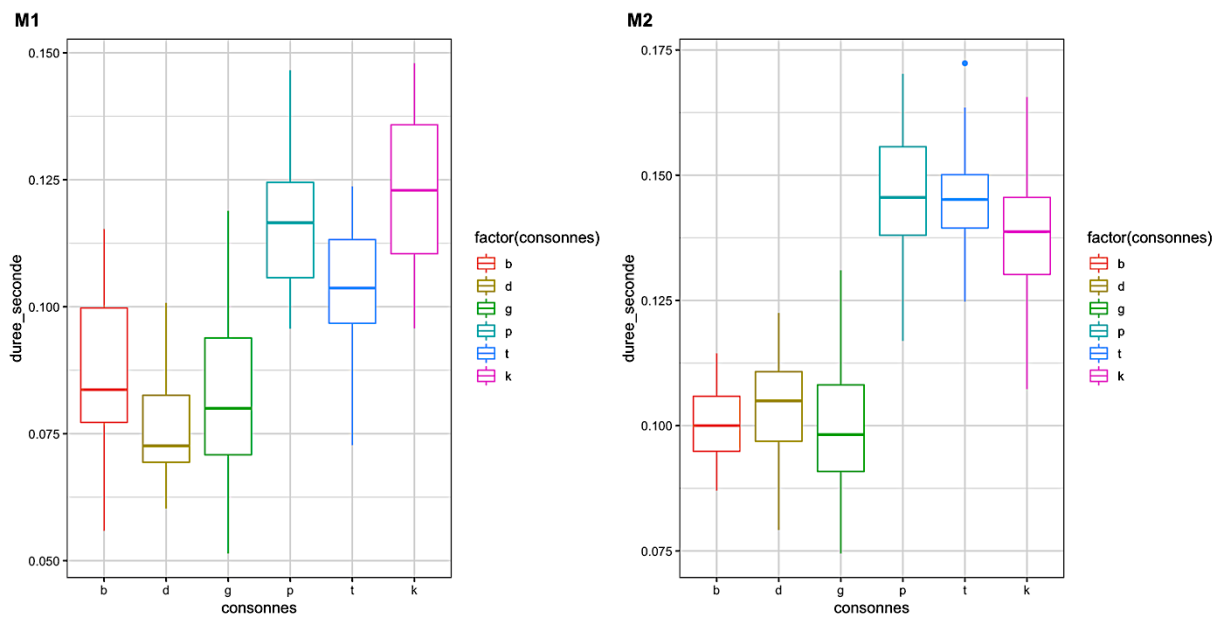


Figure 32 : Diagrammes en boîtes à moustache représentant les durées en ms des consonnes voisées (b,d,g) et non voisées (p,t,k) du français pour le sujet M1 à gauche et M2 à droite.

L'analyse des durées consonantiques met en exergue qu'il existe une différence significative pour tous les sujets entre la durée des consonnes voisées et celle des non voisées. En effet, si on regarde les diagrammes en boîtes à moustache (cf. figure 30 et 31), on constate que les consonnes non voisées sont plus longues que les voisées. En ce qui concerne les lieux d'articulation, il y a une différence de durée entre les bilabiales et alvéolaires pour les sujets F1 et M1 et entre les vélares et les alvéolaires pour les sujets féminins (cf. tableau 37). Pour les consonnes non voisées, il y a une différence de durée pour le sujet M1 entre les bilabiales et alvéolaires et pour les sujets F1, M1 et M2 entre les vélares et les alvéolaires (cf. tableau 38). Ces résultats sont intéressants et pertinents pour la classification automatique de lieux d'articulation car jusqu'ici l'opposition entre vélares et alvéolaires n'avait pas été attestée dans les mesures de pression. La durée est donc un critère pertinent à la fois pour l'opposition de voisement mais aussi pour contribuer à la distinction des lieux d'articulation grâce à l'opposition observée entre les vélares et les alvéolaires.

### 5.3 Classification automatique

A l'issue des analyses réalisées précédemment, nous obtenons pour la classification automatique 8 descripteurs : les sujets (F1, F2, M1, M2), l'opposition de voisement (voisées, non voisées), les points d'extraction (0 à 9), les intervalles temporels, les mesures de PIO, les mesures de PSG, la durée et le contexte intervocalique (a,i,u) ce qui représente un total de 14 879 instances. Dans le cadre de la classification supervisée, les algorithmes implémentés dans WEKA vont utiliser ces descripteurs de manière à entraîner le modèle (phase d'apprentissage) afin de pouvoir par la suite classer un nouveau jeu de données en fonction de la classe prédéfinie : lieux d'articulation (phase de test). Nous souhaitons tester différents algorithmes afin d'observer les différences et les similitudes au niveau des performances en changeant la nature et la quantité des descripteurs. En effet, les analyses réalisées précédemment nous ont permis d'avoir un aperçu de la pertinence de chaque descripteur. Ainsi, nous avons des indications sur ceux qui peuvent servir les performances du système ou au contraire y nuire ce qui nous permettra de comprendre les résultats obtenus et de contribuer à les améliorer en les faisant varier. Cependant, le but est de faire varier ceux qui reflètent un phénomène global (commun au sujet ou au paramètre étudié) et non pas de retirer certaines données des descripteurs afin d'améliorer les résultats. Par exemple, nous avons observé que la PIO et la durée étaient des attributs robustes de l'opposition de voisement contrairement à la PSG. Nous allons donc observer dans quelle mesure l'ajout ou le retrait de ce descripteur affecte les performances du système mais nous n'allons pas juste retirer les données de ce descripteur qui faussent l'analyse et garder celles qui fonctionnent.

Pour classifier automatiquement les données à partir de ces attributs, nous avons testé la totalité des algorithmes implémentés dans le logiciel WEKA (classification bayésienne,

arbres, règles et fonction etc.), avec le but de trouver l'algorithme le plus performant. Pour cela, nous avons utilisé la méthode de validation croisée permettant de découper l'ensemble des exemples en  $n$  sous-ensembles disjoints (par exemple,  $n = 10$ ). Chaque expérience consiste à prendre  $n - 1$  des sous-ensembles pour l'entraînement, et le  $n$ ème sous-ensemble restant pour le test. On obtient ainsi 10 évaluations distinctes. L'avantage de cette méthode est qu'en réalisant l'apprentissage sur 9/10 des exemples, on se prive de peu de données tout en s'assurant de fournir une évaluation peu biaisée car elle est en fait une moyenne de plusieurs expériences. Les résultats obtenus sur la partie « test » sont évalués en termes de classification correcte et de coefficient kappa, nous citerons à chaque fois le pourcentage d'éléments correctement classifiés accompagné du coefficient kappa. Le coefficient kappa mesure ici la concordance entre la classification automatique par WEKA et la classe réelle des lieux d'articulation avec les paramètres la caractérisant. Il varie entre -1 (désaccord total) et 1 (accord total), en passant par 0 (classification au hasard). Pour une classe donnée, connaissant la référence (les éléments annotés) et disposant d'une hypothèse (les sorties d'un système), il est possible de compter : les vrais positifs (nombre d'éléments avec la même étiquette dans la référence et l'hypothèse), les faux positifs ("faux succès", nombre d'éléments déclarés positifs alors qu'ils ne l'étaient pas), les faux négatifs ("faux rejets", nombre d'éléments déclarés négatifs alors qu'ils étaient positifs), et les vrais négatifs (nombre d'éléments correctement déclarés comme négatifs). Ces résultats sont présentés dans les matrices de confusions. Les mesures de rappel, précision et F-mesure synthétisent ces informations. Le rappel mesure la capacité d'un système à sélectionner tous les documents pertinents c'est-à-dire parmi les occurrences pertinentes combien ont été retrouvées ( $VP/VP+FN$ ). Cependant, il ne s'agit pas d'une mesure suffisante à prendre en compte car elle ne montre pas les erreurs du système. En effet, si le système retourne la totalité des instances



sachant que toutes les occurrences pertinentes seront retrouvées, le rappel sera de 1 alors que, parmi ces occurrences, il y en a des non pertinentes. La mesure d'évaluation qui évalue la capacité d'un système à sélectionner uniquement les documents pertinents est la précision. Elle mesure le nombre d'éléments correctement étiquetés par le système (vrais positifs) rapporté au nombre total d'éléments étiquetés par le système (vrais et faux positifs). Si la précision est de 1, cela signifie que le système n'a fait aucune erreur et qu'il retourne uniquement les instances correctement classifiées. Pour finir, la F-mesure, est la moyenne harmonique pondérée du rappel et de la précision. Ces résultats seront présentés sous forme d'histogramme. En fonction des scores de classification et du coefficient kappa, nous sélectionnerons et présenterons les algorithmes permettant la meilleure discrimination des lieux d'articulation. Avant cela, nous testerons également les classes voisement et contexte afin de savoir si les différents descripteurs extraits permettent de les discriminer.

### **5.3.1 L'opposition de voisement**

Nous avons testé la totalité des algorithmes implémentés dans WEKA afin de classer automatiquement les données selon l'opposition de voisement. Les algorithmes ayant obtenus les meilleurs résultats sont : PART et Random forest. L'algorithme PART permet de produire des règles par la génération itérative d'arbres de décision partiel en se basant sur l'algorithme C4.5 et la technique d'apprentissage automatique de règles « diviser pour régner ». Cette technique fonctionne de la manière suivante ; elle découpe un problème initial en sous problème, elle résout ensuite les sous-problèmes (récursivement ou directement) et enfin elle calcule une solution au problème initial à partir des solutions des sous-problèmes. L'algorithme PART procède donc de la façon suivante : il construit une règle en divisant des règles trop

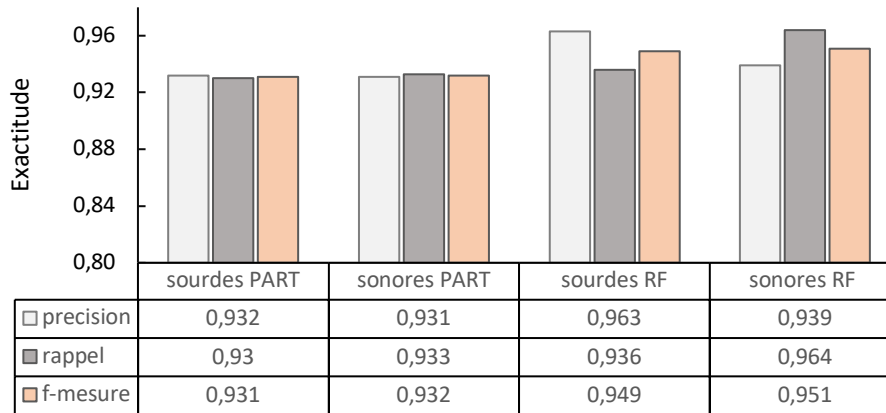
générales en un ensemble de règles qui correspondent à des sous-ensembles de conjonction des exemples. Il continue ensuite récursivement avec ces règles pour lesquelles les sous-ensembles correspondants contiennent à la fois des exemples positifs et négatifs. La règle finale comprend toutes les règles spécialisées pour lesquelles les ensembles correspondants contiennent des exemples positifs. L'algorithme Random forest selon (Shotton, 2013) « effectue un apprentissage en parallèle sur de multiples arbres de décision construits aléatoirement et entraînés sur des sous-ensembles de données différents. Les prédictions sont ensuite moyennées lorsque les données sont quantitatives ou utilisés pour un vote pour des données qualitatives, dans le cas des arbres de classification. L'algorithme des forêts aléatoires est connu pour être un des classifieurs les plus efficaces « out-of-the-box » (c'est-à-dire nécessitant peu de prétraitement des données). »

*Tableau 47 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique de l'opposition de voisement.*

PART		
Classes prédites	Sonores	Sourdes
Vraies Classes		
Sonores	7003	500
Sourdes	517	6859

Random Forest		
Classes prédites	Sonores	Sourdes
Vraies Classes		
Sonores	7234	269
Sourdes	469	6907

### Évaluation de la performance des algorithmes PART & Random Forest

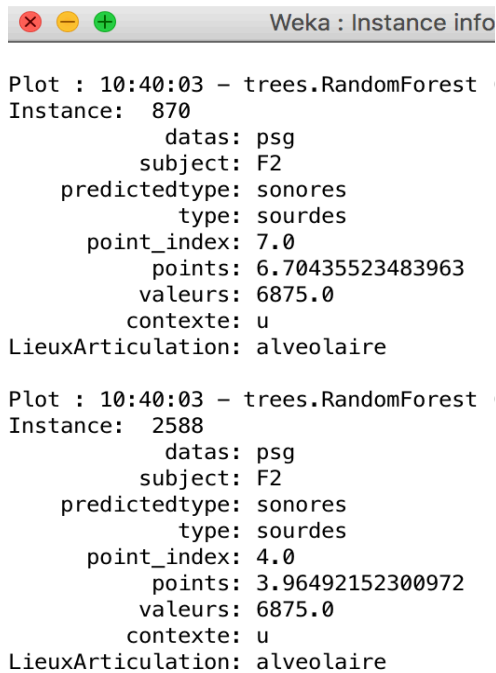


*Figure 33 : Évaluation des performances des algorithmes PART et Random forest (RF) avec la totalité des descripteurs pour la classification automatique de l'opposition de voisement avec les mesures de Précision, Rappel et F-mesure.*

Nous obtenons avec l'algorithme PART pour la phase de test avec la validation croisée, un total de 93,1649 % d'instances correctement classifiées contre 6,8351%, ce qui représente à l'échelle de nos données 13 862/14 879 instances avec un coefficient kappa de 0,8633. Si l'on regarde la matrice de confusion (cf. tableau 39), on constate que le nombre de vrais positifs (VP) pour les consonnes sonores est de 7 003 ce qui correspond aux instances correctement détectées avec une F-mesure de 0,932. En revanche, il y a 500 instances qui ont été déclarées comme sourdes alors qu'ils s'agissaient de consonnes sonores (faux négatifs) et 517 instances déclarées comme étant des consonnes sonores alors qu'il s'agissait de consonnes sourdes (faux positifs). Pour finir, 6 859 instances ont correctement été classifiées comme sourdes avec 517 faux négatifs (FN), 500 faux positifs (FP) et une F-mesure de 0,931. On constate donc que les consonnes sonores ont un taux plus important de VP que les consonnes sourdes mais cette différence reste minime (cf. figure 32).

Avec l'algorithme Random Forest, nous obtenons un total de 95,04 % d'instances correctement classifiées contre 4,96 %, ce qui représente à l'échelle de nos données 14 141/14 879 avec un coefficient kappa de 0,9008. On obtient donc de meilleurs résultats avec cet algorithme. Pour les consonnes sonores, nous obtenons un total de 7 234 instances correctement classifiées et pour les consonnes sourdes 6 907. Les consonnes sourdes obtiennent une meilleure précision que les consonnes sonores bien que le nombre de VP soit plus faible (tableau 32). En effet, il y a 469 faux positifs pour les consonnes sonores contre 269 pour les consonnes sourdes ce qui signifie que plus d'instances sont détectées pour les consonnes sonores (rappel = 0,964 contre 0,936) mais que le système est moins précis puisqu'il retourne plus d'occurrences déclarées comme étant des sonores alors que ce n'était pas le cas (précision = 0.939 contre 0.963). Ces résultats nous montrent l'importance de prendre en compte les mesures de rappel et de précision afin de ne pas tirer de conclusion trop hâtive. En se basant sur les valeurs de F-mesure, qui combinent précision et rappel, les résultats sont très légèrement meilleurs pour les sonores que pour les sourdes (0,951 > 0.949).

Une fois la classification terminée, nous avons regardé les erreurs générées par les algorithmes afin d'identifier la source de celles-ci (cf. figure 33).



```
Plot : 10:40:03 - trees.RandomForest
Instance: 870
  datas: psg
  subject: F2
  predictedtype: sonores
  type: sourdes
  point_index: 7.0
  points: 6.70435523483963
  valeurs: 6875.0
  contexte: u
LieuxArticulation: alveolaire

Plot : 10:40:03 - trees.RandomForest
Instance: 2588
  datas: psg
  subject: F2
  predictedtype: sonores
  type: sourdes
  point_index: 4.0
  points: 3.96492152300972
  valeurs: 6875.0
  contexte: u
LieuxArticulation: alveolaire
```

*Figure 34 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification de l'opposition de voisement.*

Sans surprise, nous constatons que les erreurs proviennent du descripteur PSG (cf. figure 33). Plus particulièrement, les erreurs concernent les sujets F1 et F2. Ce résultat était prévisible car nous avons remarqué lors de l'analyse de la PSG que les sujets féminins avaient des mesures de PSG quasiment similaires entre consonnes voisées et non voisées et qu'il n'y avait aucune différence significative attestée pouvant marquer l'opposition de voisement. Nous avons donc retiré provisoirement ce descripteur de notre corpus afin d'observer si cela nous permet d'améliorer les résultats obtenus bien qu'ils soient déjà très satisfaisants.

Tableau 48 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique de l'opposition de voisement sans la PSG.

PART			Random Forest		
Classes prédites	Sonores	Sourdes	Classes prédites	Sonores	Sourdes
Vraies Classes			Vraies Classes		
Sonores	3759	205	Sonores	3856	108
Sourdes	215	3535	Sourdes	139	3711

Évaluation de la performance des algorithmes  
PART & Random Forest sans la PSG

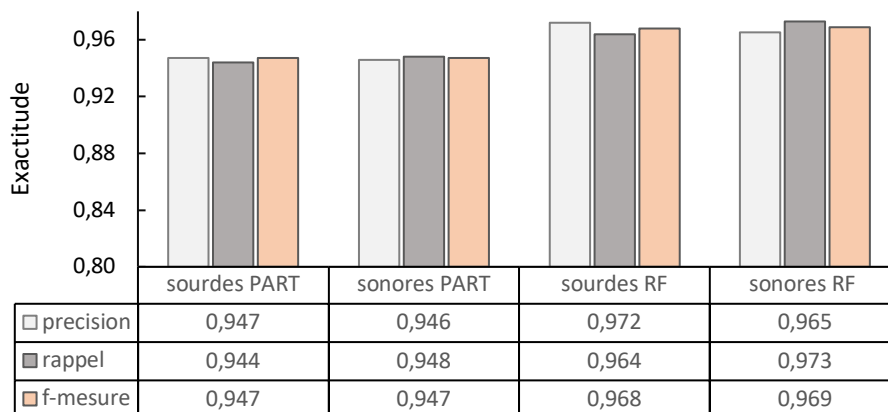


Figure 35 : Évaluation des performances des algorithmes PART et Random forest (RF) sans les mesures de PSG pour la classification automatique de l'opposition de voisement avec les mesures de Précision, Rappel et F-mesure.

Nous obtenons avec l'algorithme PART sans le descripteur PSG, un total de 94,625 % d'instances correctement classifiées contre 5,375 % (7 394/7 814), avec un coefficient kappa de 0,8925. Avec l'algorithme Random Forest, nous obtenons un total de 96,839 % des instances correctement classifiés contre 3,161 % (7 567/7 814) avec un coefficient kappa de

0,9368. Une nouvelle fois, nous obtenons de meilleurs résultats avec l'algorithme Random Forest qui semble être le plus pertinent pour la classification de données aérodynamiques.

La suppression du descripteur PSG s'avère pertinente puisqu'on passe pour l'algorithme le plus performant de 95,04 % à 96,839 % de réussite pour la classification de l'opposition de voisement. De plus, cela nous a permis d'améliorer la précision pour les consonnes sourdes et sonores avec toujours une précision plus élevée pour les consonnes sourdes (Précision = 0,965 < 0,972). La pression intra-orale apparaît donc comme le descripteur le plus robuste pour détecter l'opposition de voisement avec un niveau de réussite très satisfaisant. Ces résultats permettent de confirmer les analyses réalisées précédemment. La durée s'avère également être un descripteur pertinent car si on la retire les performances du système sont moins bonnes mais globalement la plupart des occurrences sont classifiées grâce à la PIO.

### **5.1.2 Le contexte intervocalique**

Les algorithmes ayant obtenus les meilleurs résultats pour la classification automatique du contexte intervocalique sont Random forest et BFTree. BFTree (best first tree) est un algorithme d'arbre de décision. Dans un arbre best-first, un attribut est placé à la racine de l'arbre et des branches sont créées en fonction de plusieurs critères. Les instances sont divisées en sous-ensembles, un pour chaque branche. Le processus est répété pour la branche avec le "meilleur" sous-ensemble en utilisant à chaque fois uniquement les attributs qui atteignent la branche. Le processus de construction s'arrête lorsque tous les nœuds sont purs.

Tableau 49 : Matrices de confusion des algorithmes BFTree et Random forest pour la classification automatique du contexte intervocalique.

BFT				Random Forest			
Classes prédites	a	i	u	Classes prédites	a	i	u
Vraies Classes				Vraies Classes			
a	4436	173	97	a	4577	97	32
i	244	4138	365	i	144	4439	164
u	90	314	4316	u	51	249	4420

Évaluation de la performance des algorithmes  
BFTree & Random Forest

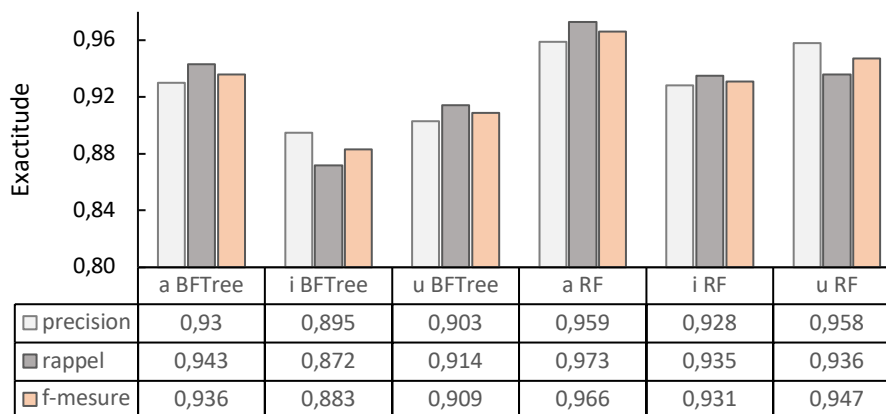


Figure 36 : Évaluation des performances des algorithmes BFTree et Random forest (RF) avec tous les paramètres sauf la durée pour la classification automatique du contexte intervocalique avec les mesures de Précision, Rappel et F-mesure.

Pour la classification automatique du contexte intervocalique, nous n'avons pas pris en compte la durée car le contexte intervocalique n'était pas précisé pour ce paramètre. Nous obtenons donc comme résultat avec tous les paramètres sauf la durée pour l'algorithme BFTree un total de 90,9476 % des instances correctement classifiés contre 9,0524 %



(12 890/14 173) avec un coefficient kappa de 0,8642. Le contexte qui obtient les meilleurs résultats est le contexte /a/ avec une F-mesure de 0,93 et 4 436 instances correctement détectées suivi du contexte /u/ et ensuite du contexte /i/ qui est le contexte le moins bien détecté (précision, rappel et F-mesure inférieurs à 0,9).

Avec l'algorithme Random Forest, nous obtenons un total de 94,8 % des instances correctement classifiées contre 5,2 % (13 436/14 173) avec un coefficient kappa de 0,9675. On constate une nouvelle fois que le contexte intervocalique ayant les meilleurs résultats de classification est le /a/ et celui ayant les moins bonnes performances et le /i/, bien que les mesures de précision, rappel et F-mesure soient meilleures. Ces résultats sont intéressants car ils montrent que le contexte intervocalique a une forte influence sur les mesures de PSG et de PIO. Ce fait avait déjà été constaté lors des analyses des mesures de pression où nous avons remarqué que les mesures de PSG et de PIO étaient inférieures dans le contexte /a/ par rapport à /i/ et /u/ et qu'il existait une différence significative entre les différents contextes intervocaliques. Cependant, nous ne pensions pas obtenir d'aussi bons résultats de classification ce qui met en exergue l'importance de prendre en compte le contexte intervocalique qui a une forte influence sur les mesures des segments consonantiques.

Lorsque nous avons regardé les erreurs générées par les algorithmes nous avons remarqué qu'elles provenaient des mesures de PIO notamment celles concernant la prédiction du /i/ (cf. figure 36). On remarque que l'algorithme Random Forest prédit un /u/ à la place du /i/ lorsqu'il se base sur les mesures de PIO. Si on se reporte aux résultats obtenus lors de l'analyse de la pression intra-orale, nous avons constaté que seul le sujet F2 marquait une opposition hautement significative entre ces deux contextes pour les consonnes non voisées.

Les erreurs de l'algorithme se portent majoritairement sur les sujets masculins pour lesquels aucune différence n'avait été attesté et sur les consonnes sonores (cf. figure 36). En revanche, nous avons obtenu des résultats plus satisfaisants pour les mesures de PSG avec une différence hautement significative entre la plupart des contextes notamment entre /u/ et /i/ qui semblent générer le plus d'erreurs ici. Nous avons donc retiré provisoirement ce paramètre afin de tester les performances des différents algorithmes sans la PIO.



```
Weka : Instance info
Plot : 11:37:34 - trees.RandomForest (CF-weka.filters.unsupervised.attrib
Instance: 12982
  datas: pio
  subject: M2
  type: sonores
  point_index: 3.0
  points: 8.05295782597565
  valeurs: 3.857421875
predictedcontexte: u
contexte: i
LieuxArticulation: bilabiale

Plot : 11:37:34 - trees.RandomForest (CF-weka.filters.unsupervised.attrib
Instance: 12988
  datas: pio
  subject: M2
  type: sonores
  point_index: 2.0
  points: 4.06309794015693
  valeurs: 2.568359375
predictedcontexte: u
contexte: i
LieuxArticulation: alveolaire

Plot : 11:37:34 - trees.RandomForest (CF-weka.filters.unsupervised.attrib
Instance: 13041
  datas: pio
  subject: M2
  type: sonores
  point_index: 2.0
  points: 8.23359992100184
  valeurs: 2.802734375
predictedcontexte: u
contexte: i
LieuxArticulation: alveolaire

Plot : 11:37:34 - trees.RandomForest (CF-weka.filters.unsupervised.attrib
Instance: 13051
  datas: pio
  subject: F2
  type: sonores
  point_index: 8.0
  points: 3.81746782800494
  valeurs: 2.4072265625
predictedcontexte: u
contexte: i
LieuxArticulation: bilabiale
```

Figure 37 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification du contexte intervocalique.

Tableau 50 : Matrices de confusion des algorithmes BFTree et Random forest pour la classification automatique du contexte intervocalique avec la PSG uniquement.

BFTree				Random Forest			
Classes prédites	a	i	u	Classes prédites	a	i	u
Vraies Classes				Vraies Classes			
a	2254	36	15	a	2280	18	7
i	44	2278	74	i	30	2328	38
u	21	78	2306	u	5	56	2344

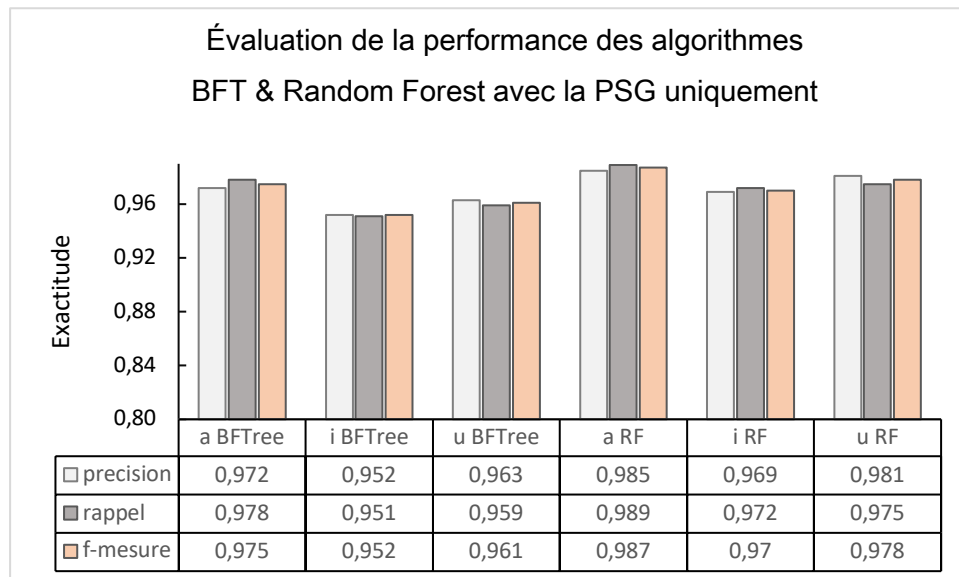


Figure 38 : Évaluation des performances des algorithmes BFTree et Random forest (RF) avec la PSG uniquement pour la classification automatique du contexte intervocalique avec les mesures de Précision, Rappel et F-mesure.

Nous obtenons avec l'algorithme BFTree un total de 96,2285 % des instances correctement classifiés contre 3,7715 % (6 838/7 110) avec un coefficient kappa de 0,9434. Avec l'algorithme Random Forest, nous obtenons un total de 97,8328 % des instances correctement classifiées contre 2,1672 % (6 952/7 110) avec un coefficient kappa de 0,9675. Nous constatons donc que nous parvenons à considérablement augmenter les performances

lorsqu'on nous retirons les mesures de PIO puisque nous passons de 94,8 % à 97,8328 % avec l'algorithme Random Forest. Le contexte /a/ est toujours celui qui est le mieux discriminé avec une F-mesure de 0,98 pour l'algorithme Random forest ce qui est un résultat vraiment excellent. La détection du contexte /i/ s'améliore même s'il s'agit toujours du contexte le moins bien discriminé (cf. figure 39). Les erreurs qui persistent pour ce contexte concernent les sujets pour lesquels aucune différence n'avait été attesté entre /i/ et /u/ ce qui concerne le sujet F2 ainsi que les consonnes sourdes dans le cas du sujet M2.

Au regard des résultats, la PSG s'avère être un paramètre particulièrement robuste puisqu'on atteint une performance qui se rapproche des 98 % pour la classification du contexte intervocalique avec une précision toujours supérieure à 0,90 ce qui signifie que les systèmes sont très précis. Ces performances sont supérieures à celles obtenues pour l'opposition de voisement ce qui nous indique que le contexte intervocalique est un descripteur particulièrement pertinent.

```

Plot : 11:06:29 - trees.RandomForest
Instance: 536
  datas: psg
  subject: F2
  type: sonores
  point_index: 8.0
  points: 6.79435484086102
  valeurs: 7.607421875
predictedcontexte: u
contexte: i
LieuxArticulation: velaire

Plot : 11:06:29 - trees.RandomForest
Instance: 575
  datas: psg
  subject: M2
  type: sourdes
  point_index: 4.0
  points: 0.811081772583053
  valeurs: 10.4736328125
predictedcontexte: u
contexte: i
LieuxArticulation: velaire

```

Figure 39 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification du contexte intervocalique avec la PSG uniquement.

### 5.3.3 Les lieux d'articulation

Pour la classification automatique des lieux d'articulation, les algorithmes ayant obtenus les meilleurs résultats sont une nouvelle fois PART et Random Forest. Les résultats obtenus sont les suivants :

Tableau 51 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique des lieux d'articulation.

PART				Random Forest			
Classes prédites \ Vraies Classes	bilabiale	alvéolaire	vélaire	bilabiale	alvéolaire	vélaire	
bilabiale	4264	386	324	4496	254	224	
alvéolaire	382	3928	374	320	4062	302	
vélaire	404	397	4420	312	364	4545	

#### Évaluation de la performance des algorithmes PART & Random Forest

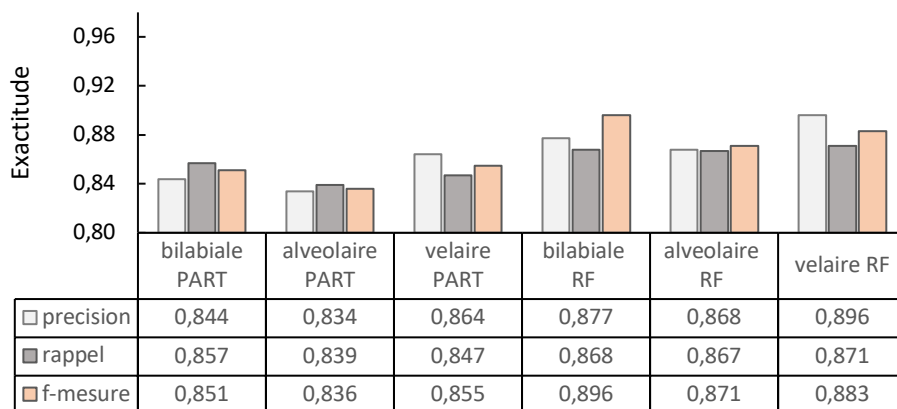


Figure 40 : Évaluation des performances des algorithmes PART et Random forest (RF) pour la classification automatique des lieux d'articulation avec les mesures de Précision, Rappel et F-mesure.

Nous obtenons avec l'algorithme PART un total de 84,7638 % des instances correctement classifiées contre 15,2362 % (12 612/14 879) avec un coefficient kappa de 0,7713. Nous constatons que ce sont les consonnes vélares qui obtiennent les meilleurs résultats de classification ( F-mesure = 0,851, VP = 4 420, P = 0,864) suivies des consonnes bilabiales (F-mesure = 0,844, VP = 4 264, P=0,844) et pour finir celles qui obtiennent les moins bons résultats sont les consonnes alvéolaires (F-mesure = 0,834, VP = 3 928, P = 0,836).

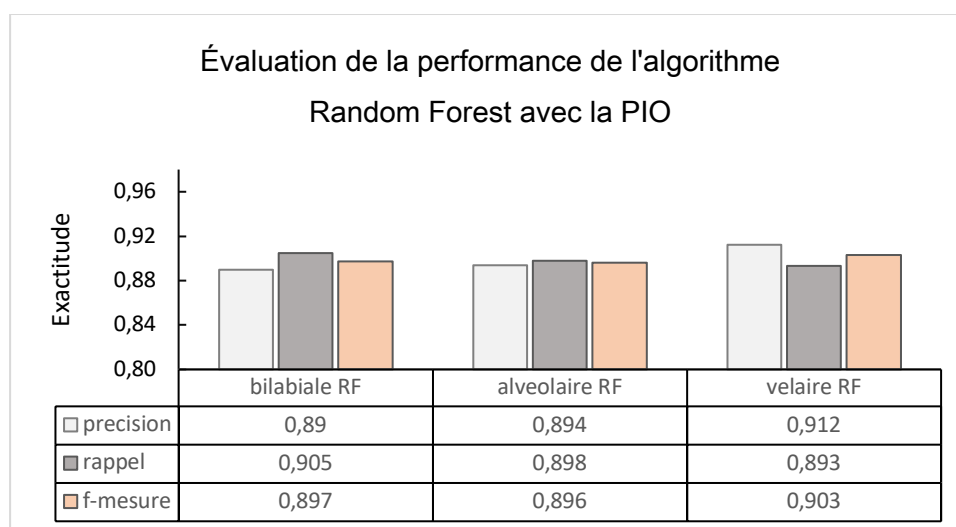
Avec l'algorithme Random Forest, nous obtenons un total de 88,0637 % des instances correctement classifiées contre 11,9363 % ce qui représente à l'échelle de nos données 13 103/14 879 avec un coefficient kappa de 0,8208. Pour cet algorithme, ce sont les consonnes vélares qui obtiennent les meilleurs résultats (F-mesure = 0,896, VP = 3 928) suivies des bilabiales et ensuite des alvéolaires qui une nouvelle fois obtiennent les moins bons résultats de classification (F-mesure = 0,871, VP = 4 062). On peut constater que les consonnes bilabiales obtiennent une F-mesure supérieure à celle des vélares (0,896 > 0,883). Pour autant, nous considérons que sont les consonnes vélares qui obtiennent les meilleurs résultats car le taux de VP est plus élevé (4 545 contre 4 496) avec une meilleure précision (0,896 contre 0,877) ce qui signifie que la détection est plus précise pour les vélares. Nous ne sommes pas étonnés que ce soient les vélares ainsi que les bilabiales qui obtiennent les meilleurs résultats car lors des analyses des mesures de PSG et de PIO les principales différences avaient été établies entre les bilabiales et les vélares.

Ces résultats sont un peu moins bons que ceux des tâches précédentes (voisement, contexte intervocalique) puisque nous obtenons une performance inférieure à 90%, mais ils

restent très satisfaisants. Les erreurs des algorithmes concernent à la fois les mesures de PSG et de PIO mais avec un nombre plus important d'erreurs pour les mesures de PSG. Afin de savoir si un des deux paramètres permet d'améliorer les résultats nous avons relancé l'algorithme Random forest avec la PSG et la durée puis avec la PIO et la durée. Nous avons obtenu une performance de 82,761 % pour la PSG ce qui réduit les performances. En revanche, nous avons atteint une performance de 89,8821 % contre 10,1179 % avec un coefficient kappa de 0.8482 pour la PIO. Les résultats obtenus sont les suivants :

*Tableau 52 : Matrices de confusion de l'algorithme Random forest pour la classification automatique des lieux d'articulation avec la PIO.*

Random Forest			
Classes prédites	bilabiale	alvéolaire	vélaire
Vraies Classes			
bilabiale	2090	107	112
alvéolaire	130	2034	100
vélaire	129	134	2201



*Figure 41 : Évaluation des performances de l'algorithme Random forest (RF) avec la PIO pour la classification automatique des lieux d'articulation avec les mesures de Précision, Rappel et F-mesure.*

Les consonnes vélares obtiennent une nouvelle fois les meilleurs résultats de classification avec 2 201 vrais positifs, une précision de 0,912 et une F-mesure de 0,903. Ensuite, ce sont les consonnes bilabiales avec 2 090 vrais positif ( $P = 0,89$ , F-mesure = 0,897) et pour finir les alvéolaires avec 2 034 vrais positifs ( $P = 0,894$ , F-mesure = 0,896). On remarque que ce sont toujours les consonnes alvéolaires qui obtiennent les moins bons résultats mais l'écart avec les bilabiales s'est nettement creusé car les F-mesure sont dorénavant quasiment équivalentes. On remarque d'ailleurs que les alvéolaires ont une meilleure précision que les bilabiales lorsqu'on retire les mesures de PSG. Ce résultat peut s'expliquer par le fait qu'une différence entre les consonnes vélares et alvéolaires avait été attesté lors de l'analyse de pression intra-orale ainsi que lors de l'analyse de la durée qui était d'ailleurs le paramètre qui permettait de mieux discriminer la différence entre ces deux contextes. Il était également prévisible d'obtenir des meilleurs résultats pour la PIO car c'était le paramètre pour lequel il y avait le plus de différences entre les différents lieux d'articulation pour les quatre sujets d'étude. Ce résultat confirme que la PIO reflète la position et le mouvement des articulateurs dans la cavité buccale et qu'elle est sensible à leurs mouvements et aux altérations des constrictions.

A l'issue de ces différentes tâches de classification, l'algorithme ayant obtenu les meilleurs résultats, parmi tous les algorithmes disponibles dans l'interface WEKA est Random forest pour les trois tâches de classification. La tâche ayant obtenu les meilleures performances est la classification du contexte intervocalique avec 97,8328 % d'instances correctement classifiées en utilisant la PSG comme paramètre principal. Ensuite, il s'agit de la tâche de classification automatique de l'opposition de voisement avec une performance de



96,839 % en utilisant les mesures de PIO et de durée. Pour finir, nous avons obtenu pour la classification automatique de lieux d'articulation un résultat de 89,8821 % d'instances correctement classifiées avec les mesures de PIO et de durée. De plus, nous avons constaté que les différents algorithmes exploitaient systématiquement le descripteur « point index » qui correspond au points d'extraction des mesures en sélectionnant à chaque fois celui qui correspondait à la mesure de PIO ou de PSG la plus pertinente. Lorsque nous avons voulu donner au système uniquement les points du milieu, les résultats étaient nettement moins bons ce qui montre l'avantage d'obtenir plusieurs points d'extraction tout au long de la durée de l'occlusive. Nous avons également été étonnés de constater que les intervalles temporels avaient contribuées à améliorer les performances. Ces résultats sont extrêmement satisfaisants car la classification se base sur un nombre peu élevé d'attributs et d'instances ce qui met en exergue le fort potentiel des mesures aérodynamiques pour la classification automatique de phonèmes. Nous ne pouvons malheureusement pas comparer ces résultats avec ceux existants dans la littérature car nous sommes les seuls à avoir utilisé exclusivement des paramètres aérodynamiques pour réaliser de la classification automatique de phonèmes.

## 6. Conclusion

Ce travail de recherche a permis de mettre en avant l'intérêt d'utiliser les données aérodynamiques pour la classification automatique des consonnes occlusives du français. Nous avons volontairement limité l'analyse aux données aérodynamiques car celles-ci n'avaient pas été exploitées dans des tâches de classification de phonèmes. Or, c'est au stade de la phase aérodynamique que les sons de la parole sont générés. Nous avons montré lors de cette étude la pertinence de ces mesures pour la détection et la description de l'opposition de voisement, l'influence du contexte intervocalique et la distinction des différents lieux d'articulation. En effet, nous avons obtenu lors de classification supervisée une performance de 97,8328 % pour le contexte intervocalique, 96,839 % pour l'opposition de voisement et 89,8821 % pour la classification automatique des lieux d'articulation. Ces tâches de classification ont révélé que la pression sous-glottique permettait à elle seule de détecter le contexte intervocalique. Il en est de même pour la pression intra orale couplée à la durée qui est apparue comme le paramètre le plus pertinent pour détecter l'opposition de voisement et les lieux d'articulation. Il s'agit donc de descripteurs extrêmement robustes à prendre en compte pour l'analyse des occlusives du français. Un second apport de ce travail est la mise en place d'une méthode automatique pour le traitement des données aérodynamiques qui nous a permis de traiter un plus grand nombre de données et d'obtenir des descripteurs montrant l'évolution des mesures de pression au cours de la réalisation de l'occlusive. Nous avons pu ainsi apporter un nombre important de données de référence qui existaient en nombre limité dans la littérature.

## A. Bibliographie

1° - (Auran et al., 2003a) Auran, C., Bouzon, C. (2003a). *Phonotactique prédictive et alignement automatique: application au corpus MARSEC et perspectives*. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence, vol. 22, p. 33-63.

2° - (Baken and Orlikoff, 2000) Baken, R.J., Orlikoff, R.F. (2000). *Clinical Measurement of Speech and Voice*. Singular, San Diego.

3° - (Bazillon et al., 2008a) Bazillon, T., Estève, Y., Luzzati, D. (2008a). *Transcription manuelle vs assistée de la parole préparé et spontanée*. Revue TAL.

4° - (Bigi et Meunier, 2018) Bigi, B., Meunier, C. (2018). *Automatic speech segmentation of spontaneous speech*. In Revista de Estudos da Linguagem. International Thematic Issue: Speech Segmentation. Editors: Tommaso Raso, Heliana Mello, Plinio Barbosa, vol. 26, no 4, e-ISSN 2237-2083.

5° - (Catford, 1977) Catford, J.C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

6° - (Chandra and Senthildevi, 2015) Chandra, E., Senthildevi, K. A. (2015). *Keyword Spotting: An Audio Mining Technique in Speech Processing – A Survey*. IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 4, Ver. II (Jul-Aug. 2015), PP 22-27 e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

7° - (Comite, 2017a) Comité d’Ethique Erasme - Université Libre de Bruxelles. (2017a). *Loi relative aux expérimentations sur la personne humaine (mai 2004) and AAHRPP standard*. Retrieved from <https://www.erasme.ulb.ac.be/fr/ethique> on Feb 22, 2017.

8° - (Demolin et al., 2019) Demolin, D., Hassid, S., Ponchard, C., Yu, S., Trouville, R. (2019). *Speech aerodynamics database*. Laboratoire de phonétique et de phonologie, CNRS-MR 7018, Sorbonne Nouvelle, Paris 3, ILPGA.

9° - (Galindo et Teston, 1989) Galindo, B., Teston, B. (1989). *Physiologia : un logiciel d'analyse des paramètres physiologiques de la parole. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, 13 :197–217.

10° - (Ghio, 2002) Ghio, A. (2002). *Phonedit : Multiparametric speech analysis. LPL (Aix-en-Provence, France)* [www.lpl.univ-aix.fr](http://www.lpl.univ-aix.fr).

11° - (Eibe et al., 2016) Eibe, F., Hall, M., Witten, I.H.(2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.

12° - (Elghazel, 2007a) Elghazel, H. (2007a). *Classification et Préviation des Données Hétérogènes : Application aux Trajectoires et Séjours Hospitaliers*. Mémoire de thèse, Université Claude Bernard Lyon 1, Lyon, France.

13° - (Hosom, 2008b) Hosom, J-P. (2008b). *Speaker-independent phoneme alignment using transition-dependent states*. *Speech Commun.* 51, 352–368.10.1016/j.specom.2008.11.003.

14° - (Landron, 2017b) Landron, S. (2017b). *L'opposition de voisement des occlusives orales du français par des locuteurs taiwanais*. Thèse de Doctorat. Université Sorbonne Nouvelle, Paris.

15° - (Lefébure et Ventari, 2001) Lefébure, R., Ventari, G. (2001). *Data Mining : Gestion de la relation client, Personnalisation des sites Web*. Eds Eyrolles.

16° - (Le Roux, 2014) Le Roux, B. (2014). *Analyse géométrique des données multidimensionnelles*. Chapitre 10 : Introduction à la classification, pages 321 à 344. *Psycho Sup*, Dunod, 2014.

17° - (Nguyen et Adda-Decker, 2013) Nguyen, N., Adda-Decker M. (2013) *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*. France. Hermes Science Publications, Cognition et Traitement de l'Information, 978-2746245303, hal-01135116.

18° - (Pitt et al., 2005a) Pitt, M., Johnson, K., Hume, E., Kiesling, S., Raymond, W. (2005a). *The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability*. *Speech Communication*, vol.45, p. 89-95.

19° - (Preux, 2011) Preux, P. (2011) *Fouille de données* (notes de cours). <http://www.grappa.univ-lille3.fr/ppreux/Documents/notes-de-cours-de-fouille-de-donnees.pdf>.

20° - (Reichel, 2012) Reichel, U.D. (2012). *PermA and Balloon: Tools for string alignment and text processing*. Interspeech. Portland, Oregon, paper no. 346.

21° - (Schiel, 1999) Schiel, F. (1999). *Automatic Phonetic Transcription of Non-Prompted Speech*. Proc. of the ICPHS 1999. San Francisco, pp. 607-610.

22° - (Sébillot, 2014) Sébillot, P. (2014). *Le traitement automatique des langues face aux données textuelles volumineuses et potentiellement dégradées : qu'est-ce que cela change ?*. Lisette Calderan; Pascale Laurent; HélèneLowinger; Jacques Millet. *Big data : nouvelles partitions de l'information*. Actes du séminaire ISTINRIA, De Boeck, pp.43-60, Information et stratégie, 978-2804189150, hal-01056396.

23° - (Seidman, 2001) Seidman, C. (2001). *Data Mining with Microsoft SGL Server 2000 Technical Reference*. Microsoft Press, Redmond,WA.

24° - (Shotton, 2013) Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Moore, R. (2013). *Real-time human pose recognition in parts from single depth images*. *Communications of the ACM*, 56 (1), 116-124.

25° - (Sjölander, 2003b) Sjölander, K. (2003b). *An HMM-based system for automatic segmentation and alignment of speech*. *Phonum*, vol. 9, p. 93-96.

26° - (Teston, 2007b) Teston, B. (2007b). *L'étude instrumentale des gestes dans la production de la parole: Importance de l'aérophonométrie*. P. Auzou. Les Dysarthries, SOLAL, pp.115-117,hal-00173547v2.

27° - (Vaissière, 2006) Vaissière, J, (2006) *La phonétique. Que-Sais-Je ?* No. 637, PUF (in translation in Japanese and Arabic).

28° - (Warren, 1982) Warren, DW. (1982). *Aerodynamics of speech*. In: LassNJ, McReynolds LV, Northern JL, Yoder DE, eds. *Speech, language and hearing*. Philadelphia: WB Saunders.

29° - (Wells, 1940) Wells, H.G. (1940) *Mankind in the making*. Kessinger Publishing Co.

## B. Liste des figures

<i>Figure 1 : Schéma représentant les différentes phases de la production de la parole introduites par (Catford, 1977).....</i>	<i>8</i>
<i>Figure 2 : Dynamique et lieux de constriction des principaux organes articulateur du conduit vocal et situation anatomique des paramètres aérodynamiques (Teston, 2007b).....</i>	<i>10</i>
<i>Figure 3 : Dispositif EVA 2 et position des sujets pour l'enregistrement des données acoustiques (fréquence fondamentale et intensité) et aérodynamiques (pressions sous-glottique et intra-orale, débit oral d'air buccal). Station de travail « Physiologia ».....</i>	<i>21</i>
<i>Figure 4 : Page d'accueil du site de la base données aérodynamiques développé au Laboratoire de Phonétique et Phonologie à l'Université Sorbonne Nouvelle, Paris 3. ....</i>	<i>22</i>
<i>Figure 5 : Spectrogramme, signal acoustique, pression intra-orale (PIO), pression sous-glottique (Psg) et intensité pour le sujet masculin disant 'papa'. Les flèches indiquent le moment où les valeurs ont été prises.....</i>	<i>25</i>
<i>Figure 6 : Résultat de l'alignement automatique effectué avec MAUS.....</i>	<i>27</i>
<i>Figure 7 : Formats des fichiers d'étiquettes avec à gauche le fichier TextGrid et à droite le fichier Mrk. ....</i>	<i>28</i>
<i>Figure 8 : TextGrid fusionné contenant la référence (reference) et le résultat de la segmentation automatique (MAU).....</i>	<i>29</i>
<i>Figure 9 : Résultat du programme python calculant le décalage moyen pour chaque occurrence de la consonne « p » entre les intervalles de la segmentation de référence (debut_ref et fin_ref) par rapport à ceux obtenus à l'issu de la segmentation automatique.....</i>	<i>30</i>
<i>Figure 10 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence pour les données du sujet masculin à gauche et féminin à droite. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 180 (M1), N = 180 (F1)). ....</i>	<i>31</i>

<i>Figure 11 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence par consonnes pour le sujet féminin. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 180, N = 30 (consonnes)).....</i>	<i>32</i>
<i>Figure 12 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence par consonnes pour le sujet masculin. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N=180, N = 30 (consonnes)).....</i>	<i>32</i>
<i>Figure 13 : Décalage moyen en milliseconde (ms) entre le résultat de la segmentation automatique par rapport à la segmentation de référence pour l'opposition de voisement. Les décalages se situant en dessous du seuil de 20 ms sont représentés en bleu clair, ceux étant supérieurs sont en bleu foncé (N = 90)...</i>	<i>33</i>
<i>Figure 14 : Visualisation de la procédure d'extraction manuelle sur le logiciel Phonedit. Le curseur correspond à la ligne verticale rouge en pointillé. Les mesures entourées en rouge correspondent aux intervalles temporels et celles en bleu aux mesures de PIO, PSG et d'INT lorsque nous positionnons le curseur sur l'intervalle de début de l'occlusive /b/.....</i>	<i>37</i>
<i>Figure 15 : Résultat du programme d'extraction automatique.....</i>	<i>38</i>
<i>Figure 16 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction. ....</i>	<i>42</i>
<i>Figure 17 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction. ....</i>	<i>43</i>
<i>Figure 18 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour le sujet F2. ....</i>	<i>45</i>
<i>Figure 19 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets féminins.....</i>	<i>49</i>



<i>Figure 20 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets masculins. ....</i>	<i>51</i>
<i>Figure 21 : Coupes sagittales du conduit vocal d'un locuteur standard dans la réalisation des voyelles orales du français. ....</i>	<i>53</i>
<i>Figure 22 : Coupes sagittales du conduit vocal d'un locuteur standard dans la réalisation des lieux d'articulation du français. ....</i>	<i>55</i>
<i>Figure 23 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les trois lieux d'articulation pour les sujets féminins. ....</i>	<i>57</i>
<i>Figure 24 : Courbes des valeurs moyennes des mesures de PIO (hpa) pour les trois lieux d'articulation pour les sujets masculins. ....</i>	<i>60</i>
<i>Figure 25 : Courbes des valeurs moyennes des mesures de Psg (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction. ....</i>	<i>64</i>
<i>Figure 26 : Courbes des valeurs moyennes des mesures de Psg (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores). En abscisse, les différents points d'extraction. ....</i>	<i>65</i>
<i>Figure 27 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets féminins. ....</i>	<i>69</i>
<i>Figure 28 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les consonnes non voisées en bleue (sourdes) et voisées en rose (sonores) en fonction de l'environnement vocalique pour les sujets masculins. ....</i>	<i>71</i>
<i>Figure 29 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les trois lieux d'articulation pour les sujets féminins. ....</i>	<i>76</i>
<i>Figure 30 : Courbes des valeurs moyennes des mesures de PSG (hpa) pour les trois lieux d'articulation pour les sujets masculins. ....</i>	<i>78</i>
<i>Figure 31 : Diagrammes en boîtes à moustache représentant les durées en ms des consonnes voisées (b,d,g) et non voisées (p,t,k) du français pour le sujet F1 à gauche et F2 à droite. ....</i>	<i>81</i>

<i>Figure 32 : Diagrammes en boîtes à moustache représentant les durées en ms des consonnes voisées (b,d,g) et non voisées (p,t,k) du français pour le sujet M1 à gauche et M2 à droite.....</i>	<i>82</i>
<i>Figure 33 : Évaluation des performances des algorithmes PART et Random forest (RF) avec la totalité des descripteurs pour la classification automatique de l'opposition de voisement avec les mesures de Précision, Rappel et F-mesure. ....</i>	<i>87</i>
<i>Figure 34 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification de l'opposition de voisement. ....</i>	<i>89</i>
<i>Figure 35 : Évaluation des performances des algorithmes PART et Random forest (RF) sans les mesures de PSG pour la classification automatique de l'opposition de voisement avec les mesures de Précision, Rappel et F-mesure. ....</i>	<i>90</i>
<i>Figure 36 : Évaluation des performances des algorithmes BFTree et Random forest (RF) avec tous les paramètres sauf la durée pour la classification automatique du contexte intervocalique avec les mesures de Précision, Rappel et F-mesure.....</i>	<i>92</i>
<i>Figure 37 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification du contexte intervocalique. ....</i>	<i>94</i>
<i>Figure 38 : Évaluation des performances des algorithmes BFTree et Random forest (RF) avec la PSG uniquement pour la classification automatique du contexte intervocalique avec les mesures de Précision, Rappel et F-mesure. ....</i>	<i>95</i>
<i>Figure 39 : Exemples d'erreurs générées par l'algorithme Random forest pour la classification du contexte intervocalique avec la PSG uniquement. ....</i>	<i>96</i>
<i>Figure 40 : Évaluation des performances des algorithmes PART et Random forest (RF) pour la classification automatique des lieux d'articulation avec les mesures de Précision, Rappel et F-mesure. ....</i>	<i>97</i>
<i>Figure 41 : Évaluation des performances de l'algorithme Random forest (RF) avec la PIO pour la classification automatique des lieux d'articulation avec les mesures de Précision, Rappel et F-mesure. ....</i>	<i>99</i>

## C. Liste des tableaux

<i>Tableau 1 : Consonnes occlusives du français.</i> .....	23
<i>Tableau 2 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet F1. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900 (NV), N=900(V)).</i> .....	41
<i>Tableau 3 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet F2. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i> .....	42
<i>Tableau 4 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet M1. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i> .....	43
<i>Tableau 5 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) et non voisées (NV) pour le sujet M1. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i> .....	43
<i>Tableau 6 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	47
<i>Tableau 7 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	47
<i>Tableau 8 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet</i>	

<i>F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	48
<i>Tableau 9 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	48
<i>Tableau 10 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	49
<i>Tableau 11 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	50
<i>Tableau 12 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	50
<i>Tableau 13 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	51
<i>Tableau 14 : Résumé des différences significatives des mesures de PIO entre les différents contextes intervocaliques</i> .....	52
<i>Tableau 15 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i> .....	55
<i>Tableau 16 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F1. La p-value</i>	

<i>est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	56
<i>Tableau 17 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	56
<i>Tableau 18 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	57
<i>Tableau 19 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	58
<i>Tableau 20 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	58
<i>Tableau 21 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	59
<i>Tableau 22 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	59
<i>Tableau 23 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet F1. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i>	63

<i>Tableau 24 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet F2. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i>	63
<i>Tableau 25 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet M1. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i>	64
<i>Tableau 26 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) et non voisées (NV) pour le sujet M2. La p-value (P) est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i>	65
<i>Tableau 27 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	67
<i>Tableau 28 : Résultat du test de Welch avec les mesures moyennes de PIO des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	67
<i>Tableau 29 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	68
<i>Tableau 30 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	68
<i>Tableau 31 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet</i>	

<i>M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>69</i>
<i>Tableau 32 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) dans les trois contextes intervocaliques pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>70</i>
<i>Tableau 33 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>70</i>
<i>Tableau 34 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (NV) dans les trois contextes intervocaliques pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>71</i>
<i>Tableau 35 : Résumé des différences significatives des mesures de PSG entre les différents contextes intervocaliques. ....</i>	<i>72</i>
<i>Tableau 36 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>74</i>
<i>Tableau 37 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>74</i>
<i>Tableau 38 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).....</i>	<i>75</i>
<i>Tableau 39 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet F2. La p-value</i>	

<i>est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	75
<i>Tableau 40 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	76
<i>Tableau 41 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M1. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	77
<i>Tableau 42 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes non voisées (NV) pour les trois lieux d'articulation pour le sujet F2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	77
<i>Tableau 43 : Résultat du test de Welch avec les mesures moyennes de PSG des consonnes voisées (V) pour les trois lieux d'articulation pour le sujet M2. La p-value est indiquée avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=30).</i>	78
<i>Tableau 44 : Résultat du test de Welch comparant les durées en ms des consonnes voisées (V) et non voisées (NV). La p-value (P) est indiqué avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001) (N=900).</i>	80
<i>Tableau 45 : Résultat du test de Welch comparant les durées en ms des consonnes voisées (V) en fonction des lieux d'articulation. La p-value (P) est indiqué avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001).</i>	80
<i>Tableau 46 : Résultat du test de Welch comparant les durées en ms des consonnes non voisées (NV) en fonction des lieux d'articulation. La p-value (P) est indiqué avec son niveau de significativité (* p &lt;0,05; ** p &lt;0,01; *** p &lt;0,001).</i>	81
<i>Tableau 47 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique de l'opposition de voisement.</i>	86



*Tableau 48 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique de l'opposition de voisement sans la PSG. ....90*

*Tableau 49 : Matrices de confusion des algorithmes BFTree et Random forest pour la classification automatique du contexte intervocalique. ....92*

*Tableau 50 : Matrices de confusion des algorithmes BFTree et Random forest pour la classification automatique du contexte intervocalique avec la PSG uniquement...95*

*Tableau 51 : Matrices de confusion des algorithmes PART et Random forest pour la classification automatique des lieux d'articulation. ....97*

*Tableau 52 : Matrices de confusion de l'algorithme Random forest pour la classification automatique des lieux d'articulation avec la PIO.....99*

## D. Annexes

### D.1 Conversion fichiers mrk en textgrid

```
import soundfile as sf
from collections import defaultdict
import re

#open .wav file
path_wav = '/Users/claraponchard/Desktop/Segmenation/Phil/Automatique/naninu.wav'
wav = sf.SoundFile(path_wav)

#duration of the .wav file
duration_wav = format(len(wav) / wav.samplerate)

# open and read a .mrk file
path = '/Users/claraponchard/Desktop/Segmenation/Phil/Manuelle /naninu.mrk'
try :
    mrk = open(path, encoding = "utf-8").read()
except UnicodeDecodeError:
    mrk = open(path,encoding="cp1252").read()

# Traitement du fichier mrk
list_mrk = mrk.split("\n") #toutes les lignes

dico_refTime = {} #pour récupérer la liste des références temporels {id : valeur}
dico_descLevel = {} #pour récupérer les noms des différents niveaux d'annotation {niveau : son nom}
#pour récupérer les valeurs des différents intervalles --> {niveau : [texte,ref_debut,ref_fin]}
dico_levelValeurs = defaultdict(lambda :list())

intervalle = re.compile(r'LBL_LEVEL_(\^[^_]+)_d+="(.)" (TS_\d+) (TS_\d+)') #pour matcher les lignes d'intervalles

for i,element in enumerate(list_mrk) : #parcours des lignes du fichier
    match_intervalle = re.search(intervalle,element) #pour repérer les lignes avec les infos des intervalles
    if element.startswith('TS_') : #lignes avec référence temporelle
        ref,time = element.split("=")
        dico_refTime[ref]=str(float(time)/1000)
    elif element.startswith('DSC_LEVEL') : #lignes avec le niveau
        level = element.split("_")[2].replace(']', '')
        name = list_mrk[i+1].split("=")[1] #le nom est sur la ligne suivante
        dico_descLevel[level]=name
    elif match_intervalle : #lignes avec les infos pour les intervalles
        level = match_intervalle.group(1)
        text = match_intervalle.group(2)
        debut = match_intervalle.group(3)
        fin = match_intervalle.group(4)
        dico_levelValeurs[level].append([text,dico_refTime[debut],dico_refTime[fin]])
print(dico_levelValeurs)
```

```

# Chaque niveau d'annotation représente tout l'axe temporel
# il faut donc insérer des segments sans annotation entre deux segments avec étiquette
# si ces segments ne partagent pas une frontière

# Rappel de structures :
# dico_levelValeurs --> {niveau d'annotation : [ ['etiquette 1','debut','fin'], ['eti 2','debut','fin'] ] }
# dico_final --> même structure que dico_levelValeurs MAIS contient les intervals sans étiquette !

dico_final = defaultdict(lambda :list())
for level in dico_levelValeurs : #pour chaque niveau d'annotation
    dico_final[level].append(['',0,dico_levelValeurs[level][0][1]]) #1er interval, qui commence à 0
    for i,valeurs in enumerate(dico_levelValeurs[level]): #pour chaque intervalle avec étiquette
        #si ce n'est pas le 1er intervalle avec étiquette (pour éviter l'erreur)
        #et que son début ne coïncide pas avec la fin de l'intervalle précédent
        if (i!=0 and
            dico_levelValeurs[level][i][1] != dico_levelValeurs[level][i-1][2] and
            dico_levelValeurs[level][i-1][2] < dico_levelValeurs[level][i][1]
        ) :

            #ajout d'un intervalle sans étiquette
            dico_final[level].append([
                '',
                dico_levelValeurs[level][i-1][2], # fin du précédent
                dico_levelValeurs[level][i][1] # début du courant
            ])
            #ajout de l'intervalle avec étiquette
            dico_final[level].append([
                dico_levelValeurs[level][i][0], #texte
                dico_levelValeurs[level][i][1], #début
                dico_levelValeurs[level][i][2] #fin
            ])
    else : #dans tous les autres cas
        dico_final[level].append([
            dico_levelValeurs[level][i][0],
            dico_levelValeurs[level][i][1],
            dico_levelValeurs[level][i][2]
        ])

# create (or overwrite) new file foroutput Textgrid
path_textGrid = "naninu.TextGrid"
textgrid = open(path_textGrid,"w",encoding='UTF-8')

# write textgrid preamble
textgrid.write("File type = \"ooTextFile\" + '\n') # text type : Organized by tier.
textgrid.write("Object class = \"TextGrid\" + '\n' + '\n')
textgrid.write("xmin = 0" + '\n') #First marked time of the file
textgrid.write("xmax = " + duration_wav + '\n') #Last marked time of the file.
textgrid.write("tiers? <exists>" + '\n')
textgrid.write("size = " + str(len(dico_descLevel)) + '\n')
textgrid.write("item []: " + '\n')

for level in dico_descLevel :

    textgrid.write("    item ["+str(list(dico_descLevel.keys()).index(level)+1)+"]:" + '\n')
    textgrid.write("        class = \"IntervalTier\" + '\n') #Transcription is marked as intervals
    textgrid.write("        name = " + "'" + dico_descLevel[level] + "'" + '\n')
    textgrid.write("        xmin = 0" + '\n')
    textgrid.write("        xmax = " + duration_wav + '\n')
    textgrid.write("        intervals: size = " + str(len(dico_final[level])) + '\n')

    for i in range(len(dico_final[level])):

        textgrid.write("            intervals [" + str(i+1) + "]:" + "\n")
        textgrid.write("                xmin = " + str(dico_final[level][i][1]) + "\n")
        textgrid.write("                xmax = " + str(dico_final[level][i][2]) + "\n")
        textgrid.write("                text = \"" + str(dico_final[level][i][0]) + "\"\n")

f = open("naninu.TextGrid").read()
print(f)

```

## D.2 Script d'extraction automatique des mesures aérodynamiques

```
import os #module pour parcourir une arborescence de fichiers
import tgt #module pour manipuler des textgrids
import re #module pour manipuler des expressions régulières

#fonctions pour parcourir les dossiers contenant les fichiers mrk et textgrid et récupérer les informations utiles
def get_fichiers_tg(base_chemin):
    all_data_tg = []
    for prenom in os.listdir(base_chemin):
        chemin = os.path.join(base_chemin, prenom)
        if prenom != '.DS_Store':
            for types in os.listdir(chemin):
                if types != '.DS_Store':
                    chemin_type = os.path.join(chemin, types)
                    for consonnes in os.listdir(chemin_type):
                        if consonnes != '.DS_Store':
                            chemin_consonnes = os.path.join(chemin_type, consonnes)
                            for datas in os.listdir(chemin_consonnes):
                                if datas != '.DS_Store':
                                    chemin_datas = os.path.join(chemin_consonnes, datas)
                                    if datas[-9:] in ['.TextGrid']:
                                        all_data_tg.append([chemin_datas, prenom, types, consonnes, datas])
    return all_data_tg

def get_fichiers_mrk(base_chemin):
    all_data = []
    for prenom in os.listdir(base_chemin):
        chemin = os.path.join(base_chemin, prenom)
        if prenom != '.DS_Store':
            for types in os.listdir(chemin):
                if types != '.DS_Store':
                    chemin_type = os.path.join(chemin, types)
                    for consonnes in os.listdir(chemin_type):
                        if consonnes != '.DS_Store':
                            chemin_consonnes = os.path.join(chemin_type, consonnes)
                            for datas in os.listdir(chemin_consonnes):
                                if datas != '.DS_Store':
                                    chemin_datas = os.path.join(chemin_consonnes, datas)
                                    if datas[-4:] in ['.mrk', '.int', '.pio', '.psg', '.MRK', '.INT', '.PIO', '.PSG']:
                                        all_data.append(chemin_datas)
    return all_data

: #fonctions pour récupérer les récupérer les intervalles et les étiquettes des consonnes de la tier MAU
def get_liste_intervalles(chemin_fichier_tg, liste_des_consonnes):

    tg_auto = tgt.read_textgrid(chemin_fichier_tg)
    tier_auto = tg_auto.get_tier_by_name('MAU')

    all_intervals = []
    for consonne in liste_des_consonnes:
        intervals_dela_consonne = tier_auto.get_annotations_with_text(consonne, n=10000)
        all_intervals += intervals_dela_consonne

    return all_intervals

#fonction qui selectionne uniquement les consonnes en position C2
def deuxieme_consonne(all_intervals):

    liste_intervals_deuxieme = []
    liste_intervals_deuxieme.extend(all_intervals[1::2])

    return liste_intervals_deuxieme

#création d'un dictionnaire avec l'index de la consonne, son interval de début, fin et la durée
def intervalles_to_dict(liste_intervalles):
    liste_intervalle_new = []
    for i, interval in enumerate(liste_intervalles):
        dico_interval_i = {'index': i, 'debut': interval.start_time, 'fin': interval.end_time, 'duree': interval.durat}
        liste_intervalle_new.append(dico_interval_i)

    return liste_intervalle_new

: #construction d'un tableau pour structurer les données avec Numpy
import numpy as np

def intervalle_dict_to_linspace(interval_dict):
    return list(np.linspace(start = interval_dict['debut'], stop = interval_dict['fin'], num = 10))
```

```

: import wsig
import numpy as np

#on récupère les valeurs de PSG, PIO correspondant aux intervalles temporels
def prendre_valeur(filename, cursorTime):

    wave = wsig.read(filename)

    # Calibration
    if wave._filetype == b'WSIG':

        signal = np.frombuffer(wave.readframes(-1), np.int16)

        calibratedSignal = (signal - wave.getzero()) * (wave.getvalueatmax() / wave.getsignaldynamic())

        framerate = wave.getframerate() # frame rate
        cursorFrame = int(cursorTime * framerate) # actual cursor frame

        valeur = calibratedSignal[cursorFrame]

        return valeur

: import pandas as pd

#creation du fichier csv en sortie
def create_fichier_analyse(base_chemin, liste_consonnes):

    all_data_tg = get_fichiers_tg(base_chemin)

    liste_total = []

    for tg_liste in all_data_tg:
        fichier, prenom, types, consonnes, datas = tg_liste

        liste_intervalles = get_liste_intervalles(fichier, liste_consonnes)
        deuxieme = deuxieme_consonne(liste_intervalles)
        liste_intervalles_dicts = intervalles_to_dict(deuxieme)

        les_fichiers_data = [fichier[:-9] + x for x in ['.psg']]

        for consonne_dict in liste_intervalles_dicts:
            les10points = intervalle_dict_to_linspace(consonne_dict)
            for data_file in les_fichiers_data:
                for point_index, cursortime in enumerate(les10points):

                    liste_total.append([datas, data_file[-3:], prenom, types, consonnes, point_index, cursortime, pr

    analyse = pd.DataFrame.from_records(liste_total)
    analyse = analyse.reset_index(drop = False) # index as column
    analyse.columns = ['index', 'tgfile', 'datas', 'subject', 'type', 'consonnes', 'point_index', 'points', 'valeurs']
    return analyse

```

```

: pd.set_option('display.max_colwidth', -1)

```

```

: analyse = create_fichier_analyse(base_chemin = '/Users/claraponchard/Desktop/clara_stage/Aerodynamics-Database',
    liste_consonnes = ["p","t","k","b","d","g"])

```

```

: analyse

```

	index	tgfile	datas	subject	type	consonnes	point_index	points	valeurs
0	0	babibu.TextGrid	psg	M1	sonores	b	0	3.031880	16.171875
1	1	babibu.TextGrid	psg	M1	sonores	b	1	3.039646	16.152344
2	2	babibu.TextGrid	psg	M1	sonores	b	2	3.047412	16.152344
3	3	babibu.TextGrid	psg	M1	sonores	b	3	3.055178	16.152344
4	4	babibu.TextGrid	psg	M1	sonores	b	4	3.062944	16.132812
5	5	babibu.TextGrid	psg	M1	sonores	b	5	3.070710	16.152344
6	6	babibu.TextGrid	psg	M1	sonores	b	6	3.078475	16.171875
7	7	babibu.TextGrid	psg	M1	sonores	b	7	3.086241	16.210938
8	8	babibu.TextGrid	psg	M1	sonores	b	8	3.094007	16.230469
9	9	babibu.TextGrid	psg	M1	sonores	b	9	3.101773	16.230469
10	10	babibu.TextGrid	psg	M1	sonores	b	0	5.749226	14.238281
11	11	babibu.TextGrid	psg	M1	sonores	b	1	5.757136	14.296875
12	12	babibu.TextGrid	psg	M1	sonores	b	2	5.765046	14.355469
13	13	babibu.TextGrid	psg	M1	sonores	b	3	5.772955	14.414062
14	14	babibu.TextGrid	psg	M1	sonores	b	4	5.780865	14.472656
15	15	babibu.TextGrid	psg	M1	sonores	b	5	5.788775	14.531250
16	16	babibu.TextGrid	psg	M1	sonores	b	6	5.796684	14.589844
17	17	babibu.TextGrid	psg	M1	sonores	b	7	5.804594	14.648438
18	18	babibu.TextGrid	psg	M1	sonores	b	8	5.812504	14.687500