

Séparation des sources acoustiques de la parole : Application à la phonétique expérimentale

Acoustic source separation in speech: application to experimental phonetics

Benjamin Elie

Laboratoire des Signaux et Systèmes
CentraleSupélec/CNRS/Université Paris-Saclay
Bâtiment Bréguet
3, rue Joliot Curie
91190 Gif sur Yvette
E-mail : benjamin.elie@centralesupelec.fr

Gilles Chardon

Laboratoire des Signaux et Systèmes
CentraleSupélec/CNRS/Université Paris-Saclay
Bâtiment Bréguet
3, rue Joliot Curie
91190 Gif sur Yvette
E-mail : gilles.chardon@centralesupelec.fr

Yves Laprie

Loria
Inria/CNRS/Université de Lorraine
Campus Scientifique
BP 239
54506 Vandoeuvre lès Nancy
E-mail : Yves.laprie@loria.fr

Résumé

Dans le but d'étudier, d'analyser, et de classer les différents sons issus de la parole humaine, la phonétique utilise une variété d'indices acoustiques extraits des signaux de la parole. Depuis la création de l'alphabet phonétique international en 1888, le nombre de catégories et de réalisations de sons répertoriés augmente continuellement grâce au développement des dispositifs expérimentaux et des méthodes d'analyse de signaux. Dans cet article, nous nous intéressons à la production des fricatives, qui sont des consonnes caractérisées par l'apparition d'un bruit de friction en aval d'un resserrement du conduit vocal. Plus particulièrement, nous développons une méthode d'analyse de signaux qui permet d'étudier finement la production des fricatives voisées. Celles-ci se caractérisent par la présence de deux sources acoustiques principales : le bruit de friction et la source voisée, due aux oscillations des cordes vocales. La méthode consiste à les séparer à partir du signal acoustique de la parole. Par rapport aux méthodes existantes, la méthode X-GLOS que nous présentons est plus robuste au bruit et au niveau de jitter. Nous présentons une application dédiée à la phonétique expérimentale permettant de valider des hypothèses issues de simulations numériques. L'expérience démontre ainsi qu'il existe une zone d'instabilité dans la production des fricatives voisées dans laquelle de faibles perturbations des configurations articulatoires et laryngées entraînent de grandes variations des indices acoustiques. Nous montrons également que les locuteurs tendent à éviter de rester dans cette zone d'instabilité lorsqu'ils produisent des fricatives voisées.

Abstract

In order to study, analyze, and classify the different sounds of human speech, phonetic studies use various acoustic features extracted from speech signals. Since the creation of the International Phonetic Alphabet in 1888, the number of classes and realizations of indexed sounds is constantly increasing thanks to the development of experimental devices and signal processing techniques. This paper focuses on the production of fricatives, which are consonant sounds that are characterized by the appearance of a frication noise downstream of a constriction in the vocal tract. More specifically, it presents a signal processing method that enables the production of voiced fricatives to be finely investigated. Voiced fricatives are characterized by the simultaneous presence of two main acoustic sources: the frication noise and the voiced source, generated by the oscillations of the vocal folds. The method consists in separating them from the audio speech signal. In comparison with the existing techniques, the X-GLOS method that we are presenting is more robust to the noise and jitter levels. The paper presents an application to experimental phonetics that validates previous hypotheses made from numerical simulations. The experiment proves that voiced fricatives are produced in an unstable zone in which small perturbations of the articulatory and laryngeal configurations lead to large variations of the acoustic features. It also proves that speakers tend to avoid spending too much time in the unstable zone when they produce voiced fricatives.

Rappels des connaissances sur la production de la parole et enjeux des travaux

Appareil vocal : un système multiphysique complexe

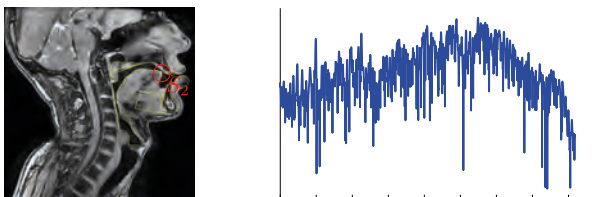
La production de la parole humaine fait intervenir un ensemble de phénomènes de natures différentes. Les phénomènes biomécaniques se traduisent par les gestes que produit le locuteur en contrôlant les mouvements des différents articulateurs (lèvres, mandibule, langue...). Ces gestes vont alors moduler temporellement la géométrie du conduit vocal qui agit comme un guide d'onde : cela fait alors intervenir des phénomènes de nature acoustique. Enfin, des phénomènes mécaniques peuvent intervenir. Ils sont généralement dus aux interactions fluide/structure au

voisinage des différents articulateurs. Ils peuvent également provoquer des phénomènes d'auto-oscillations : l'exemple le plus caractéristique étant les oscillations des cordes vocales (qu'on désigne généralement par *plis vocaux* chez les spécialistes de la parole), qui sont responsables du voisement. Tous ces phénomènes dynamiques permettent de moduler les indices acoustiques pour transmettre l'information à deux niveaux distincts : le niveau *segmental*, à savoir la production des sons d'une langue (la réalisation acoustique d'un *phonème*) qui permettent de distinguer les mots prononcés, et le niveau *suprasegmental* (ou *prosodique*), qui comprend l'ensemble des variations expressives pour prononcer les mêmes mots (accentuation, intonation, rythme...).

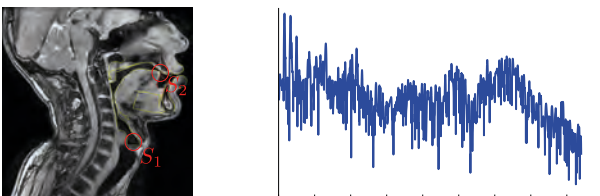
Phonétique articulatoire : classification des sons

Nous rappelons ici la base terminologique communément admise en phonétique articulatoire qui servira à la bonne compréhension de l'article. Elle consiste à classer les sons de la parole selon leur mode et leur lieu d'articulation. Nous distinguons ainsi les consonnes des voyelles par le degré plus ou moins fort du ou des rétrécissements qui apparaissent dans le conduit vocal, fort sous la forme d'une constriction, voire total au point de former une occlusion dans le conduit vocal, pour les consonnes, et faible pour les voyelles. Les consonnes sont ensuite classées selon leur mode de production, et ensuite selon leur lieu ou point d'articulation, c'est-à-dire l'endroit où apparaît la constriction principale. Par exemple, les fricatives sont caractérisées par une constriction suffisamment étroite à l'intérieur du conduit vocal pour provoquer l'apparition de turbulences du jet d'air, sans qu'il y ait de fermeture complète du conduit vocal. Acoustiquement, cela se traduit par la présence d'un bruit de friction dû aux turbulences. Cet article s'intéresse aux fricatives du français, qui correspondent à trois différents points d'articulation¹ : labio-dentale (constriction formée entre la lèvre inférieure et les incisives supérieures), alvéolaire (constriction entre la pointe de la langue et l'arrière des dents supérieures), et post-alvéolaires (constriction entre la pointe de la langue et le palais dur). Pour chacun de ces trois points d'articulation, le français distingue les fricatives sourdes (sans oscillations des plis vocaux), des fricatives voisées (avec oscillations des plis vocaux). Ainsi, nous considérons deux fricatives (voisée/sourde) pour chacun des trois lieux d'articulations : /v,f/ (vert, fer) pour les labio-dentales, /z,s/ (zoo, sot) pour les alvéolaires, et /ʒ,ʒ/ (jarre, char) pour les post-alvéolaires.

Sources acoustiques de la parole



a) Fricative sourde (non-voisée)



b) Fricative voisée

Fig. 1 : Positionnement des sources acoustiques à l'intérieur du conduit vocal pour une fricative alvéolaire sourde /s/ (haut) et voisée /z/ (bas). Un exemple de spectre typique de fricative alvéolaire est représenté à droite, en haut pour /s/ et en bas pour /z/.
Location of the acoustic sources inside the vocal tract for a voiceless alveolar fricative /s/ (top) and a voiced alveolar fricative /z/ (bottom). Typical spectra of alveolar fricatives are shown on the right, top for /s/ and bottom for /z/.

Dans cet article, nous distinguons deux sources acoustiques principales dans la parole : la source glottique (ou voisée, due aux oscillations des plis vocaux), et la source supraglottique (ou bruit de friction, due à l'apparition de turbulences du jet d'air en aval de la constriction supraglottique). La source glottique est de nature périodique, de spectre harmonique, alors que la source supraglottique est de type stochastique, apériodique. Dans le cas des fricatives voisées, les deux sources sont activées de manière simultanée (cf. Figure 1). Dans ce cas, le spectre du son généré contient des composantes harmoniques, ainsi que des composantes bruitées, d'aspect large bande.

Recherches en phonétique et en production de la parole

Les recherches concernant la production de la parole s'intéressent à l'ensemble des phénomènes décrits au premier paragraphe.

D'un point de vue modélisation, les modèles biomécaniques [1] et articulatoires [2, 3] permettent d'étudier comment s'articulent les différents organes articulatoires pour générer les suites de phonèmes, ainsi que les phénomènes transitoires et anticipatoires (coarticulation). Les simulations acoustiques [4, 5, 6] permettent d'étudier l'impact acoustique des configurations articulatoires et laryngées sur la production de certains sons de la parole. Les modèles mécaniques auto-oscillants des plis vocaux [7, 8] ou de la pointe de la langue [9, 10] permettent, quant à eux, d'étudier les conditions d'apparition de phénomènes d'auto-oscillations qui sont essentiels pour produire certains sons.

La phonétique expérimentale s'intéresse à l'analyse des traits distinctifs permettant de classer les sons d'une langue (niveau segmental), ou les traits prosodiques (niveau supra-segmental). Pour effectuer ces classifications, elle utilise des indices acoustiques dont la complexité et la finesse ont augmenté avec le développement de nouvelles techniques de mesures et d'analyse des signaux. Un exemple caractéristique est la complexité croissante de l'*Alphabet Phonétique International* [11] depuis sa création en 1888, et qui a vu le nombre de catégories et de réalisations de sons répertoriés augmenter continuellement. Il est donc important de proposer des méthodes d'analyse des signaux acoustiques permettant d'extraire le maximum d'informations utiles à l'identification et la compréhension des phénomènes impliqués lors de la production de la parole. C'est dans ce but que la méthode d'identification et de séparation des sources acoustiques de la parole, qui est détaillé dans cet article, a été conçue.

But et enjeux de l'étude présentée

Cette étude s'intéresse au contraste de voisement dans les fricatives. La grande majorité des études précédentes qui s'intéressent aux différences perceptives entre les fricatives voisées et sourdes se focalisent sur les temps relatifs que durent les segments voisés et non-voisés lors de la production des fricatives [12, 13, 14, 15]. Le voisement est alors vu comme binaire. Or, le rapport énergétique entre les contributions de la source glottique et celles du bruit de friction varie de manière continue.

1- A noter que le «r» français est communément considéré comme une fricative (fricative uvulaire voisée, notée /ʀ/), mais nous l'excluons de l'étude car il n'est pas prononcé comme une fricative par l'ensemble de la communauté linguistique francophone : il existe différentes façons de prononcer «r» librement en français qui ne sont pas toutes fricatives.

En effet, nous pouvons supposer que les mouvements des articulateurs supraglottiques et de l'abduction graduelle de la glotte, qui sont relativement lents devant les oscillations des plis vocaux, modifient de manière graduelle à la fois l'amplitude du bruit de friction et l'amplitude des contributions de la source glottique. L'impact acoustique des configurations du conduit vocal au niveau supraglottique sur le bruit de friction a été étudié [16, 17], mais très peu d'attention a été portée sur l'impact acoustique des configurations à la glotte. Dans une précédente étude [18], nous avons démontré de manière numérique que les configurations à la glotte avaient un impact important sur la production des fricatives en contrôlant leurs régimes de production. Ces points sont détaillés dans la section « *Mise en évidence des régimes de production des fricatives* ». La découverte de l'existence de ces régimes offre une nouvelle vision sur les fricatives, notamment par l'intermédiaire d'une approche continue du contraste de voisement, en opposition à l'approche binaire qui prévaut dans les études précédentes.

Néanmoins, dans l'optique d'études expérimentales permettant de vérifier les hypothèses issues des simulations numériques (Cf le paragraphe « *Hypothèses sur les stratégies de production des fricatives* », page suivante), il nous faut une méthode efficace et robuste permettant de séparer les contributions des deux sources principales de la parole lors de la production des fricatives, à savoir la source glottique (voisement) et la source supraglottique (bruit de friction), même en cas de conditions difficiles propres aux fricatives voisées (coloration du bruit, fort niveau de bruit, fort niveau de jitter...). Le principe de la méthode X-GLOS [19] que nous avons développée pour répondre à ces besoins est détaillé à la suite de l'article. Cette méthode s'appuie sur l'estimation des paramètres instantanés des sinusoides qui correspondent aux contributions de la source glottique. Contrairement aux méthodes concurrentes (par exemple PAP [20] ou PSHF [21]), qui utilisent un modèle localement stationnaire, l'approche non-stationnaire de la source glottique permet une meilleure robustesse aux variations locales de la fréquence fondamentale observées en pratique dans la parole naturelle. La méthode est appliquée pour une étude expérimentale préliminaire permettant d'étudier la validité des hypothèses issues des études numériques [18]. Les résultats seront détaillés et discutés dans le chapitre suivant au paragraphe « *Résultats* », page suivante.

Étude de production des fricatives

Cette section présente une application directe de la méthode de séparation des sources acoustiques de la parole pour répondre à des problèmes d'ordre phonétique. Ici, nous nous intéressons à la production des fricatives, et plus précisément à la quantification du rapport voisement/bruit de friction contenus dans le signal acoustique des fricatives voisées. L'intérêt étant alors de pouvoir étudier cette quantité à travers son évolution temporelle, notamment lors des transitions voyelles/consonnes, ainsi que pendant le segment fricatif, et de pouvoir ainsi dégager des manières et stratégies différentes de produire ces sons. Pour cela, nous avons mesuré le quotient de voisement (VQ), estimé à l'aide de la méthode X-GLOS.

Mise en évidence des régimes de production des fricatives

Nous rappelons ici brièvement les résultats de nos récents travaux qui ont mis à jour l'existence de régimes de production des fricatives [18, 30].

Dans [18], nous avons effectué des simulations numériques à partir de configurations statiques du conduit vocal, dérivées d'IRM d'un sujet, correspondant à la production de fricatives au trois points d'articulation du français, à savoir labio-dental (/f,v/), alvéolaire (/s,z/), et post-alvéolaire (/ʃ,ʒ/). Plusieurs simulations ont ensuite été effectuées avec des degrés d'ouverture glottique différents pour étudier l'impact acoustique de l'abduction partielle de la glotte sur la production de fricatives.

L'impact acoustique a notamment été analysé à l'aide d'un indicateur continue de la quantité de voisement. Celui-ci est mesuré à l'aide du quotient de voisement, noté VQ, qui quantifie le pourcentage des contributions de la source voisée (glottique) dans le signal acoustique produit, à savoir :

$$VQ = 100 \times \frac{\|s_p(t)\|^2}{\|s(t)\|^2}.$$

Une valeur de 100% indique un signal purement voisé, une valeur de 0% indique un signal purement non-voisé (apériodique), et une valeur de 50% indique que les contributions de la source glottique et de la source supraglottique possèdent la même énergie.

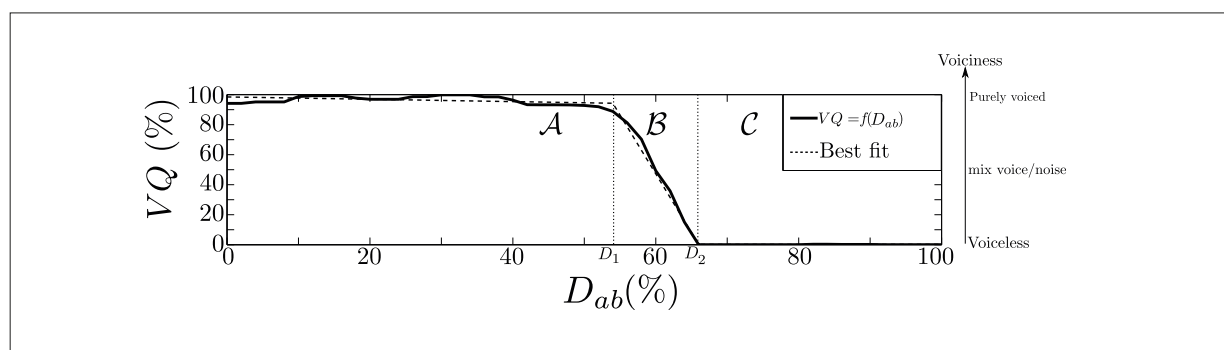


Fig. 2 : Exemple de courbe typique présentant le quotient de voisement VQ en fonction du degré d'abduction partielle de la glotte D_{ab} . Les ouvertures critiques D_1 et D_2 représentent les frontières des différents régimes, A, B et C. Typical curve of the voicing quotient as a function of the partial abduction degree of the glottis D_{ab} . Critical openings D_1 and D_2 are the boundaries of the production regimes, A, B and C.

Le résultat principal de l'étude est la mise en évidence de régimes de production des fricatives qui sont pilotés par le degré d'abduction partielle de la glotte. Ces régimes, notés \mathcal{A} , \mathcal{B} et \mathcal{C} sont représentés sur la figure 2. Le régime \mathcal{A} est caractérisé par une valeur de VQ constamment élevée, et faiblement dépendante de l'ouverture glottique. C'est un régime stable pour lequel le signal est presque exclusivement voisée. À partir d'une certaine ouverture glottique critique, notée D_1 , nous passons au régime \mathcal{B} , caractérisé principalement par son instabilité : de faibles perturbations de l'ouverture glottique donnent lieu à de fortes variations des indices acoustiques. Ce régime correspond à celui qui doit être produit pour les fricatives voisées, à savoir le régime pour lequel le signal acoustique contient à la fois des contributions de la source voisée et du bruit de friction à des niveaux d'énergie comparables. À partir d'une deuxième ouverture glottique critique, notée D_2 , nous passons au régime \mathcal{C} , caractérisé par l'absence de contributions de la source voisée, d'où un niveau de VQ quasi-nul et constant. Les indices acoustiques sont alors stables. C'est le régime des fricatives sourdes.

Hypothèses sur les stratégies de production des fricatives

Lors de la production de fricatives en contexte intervocalique, le locuteur peut donc effectuer plusieurs parcours différents parmi ces régimes de production. Pour les fricatives sourdes, il n'existe qu'un seul chemin possible : le locuteur finit la voyelle en \mathcal{A} , puis doit transiter rapidement en \mathcal{B} pour aller en \mathcal{C} , correspondant au régime des fricatives sourdes. Le régime \mathcal{C} étant stable, le locuteur ne rencontre théoriquement aucune difficulté pour maintenir ce régime. Pour les fricatives voisées, le locuteur devrait théoriquement aller \mathcal{B} de \mathcal{A} en, maintenir ce dernier, et revenir en \mathcal{A} pour la voyelle suivante. Or, \mathcal{B} étant instable, cette solution peut être difficile à réaliser. Les alternatives sont alors, soit de rester dans \mathcal{A} , qui a l'avantage de rester stable, tout en favorisant le voisement, et donc le contraste entre fricatives sourdes et voisées, soit d'effectuer un très court passage aller-retour $\mathcal{A}/\mathcal{B}/\mathcal{A}$ voire $\mathcal{A}/\mathcal{B}/\mathcal{C}/\mathcal{B}/\mathcal{A}$, sans maintien des positions articulaires et laryngées. Cette solution expliquerait alors le fait que les fricatives voisées sont généralement plus courtes que les sourdes [12].

Nous vérifions dans cette étude ces hypothèses d'instabilité du régime des fricatives voisées. Celle-ci suppose que le régime \mathcal{B} étant instable, les locuteurs devraient avoir des difficultés à le maintenir. Pour cela, nous comparons la production de fricatives voisées entre deux voyelles sans consigne particulière, avec la production de ces mêmes fricatives dont on aurait demandé au locuteur de les maintenir pendant un temps exagérément long (environ 1 s). Si l'hypothèse est vérifiée, nous devrions observer une tendance chez les locuteurs à maintenir un régime stable lors des fricatives longues, avec une préférence pour le régime \mathcal{A} car il permet d'assurer le contraste de voisement.

Corpus

Des fricatives ont été enregistrées en contexte intervocalique (entre deux voyelles) formant des logatomes de la forme voyelle-fricative voisée-voyelle (V_1FvV_2), où V_1 et V_2 sont choisies parmi les trois voyelles cardinales /i,a,u/ et Fv est choisie parmi les trois fricatives voisées du français, /v,z,ʒ/.

Cela donne donc un ensemble de vingt-sept combinaisons possibles de logatomes V_1FvV_2 . Dans le but de vérifier l'hypothèse selon laquelle les locuteurs évitent ou raccourcissent les zones d'instabilité, nous enregistrons des versions de chaque logatome au sein duquel la fricative centrale est exagérément rallongée, d'environ une seconde. Dans ce cas, les locuteurs devraient alors choisir de rester dans le régime \mathcal{A} afin de favoriser le voisement tout en restant dans une zone stable. Chaque logatome est enregistré trois fois en version fricative courte, et trois fois également en version fricative longue, soit un total de $27 \times 2 \times 3 = 162$ acquisitions par locuteur, et 54 par fricative et locuteur.

Les locuteurs sont au nombre de 13, allant de 20 à 61 ans au moment de l'enregistrement. Les détails de l'étude, et notamment l'existence des régimes de production des fricatives et l'hypothèse d'instabilité des fricatives voisées n'ont pas été communiqués aux locuteurs avant les enregistrements afin de garder une production la plus naturelle possible. Aucun des locuteurs ne souffre de trouble du langage reconnu et tous sont volontaires pour l'expérience.

Les signaux ont été enregistrés dans une salle calme, à l'aide d'un micro casque et d'une carte d'acquisition, et ont été acquis avec une fréquence d'échantillonnage de 20 kHz.

Analyse

La segmentation phonétique nécessaire pour isoler les portions de signal correspondant aux fricatives a été faite automatiquement à l'aide d'une méthode d'alignement texte-parole automatique [31], et les éventuelles erreurs corrigées à la main ensuite. La méthode de séparation X-GLOS a été appliquée à l'ensemble des signaux, et l'évolution temporelle du quotient de $VQ(t)$ est ensuite calculée. Nous définissons ensuite un facteur de dévoisement ΔVQ , qui correspond à l'écart en % entre les valeurs maximale et minimale du quotient de voisement pendant la production de la fricative.

Résultats

Mise en évidence expérimentale des différentes stratégies pour les fricatives courtes

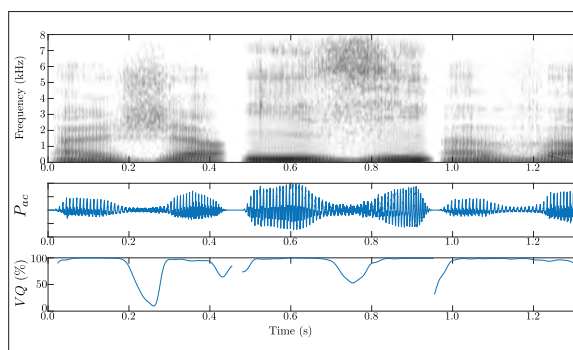


Fig. 3 : Spectrogramme large-bande (haut), signal de pression acoustique (milieu), et quotient de voisement VQ (bas) de trois réalisations (/aʒa/-/izi/-/ava/), illustrant bien l'existence de manières différentes de réaliser les fricatives voisées
Wide-band spectrogram (top), speech audio signal (middle), and voicing quotient VQ (bottom) of three tokens (/aʒa/-/izi/-/ava/), showing the existence of different ways to produce voiced fricatives

L'analyse des fricatives courtes révèle bien la présence de stratégies multiples de production des fricatives. La figure 3 représente ces différentes stratégies. Pour la première réalisation ($/a_3a/$), le locuteur dévoise de manière très importante sa fricative post-alvéolaire, le quotient de voisement chute brutalement, atteint de très faibles valeurs, avant de remonter rapidement. Pour la deuxième ($/izi/$), nous avons un schéma similaire, mais cette fois-ci le dévoisement de la fricative est beaucoup moins marqué : VQ est toujours supérieur à 50%. Enfin, pour le logatome $/ava/$, le locuteur maintient le régime \mathcal{A} , VQ étant constamment proche des 100%.

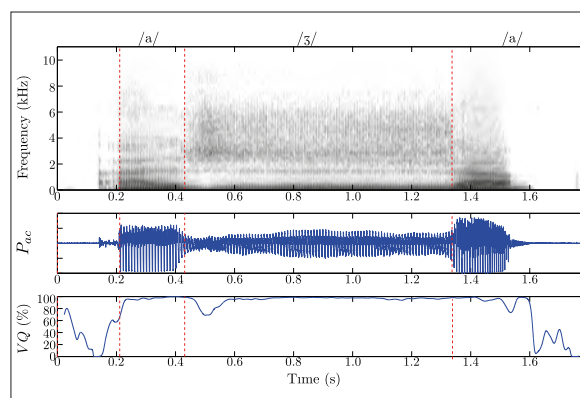
Analyse des fricatives longues

L'analyse qualitative des fricatives longues confirme l'hypothèse d'instabilité du régime \mathcal{B} . En effet, la très grande majorité des locuteurs choisissent de maintenir le régime \mathcal{A} pour favoriser le voisement. Cela se traduit alors par des réalisations similaires à la figure 4(a) : le quotient de voisement est constamment proche de 100% pendant la production de la fricative. À noter que dans ce cas précis, le locuteur effectue un très court passage dans \mathcal{B} , puis maintient \mathcal{A} pour retrouver la stabilité lui permettant de prolonger sa fricative voisée. Parmi l'ensemble des treize locuteurs, un seul a maintenu le régime \mathcal{B} pendant la production des fricatives longues. Un exemple est donné sur la figure 4(b). Il l'a effectué pour l'ensemble des réalisations des fricatives longues post-alvéolaires et alvéolaires, et jamais pour les fricatives labio-dentales. À noter que malgré le fait qu'il maintienne le régime \mathcal{B} , le quotient de voisement est erratique et présente de grandes fluctuations.

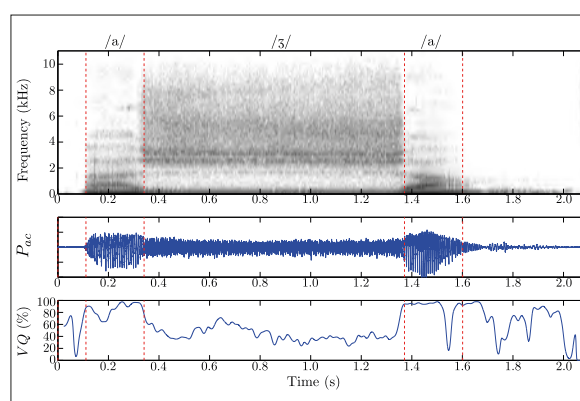
Comparaison fricatives courtes vs fricatives longues

Après cette introduction des résultats à travers quelques exemples, la figure 5 présente des résultats quantitatifs et statistiques. Ils confirment le fait que les locuteurs choisissent très majoritairement de rester dans le régime \mathcal{A} dès lors que la fricative est longue. En effet, à de rares exceptions près, qui proviennent d'ailleurs toutes du même locuteur, les valeurs du coefficient de dévoisement est toujours très bas pour les fricatives longues. Ces exceptions sont d'ailleurs considérées comme des données statistiquement aberrantes, distantes de la valeur médiane de plus de trois fois l'écart-type médian.

Pour les fricatives courtes, les valeurs de ΔVQ sont très dispersées, allant de 0 à 100%. Cette grande dispersion souligne l'existence de différentes stratégies ou manières de prononcer les fricatives voisées du français en contexte intervocalique : elles peuvent être réalisées à la manière d'approximantes (le voisement est nettement prédominant, d'où un ΔVQ quasi-nul), elles peuvent être partiellement dévoisées (niveaux de voisement et de bruit de friction équivalents), voire complètement dévoisées pendant un bref intervalle de temps ($\Delta VQ \approx 100\%$). Le centre de gravité des ΔVQ est par ailleurs plutôt bas, indiquant que les locuteurs choisissent de favoriser le voisement au détriment du bruit de friction, très probablement pour favoriser le contraste voisée/non-voisée. Ces remarques sont en revanche moins nettes pour les fricatives labio-dentales, qui sont quasi-exclusivement prononcées avec un faible ΔVQ . Pour les fricatives labio-dentales longues, ΔVQ ne dépasse jamais 5%.



a)



b)

Fig. 4 : Spectrogramme large-bande (haut), signal de pression acoustique (milieu), et quotient de voisement VQ (bas) de deux réalisations avec fricatives longues ($/a_3a/$) effectuées par deux locuteurs différents
Wide-band spectrogram (top), audio speech signal (middle), and voicing quotient VQ (bottom) of two tokens with a long fricative ($/a_3a/$) produced by two different speakers

Ces résultats valident donc expérimentalement deux hypothèses issues de l'analyse numérique de notre étude précédente [18]. Premièrement, le régime de production théorique des fricatives voisées (\mathcal{B}) est instable, ce qui contraint les locuteurs à rester dans le régime stable \mathcal{A} en cas de prolongement de la fricative, dans le but d'assurer le contraste voisée/non-voisée (ils favorisent la production de voisement plutôt que le bruit de friction). Deuxièmement, les différents régimes de production permettent aux locuteurs d'employer des stratégies différentes de production des fricatives.

Conclusions

Nous avons présenté dans cet article une méthode de traitement de signal permettant de séparer les contributions des sources glottiques des sources supraglottiques dans la parole. Des tests sur des signaux numériques montrent qu'elle est plus performante que les méthodes concurrentes [20, 21], qu'elle est très peu sensible au niveau de

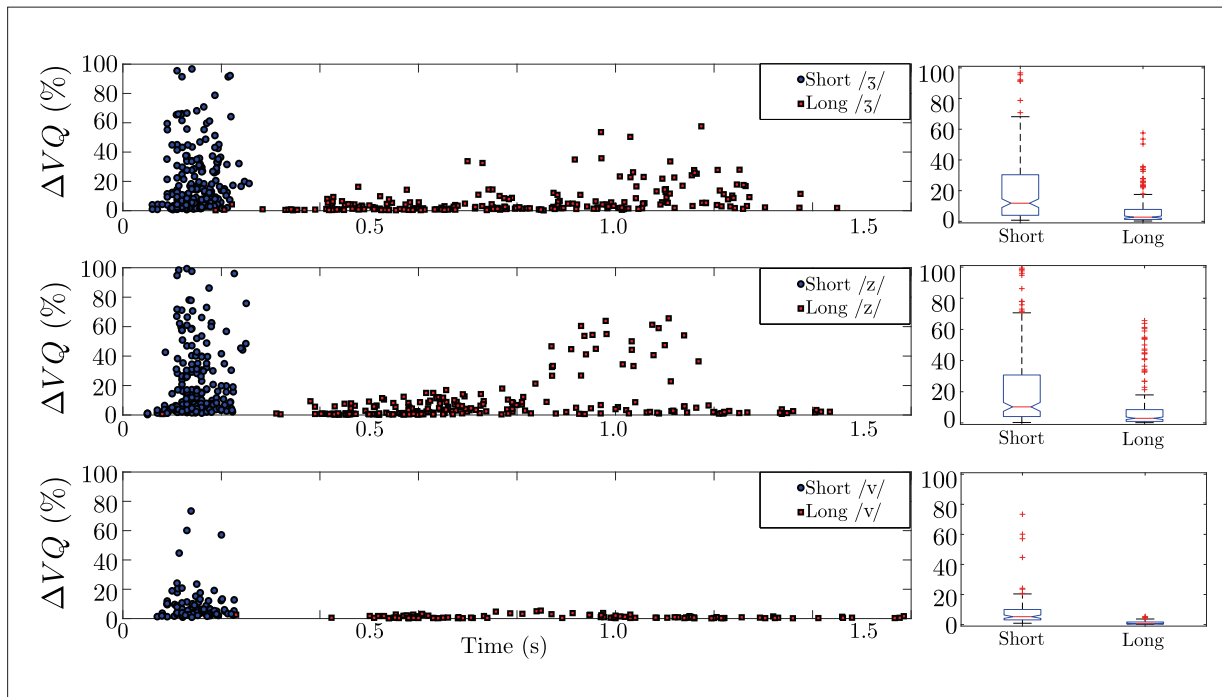


Fig. 5 : Coefficient de dévoisement ΔVQ pour l'ensemble des signaux analysés, regroupés par point d'articulation de la fricative voisée. De haut en bas : palato-alvéolaire /ʒ/, alvéolaire /z/, et labio-dental /v/. Gauche : nuage de points ΔVQ en fonction de la durée de la fricative. Droite : diagramme en boîte. Pour chaque boîte, la ligne centrale indique la valeur médiane, les valeurs supérieures et inférieures de la boîte indiquent les 75e et 25e percentiles, respectivement. Les lignes traitillées indiquent les valeurs extrêmes non considérées comme aberrantes statistiquement, et les croix '+' indiquent les valeurs aberrantes statistiquement

Devoicing index ΔVQ for all of the analyzed signals, gathered by place of articulation of the voiced fricative. From top to bottom: palatoalveolar /ʒ/, alveolar /z/, and labio-dental /v/. Left: scatter plot ΔVQ as a function of the duration of the fricative. Right: boxplot. For each box, the central line is the median value, upper and lower values of the box are the 75th and 25th percentiles, respectively. Dashed lines are the extreme values that are not considered as outliers, and cross markers '+' are the outliers

bruit et au niveau de jitter. Elle permet donc une estimation précise du quotient de voisement VQ , même lorsque celui est théoriquement très faible, tel qu'aux frontières voyelle-consonne, par exemple, où lorsque les variations locales de fréquence fondamentale sont très fortes (fort jitter). Ceci est notamment dû au fait de la prise en compte des non-stationnarités du signal glottique par l'estimation des paramètres instantanés des sinusoides des contributions glottiques.

L'application de la méthode à des signaux de parole réels issus d'un corpus spécifiquement enregistré pour cette étude a permis de valider expérimentalement des hypothèses issues d'une précédente étude numérique [18]. Ces hypothèses suggèrent qu'il existe une zone d'instabilité lors de la production des fricatives qui correspond aux fricatives voisées : dans cette zone, de faibles perturbations de l'ouverture glottique entraînent de grandes variations des indices acoustiques, et notamment du quotient de voisement. L'existence de cette instabilité perturbe alors les locuteurs lorsqu'ils doivent maintenir des fricatives voisées, ou bien les contraignent à raccourcir les fricatives voisées par rapport à leur contreparties sourdes, qui sont produites dans un régime stable.

De manière qualitative, cela se traduit par une diminution nette du coefficient de dévoisement ΔVQ lorsque les fricatives voisées sont maintenues, par rapport au cas où elles sont produites proches des conditions de parole

naturelle. À l'exception d'un locuteur parmi les treize participants, tous ont tenu le régime stable de fort voisement pour les fricatives longues, alors que des passages dans le régime instable de voisement intermédiaire ont été observés de manière plus fréquente pour les fricatives courtes. L'instabilité des fricatives voisées apporte un nouvel éclairage sur des phénomènes couramment observés en pratique, comme le dévoisement des fricatives en fin de mot en allemand ou en russe, par exemple. Des études comparatives avec des locuteurs de langues différentes pourraient ainsi fournir des éléments de réponses sur les raisons qui amènent certaines langues à neutraliser le contraste de voisement dans des contextes phonologiques particuliers.

Cette étude a également permis de souligner les différences entre fricatives voisées post-alvéolaires et alvéolaires /ʒ,z/ et les fricatives labio-dentales /v/ : ces dernières sont systématiquement produites dans le régime de fort voisement, alors que les autres sont produites avec des coefficients de dévoisement très variables, allant de 0% à 100%. Ces remarques posent de nouvelles questions relatives à la classification des phonèmes du français, et de manière générale à la classification des différentes réalisations des fricatives.