



HAL
open science

Interaction entre durée et position dans la perception des fricatives voisées chuchotées

Yohann Meynadier, Noël Nguyen, Sophie Dufour

► **To cite this version:**

Yohann Meynadier, Noël Nguyen, Sophie Dufour. Interaction entre durée et position dans la perception des fricatives voisées chuchotées. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition), 2020, Nancy, France. pp.442-450. hal-02798571v1

HAL Id: hal-02798571

<https://hal.science/hal-02798571v1>

Submitted on 7 Jun 2020 (v1), last revised 23 Jun 2020 (v3)

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Interaction entre durée et position dans la perception des fricatives voisées chuchotées¹

Yohann Meynadier, Noël Nguyen, Sophie Dufour
Aix Marseille Univ, CNRS, LPL, UMR7309, Aix-en-Provence, France
Aix Marseille Univ, ILCB, Marseille, France

{yohann.meynadier ; noel.nguyen-trong ; sophie.dufour}@univ-amu.fr

RÉSUMÉ

Cette étude s'intéresse à la reconnaissance du trait de voisement en parole chuchotée. Nos travaux antérieurs (Dufour & Meynadier 2019) montrent une reconnaissance plus tardive du trait [+voisé] reposant sur un traitement probablement pré-lexical d'informations acoustiques, autres que la vibration laryngée, extraites du signal chuchoté et utilisées dans l'accès lexical. Via une tâche d'identification en perception catégorielle, cette étude révèle que la durée conditionne la perception du voisement des fricatives chuchotées : plus /ʃ/ est long, plus il est perçu [-voisé] ; plus /ʒ/ est bref, plus il est identifié [+voisé]. Cet effet de durée est modulé par le trait sous-jacent de voisement et la position dans le (non)mot. La fricative [+voisé] en position finale montre une frontière perceptive particulièrement décalée vers des durées beaucoup plus longues que les autres.

ABSTRACT

Interplay between duration and word position in voicing perception of whispered fricatives

This study focuses on the recognition of the voicing-feature in whispered speech. Our previous works (Dufour & Meynadier 2019) show a later recognition of the feature [+voiced] probably based on a pre-lexical processing of acoustic information, other than the laryngeal vibration, extracted from the whispered signal and used in lexical access. Via an identification task in categorical perception, this study reveals that duration influences the perception of the voicing of whispered fricatives: the longer /ʃ/ is, the more it is perceived [-voiced]; the shorter /ʒ/ is, the more it is identified [+voiced]. This duration effect is modulated by the underlying voicing-feature and the position within the (non-sense) word. The fricative [+voiced] in word-final position shows a perceptual boundary particularly shifted towards durations much longer than the others.

MOTS-CLÉS : voisement, parole chuchotée, perception, fricative, français

KEYWORDS: voicing, whispered speech, perception, fricative, French

1 Introduction

Le chuchotement crée une opacité du trait de voisement entre la forme phonétique et la représentation phonologique des mots en français. Cette opacité provient du fait qu'en parole chuchotée le trait de voisement ne peut plus être rattaché à la vibration des cordes vocales à un

¹ Une version résumée de ce travail a été acceptée à la 17^e International Conference on Laboratory Phonology, 6-8 juillet 2020, Vancouver, Canada (LabPhon17).

niveau phonétique, mais doit être associé à d'autres propriétés acoustiques secondaires pour être correctement reconnu.

Dans le versant de la perception, les travaux sur l'anglais rapportent que les obstruantes voisées chuchotées sont reconnues au-delà du seuil du hasard (Dannenbring 1980, Tartter 1989, Munro 1990, Gilichinskaya & Strange 2011). Pour le français, Dufour & Meynadier (2019) ont précisé le décours temporel de la reconnaissance des mots chuchotés à obstruantes voisées via un paradigme d'amorçage audiovisuel de décision lexicale. Les résultats montrent que la reconstruction du trait [+voisé] des obstruantes chuchotées prend plus de temps que la reconnaissance du trait [-voisé]. Sans être totalement homophones aux obstruantes sourdes, les obstruantes voisées chuchotées restent temporairement ambiguës. L'ambiguïté n'est levée que si un temps supplémentaire de traitement de 50 ms est accordé au sujet. L'hypothèse est que dans ce bref laps de temps, l'auditeur extrait du signal de parole chuchotée les informations acoustiques nécessaires à la reconnaissance de l'obstruante voisée.

Dans le versant de la production, des travaux récents sur le français montrent que les consonnes voisées chuchotées, bien que produites sans vibration des cordes vocales, conservent certaines traces phonétiques de leur identité sous-jacente [\pm voisé] (pour une revue Meynadier & Dufour 2016, 2018). Notamment, les durées acoustiques des consonnes et des voyelles préconsonantiques paraissent être les corrélats du voisement les plus robustes en parole chuchotée (Vercherand 2010, Meynadier & Gaydina 2013, Meynadier & Dufour 2018). Par ailleurs, en parole modale la manipulation de la durée des obstruantes influence l'identification de leur trait de voisement (Denes 1955). Reste qu'aucune étude, à notre connaissance, ne s'est intéressée à ce point en parole chuchotée.

Afin de clarifier l'opacité du contraste de voisement en parole chuchotée, trois hypothèses sont testées ici. Premièrement, sur la base des travaux précités, nous testons ici la possibilité que dans le bref laps de temps supplémentaire dans la reconnaissance des obstruantes voisées chuchotées, mis en évidence par Dufour & Meynadier (2019), le sujet est capable d'extraire de la parole les informations acoustiques relatives à la durée des consonnes afin d'identifier perceptivement leur trait de voisement. Nous nous attendons donc à ce que les résultats de perception soient en accord avec ceux de production rapportés par la littérature, à savoir que plus la consonne est longue, plus elle est associée au trait [-voisé], et inversement, plus elle est courte, plus elle est identifiée comme [+voisé]. Deuxièmement, nous cherchons à établir à quel niveau de traitement l'information de durée segmentale est recrutée pour recouvrer le voisement sous-jacent de la consonne. Suite à Dufour & Meynadier (2019), nous supposons qu'un traitement pré-lexical des indices acoustiques est impliqué dans la reconnaissance du 'voisement chuchoté'. Nous nous attendons donc à ce que l'effet de durée de la consonne soit similaire pour des mots et pour des non-mots. Enfin, nous testons ici également si un effet de position de la consonne dans le (non)mot influence l'identification de son voisement sous-jacent. Sur ce point, nous pensons que l'allongement spécifique des consonnes finales de (non)mots isolés (cf. Table 1), du fait de l'allongement final prosodique, pourrait décaler la frontière catégorielle d'identification du trait de voisement.

Pour répondre à ces questions, la durée des fricatives / ʃ / et / ʒ / en position initiale et finale de mots et de non-mots a été manipulée dans une tâche de perception catégorielle du trait [\pm voisé] de la fricative.

2 Protocole expérimental

Un test d'identification portant sur les fricatives / ʃ / versus / ʒ / a été mené pour évaluer si la durée de la consonne permet de faire basculer la perception catégorielle du voisement de la consonne chuchotée.

2.1 Matériel linguistique

Les fricatives ont été choisies pour représenter le contraste de voisement afin de simplifier la manipulation de la durée acoustique de la consonne. En effet, les plosives, présentant deux phases articulatoires distinctes, occlusion et explosion, obligeraient à un contrôle de leur durée plus complexe. Par ailleurs, les modifications de durée de l'occlusion silencieuse des plosives chuchotées ne sont pas perceptibles en initiale absolue de mot. Seul le contraste /ʃ/-ʒ/ est ici utilisé.

/ʃ/-ʒ/ ont été produites par un locuteur français dans des séquences CVC formant soit une paire minimale, en position initiale de mot : « *char* » vs « *jarre* » et en position finale de mot : « *cache* » vs « *cage* », soit une paire de non-mots se distinguant seulement par le voisement de la fricative, en initiale : « *cheur* » vs « *jeur* » et en finale « *queuche* » vs « *queuge* ». Ces quatre paires ont été enregistrées en chambre sourde à l'aide d'un microphone Senheiser HD415 (échantillonnage : 32 kHz ; 32 bits). Une distance de 30 cm a été maintenue constante entre le micro et la bouche du locuteur qui a produit naturellement les mots en voix chuchotée non projetée ou forcée (comme chuchotés à l'oreille d'un interlocuteur à distance proximale). L'absence de vibrations glottiques du signal chuchoté, ainsi que la différence de durée attendue (pour des références en voix chuchotée et modale, Dufour & Meynadier 2019) entre les consonnes cibles et entre les voyelles préconsonantiques selon le voisement de la fricative (Table 1) ont été vérifiées via Praat (Boersman & Weenink 2017).

	C	V	C		C	V	C
<i>char</i>	138	-	-	<i>cheur</i>	154	-	-
<i>jarre</i>	106	-	-	<i>jeur</i>	111	-	-
<i>cache</i>	-	185	253	<i>queuche</i>	-	199	243
<i>cage</i>	-	240	161	<i>queuge</i>	-	230	156

TABLE 1 : Durée (ms) des fricatives cibles et des voyelles préconsonantiques des mots et non-mots utilisés pour la génération des stimuli de l'expérience

À l'aide du Vocal Toolkit (Corretge 2019) de Praat, la durée acoustique de la fricative a été modifiée par resynthèse selon un continuum de 11 pas de 20 ms allant de 50 à 250 ms. L'intensité globale (RMS) de chaque fricative cible a été neutralisée à la valeur moyenne calculée entre celle de la [+voisé] et celle de la [-voisé] de la paire de (non)mots concernée. Les portions restantes *_CV* (*_ar*, *_eur*) et *CV_* (*ca_*, *queu_*) ont également été neutralisées en durée, en intensité et en fréquence. Les portions de chaque paire [\pm voisé] de (non)mots ont dans un premier temps été resynthétisées avec une durée neutre correspondant à la durée moyenne entre les deux membres de la paire. Un morphing acoustique (spectral) à mi-chemin entre ces deux mêmes exemplaires a été appliqué (via la fonction *Mix...* du Vocal Toolkit de Praat). Enfin, l'intensité a été homogénéisée de la même manière que pour les fricatives cibles. Les 4 paires x 2 (non)mots x 11 pas de durée, soit 88 stimuli, ont été contrôlés afin que seule la différence de durée de la fricative varie. À noter que le profil spectral (autre que l'intensité globale) de la fricative cible n'a pas été neutralisé par morphing, et que donc il est susceptible de différences selon le voisement sous-jacent de la consonne. Néanmoins, à notre connaissance, aucune différence spectrale systématique due au voisement de la consonne n'a encore été mise en évidence dans la littérature (Meynadier & Dufour 2018).

Enfin, les stimuli ont été reconstruits par concaténation de la fricative cible et de la portion restante manipulées du (non)mot. Afin de lisser la transition entre ces deux parties du stimulus, un amortissement initial et/ou final (fonction *Fade...* du Vocal Toolkit de Praat) sur 30 ms a été appliqué sur le signal. Ces deux parties ont alors été concaténées avec un chevauchement de même durée.

2.2 Expériences

Ainsi, 8 continua de durée consonantique, soit 2 (mots + non-mots) en paire de voisement x 2 positions x 11 pas de durée de la fricative, ont été soumis à une tâche d'identification dans un test de perception catégorielle. 32 sujets (21 F, 11 H) ont passé l'expérience sur les mots, 32 autres (26 F, 6 H) sur les non-mots. Chaque sujet a répondu sur les paires à fricative en position initiale (*char* vs *jarre* ou *cheur* vs *jeur*) et en position finale (*cache* vs *cage* ou *queuche* vs *queuge*) dans deux sessions séparées par une pause ; l'ordre de ces deux sessions a été balancé entre les sujets. À l'intérieur de chaque session, les stimuli à fricative [+voisé] et [-voisé] ont été aléatoirement ordonnés différemment pour chaque sujet. Les tests ont été implémentés sur ordinateur via le logiciel Perceval (André et al. 2003).

Les 64 sujets de l'expérience ont été rémunérés 10 € pour un temps de passation d'environ 30 min. Aucun n'a rapporté de troubles de l'audition, de la parole ou du langage, ni de troubles attentionnels. Ils ont passé le test individuellement dans un box expérimental calme. Les stimuli ont été entendus au moyen d'un casque audio circum-aural Superflux HD681B et les réponses saisies sur un boîtier à 2 boutons. La position des (non)mots à fricative [+voisé] et [-voisé] était indiquée orthographiquement sur le boîtier et demeuraient constante lors de la session. Elle était inversée entre les deux sessions du test et balancée entre les sujets de l'expérience. Le sujet était face à un écran sur lequel il a lu les consignes et où apparaissait en son milieu un point de fixation durant les 750 ms précédant l'écoute du stimulus, afin de maintenir son attention. Un silence de 1,5 sec. séparait chaque stimulus. Le niveau d'écoute a été maintenu à un volume fixe, confortable et représentatif de la parole chuchotée durant toute l'expérience et entre les sujets. La tâche du sujet consistait à identifier le (non)mot entendu en appuyant sur le bouton correspondant du boîtier-réponse immédiatement après la fin acoustique du stimulus, sans pression ou limite de temps. Chaque stimulus a été présenté 10 fois dans la même session. Chaque sujet a donc pris sa décision de voisement sur 220 stimuli pour chacune des deux sessions du test.

3 Résultats

Nous rapportons ici les résultats statistiques obtenus sur le pourcentage moyen de réponses correctes aux tests. Une réponse est dite 'correcte' quand l'auditeur a choisi le (non)mot comportant la fricative dont le trait sous-jacent de voisement correspond bien à celui de la fricative du stimulus entendu. Dans un premier temps, sont présentés les statistiques et les résultats relatifs à l'effet du continuum de durée de la fricative sur le pourcentage d'identification de son voisement. Ensuite, les analyses et les résultats statistiques concernant les différences en durée de frontière catégorielle d'identification du voisement sont présentés afin de rendre compte des interactions de la durée de la fricative avec son voisement sous-jacent, sa position dans le (non)mot et la lexicalité du stimulus.

3.1 Effet de la durée

Les réponses des participants (1 = réponse correcte, 0 = réponse incorrecte) ont été analysées grâce à un modèle de régression *logit* à effets mixtes (Baayen 2008 ; Baayen et al., 2008). Le modèle incluait la durée des consonnes (11 niveaux), le voisement ([+voisé], [-voisé]), la position (initiale, finale) et la lexicalité (mot, non-mot), ainsi que leurs interactions, comme effets fixes. Le modèle incluait également l'intercept et la pente aléatoires par participants pour l'effet de la durée des consonnes, du voisement et de la position (Barr et al., 2013). L'ensemble de données étant trop volumineux pour les paramètres par défaut de *glmer.nb*, l'option *nAGQ=0* a été implémentée afin de permettre au modèle de converger (Bates et al., 2016). Le modèle a été appliqué sur 28159 données (1 réponse n'a pas été enregistrée par le logiciel).

De façon intéressante le modèle a révélé une interaction significative entre la durée des fricatives et leur voisement [$\chi^2= 5341,61$; $p < .0001$]. Plus la fricative /f/ est longue, plus elle est perçue [-voisé], et inversement plus la fricative /ʒ/ est longue, moins elle est perçue [+voisé] (autrement dit, plus /ʒ/ est brève, mieux elle est identifiée comme [+voisé]). La Figure 1 montre que cet effet est vrai que la fricative soit produite en position initiale ou finale de mot ou de non-mot.

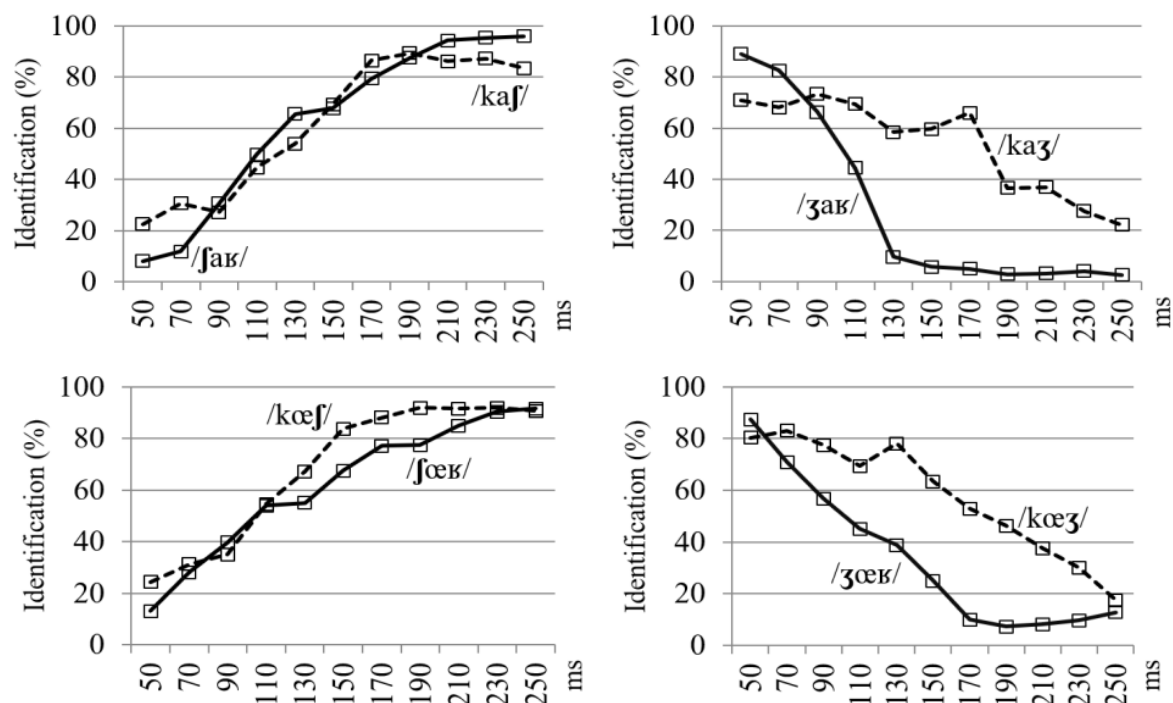


FIGURE 1 : Pourcentages moyens d'identification du trait de voisement en fonction de la durée de chaque fricative chuchotée selon sa position dans un mot (haut) ou un non-mot (bas)

3.2 Effet du voisement, de la position et de la lexicalité

Pour comprendre comment le trait de voisement, la position dans le (non)mot et la lexicalité du stimulus influencent l'identification du trait de voisement selon la durée de la fricative chuchotée, nous avons étudié les différences entre les frontières de perception catégorielle des différentes conditions. L'analyse statistique des interactions entre la durée de la fricative et les facteurs *voisement*, *position* et *lexicalité* a été effectuée sur les valeurs de durée des frontières perceptives catégorielles du trait [\pm voisé] des fricatives cibles. Pour chaque sujet (N = 64) dans chaque condition, à savoir pour chaque fricative (/f/, /ʒ/) dans chaque position (initiale, finale) de chaque (non)mot (*char*, *jarre*, *cache*, *cage*, *cheur*, *jeur*, *queuche*, *queuege*), la frontière catégorielle a été calculée à partir de la fonction spline cubique des pourcentages d'identification du voisement (à savoir les pourcentages moyens de réponses correctes) selon les 11 pas du continuum de durée de la fricative cible. Cette frontière correspond au point de la courbe où la perception du trait [\pm voisé] de la fricative bascule. Elle est déterminée par la détection de la durée en ms (x) qui conduit à l'intercept (y) entre la courbe interpolée et un taux de réponse correcte à 50%, à savoir à 50% d'identification (c'est-à-dire le seuil de hasard de réponses binaires). 11 sujets (6 pour les mots et 4 pour les non-mots) ont été exclus des analyses. Cette exclusion correspond à une absence de bascule catégorielle dans la perception du voisement de la fricative en fonction de sa durée dans au moins une des conditions du test. Soit le sujet à un comportement trop incohérent, soit ses pourcentages d'identification sont toujours inférieurs ou supérieurs au seuil de 50% quelle que soit la durée de la

fricative cible. Ainsi, 17,2% des données de frontière (soit 44 valeurs sur 256, à savoir 32 sujets x 8 (non)mots) ont été exclues.

Les valeurs moyennes de frontière (durée en ms) ont été analysées via une ANOVA à trois facteurs dont deux à mesures répétées (*voisement*, *position*) et un à mesures indépendantes (*lexicalité*). La triple interaction *voisement*position*lexicalité* est significative [$F(1, 54) = 4,32$; $p < .05$]. Afin d'étudier cette interaction, nous rapportons les effet simples de ces facteurs et les interaction à deux facteurs, ainsi que les comparaisons multiples deux à deux a posteriori entre les conditions expérimentales avec une correction de Bonferroni. Les comparaisons d'intérêt entre *voisement*, *position* et *lexicalité* sont rapportés dans la Table 2.

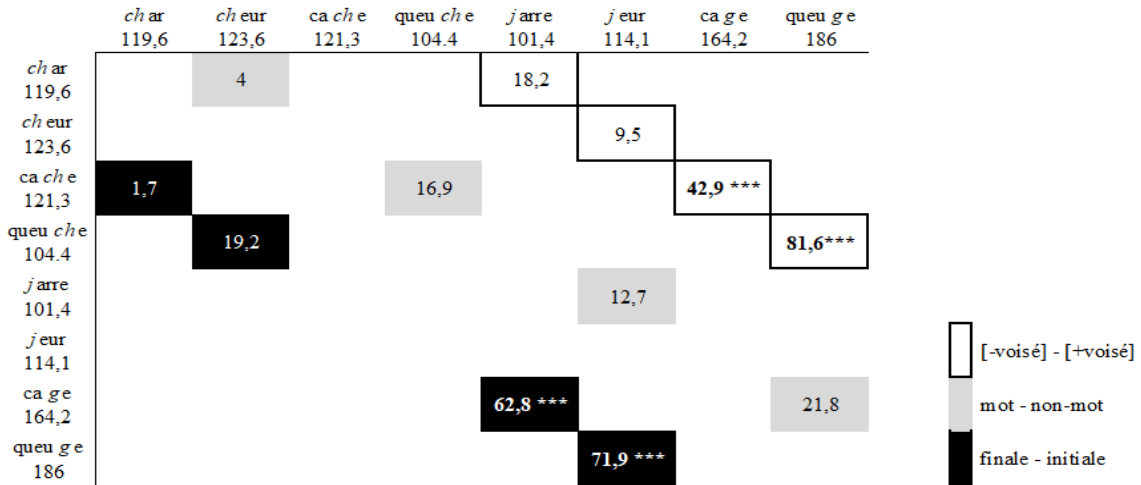


TABLE 2 : Valeurs (sous le nom du stimuli) et écarts de durée (en ms) des frontières de perception catégorielle du trait de voisement des fricatives chuchotées selon le voisement, la position dans le (non)mot et la lexicalité du stimulus. *** signale une différence significative avec $p < .0001$; l'absence d'astérisque indique une différence non significative ($p > .05$)

Toutes conditions confondues, seuls le *voisement* et la *position* de la fricative montrent un effet significatif sur les valeurs de durée des frontières catégorielles, respectivement [$F(1, 54) = 52,10$; $p < .0001$] et [$F(1, 54) = 34,87$; $p < .0001$]. Ces deux facteurs ont également une interaction significative [$F(1, 54) = 117,60$; $p < .0001$]. La Figure 2 montre ainsi que la frontière catégorielle de la fricative [+voisé] finale est spécifiquement plus haute que les autres. Elle atteint 164 ms pour le mot « cage » et 186 ms pour le non-mot « queue », alors que toutes les autres se situent globalement autour de 100 à 120 ms. La Table 2 résume les tests post-hoc de ces facteurs. Elles montrent que seule la frontière catégorielle de la fricative [+voisé] finale se distingue significativement de celle de la fricative [-voisé] finale et de celle de la fricative [+voisé] initiale, pour les mots (respectivement [$F(1, 52) = 38,01$; $p < .0001$] et [$F(1, 52) = 48,19$; $p < .0001$]) comme pour les non-mots (respectivement [$F(1, 52) = 144,46$; $p < .0001$] et [$F(1, 52) = 66,26$; $p < .0001$]). Le déplacement vers une valeur plus élevée de la frontière perceptive de la fricative [+voisé] finale par rapport aux autres oscille entre 43 et 82 ms.

Enfin, une interaction significative entre le *voisement* et la *lexicalité* est également rapportée [$F(1, 54) = 12,201$; $p < .001$]. Elle indique que toutes positions confondues seules les frontières catégorielles de la fricative [+voisé] sont un peu plus hautes pour les non-mots (queue, jeur) que pour les mots (cage, jarre). Cela est illustré par la Figure 2. L'écart de frontière est en moyenne de 22 ms entre « cage » et « queue », et de 13 ms entre « jarre » et « jeur » (Table 2). Pour la fricative [-voisé], nous n'observons pas de régularité de ce point de vue. Reste que les tests post-hoc,

rapportés dans la Table 2, ne signalent aucun de ces écarts de frontière entre mot et non-mot comme étant significatif.

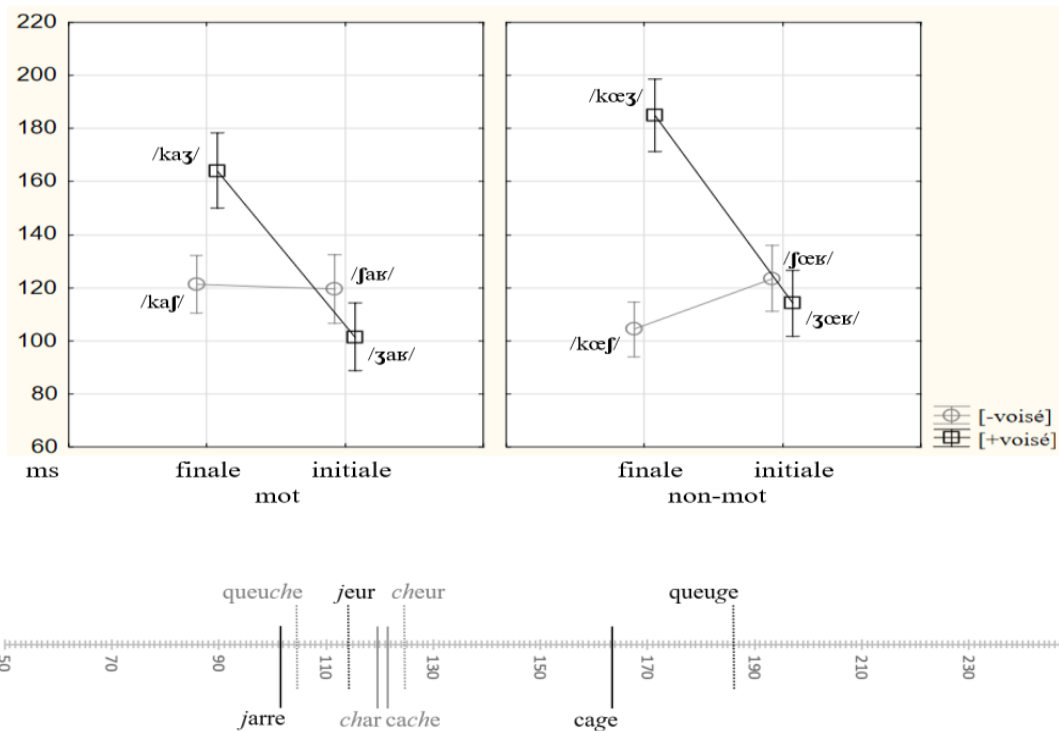


FIGURE 2 : Valeurs moyennes (en haut) et position sur le continuum (en bas) des durée des frontières de perception catégorielle du trait de voisement des fricatives chuchotées selon le voisement et la position pour les mots et les non-mots. Sur le graphique du haut, les barres représentent les intervalles de confiance à 95%

4 Discussion

Dans tous les cas, des durées plus longues favorisent toujours la perception du trait [-voisé], et plus courtes celle du trait [+voisé] (Figure 1). Les courbes d'identification moyenne du voisement montrent une influence plutôt graduelle que catégorielle. Elles ont également une forme particulière qui pourrait être une trace du biais perceptif en faveur du trait [-voisé] des obstruantes chuchotées noté dans la littérature (cf. Dufour & Meynadier 2019). En effet, elles dessinent un plateau relativement stable d'une perception en fricative [-voisé] au moins à partir de 170 ms, pour /ʃ/ mais également pour /ʒ/, sauf si ce dernier est final de (non)mot. De plus, toute fricative chuchotée, autre que /ʒ/ en position finale, à partir d'une durée de 110 ms, est majoritairement perçue en /ʃ/, et en deçà en /ʒ/. Les résultats montrent donc que la durée consonantique module la perception du trait de voisement des fricatives chuchotées. La reconnaissance de leur voisement sous-jacent s'appuierait bien sur le corrélat phonétique de durée consonantique tel qu'observé en production de la parole, à savoir sur la durée plus longue des obstruantes [-voisé] par rapport aux [+voisé]. L'hypothèse que le sujet est capable d'extraire de la parole chuchotée les informations acoustiques relatives à la durée des consonnes afin d'identifier perceptivement leur trait de voisement est donc bien confirmée. Cela plaide pour une reconstruction du trait à partir du traitement de détails phonétiques fins présents dans le signal acoustique.

Pour aller, plus loin nous avons aussi tester l'effet de la lexicalité des stimuli. Les analyses des frontières perceptives du trait de voisement selon la durée consonantique n'ont pas révélé de différences significatives entre les fricatives insérées dans des mots et celles insérées dans de non-

mot (Table 2). Ce résultat irait dans le sens d'une extraction du trait de voisement à partir des corrélats de durée segmentale présents dans le signal de parole, et donc de l'implication d'un traitement pré-lexical de ces indices acoustiques lors de la reconnaissance d'un trait, d'un phonème et/ou d'un mot dont la forme phonétique est altérée ou plus éloignée de sa forme phonologique.

Enfin, de manière intéressante, cette étude montre, pour la première fois, que l'influence de la durée consonantique n'est pas homogène, mais est fonction de l'interaction entre le trait de voisement de la fricative et sa position dans le (non)mot. En effet, la Figure 1 montre que pour une même durée et au moins à partir de 90 ms, la fricative [+voisé] est toujours mieux reconnue en finale qu'en initiale de (non)mot. Ainsi, seule la fricative [+voisé], et seulement si elle se trouve en position finale de (non)mot, a une frontière perceptive du voisement selon la durée significativement plus haute que celle des fricatives dans toute autre condition. Sa frontière apparaît donc décalée vers une durée consonantique beaucoup plus longue, de 164 (pour « cage ») à 186 ms (pour « queuge »), par rapport aux autres situées autour de 100 à 120 ms. Ainsi, par rapport une position en initiale, une fricative [+voisé] finale nécessite un allongement de 60-70 ms (Table 2) pour erronément être perçue [-voisé]. La perception du trait [+voisé] de la fricative chuchotée résisterait spécifiquement mieux à un allongement en position finale de mot.

Cette asymétrie de l'effet de durée en fonction de la position dans le mot pourrait s'expliquer par différents mécanismes, qui restent spéculatifs à ce stade de notre étude. En effet, le profil spectral de fricatives [+voisé] et [-voisé] de nos stimuli n'ont pas été neutralisé (par exemple par un morphing entre les deux spectres). La possibilité de différences spectrales corrélatives au voisement sous-jacent de la fricative présentes dans le signal et utiles à l'auditeur n'est donc pas exclue. Cette possibilité reste cependant une question ouverte sachant qu'à notre connaissance, aucune étude spectrale en parole n'a pu véritablement attester de différences systématiques liées au voisement dans le spectre de bruit des obstruantes chuchotées (cf. Meynadier & Dufour 2018). Si ces traces existent et sont exploitées par l'auditeur, il reste toujours à expliquer pourquoi seule la fricative [+voisé] et seulement en position finale montre une résistance spécifique au changement de durée, à savoir l'allongement. Une piste pourrait s'appuyer sur l'influence bien connue de la structure prosodique (dont syllabique et lexicale) sur la réalisation phonétique des phonèmes (connue sous la notion de *phonetic encoding of prosodic structure*, par exemple Cho 2016). Les segments montrent des variations acoustiques selon leur position et le poids de la frontière prosodique. L'une de ces variations concerne l'allongement en fin de constituant réalisé par une décélération du tempo de parole. Les segments finaux sont ainsi systématiquement plus longs en position finale qu'en position initiale, comme l'illustre, par exemple, les durées segmentales de nos stimuli présentées dans la Table 1. Or, des études récentes mettent en évidence que la perception d'une frontière prosodique, marquée par un allongement final, module la catégorisation des phonèmes en frontière (Steffman 2019, Mitterer et al. 2016 ; pour une revue et une discussion). Dans la lignée de ces travaux, on pourrait donc supposer ici que l'auditeur recrute ses connaissances prosodiques lors de la reconnaissance de traits segmentaux qui lui permettent d'opérer une compensation perceptive de l'allongement subi par la fricative [+voisé] chuchotée en position finale. Mais alors, reste à savoir pourquoi seule la fricative [+voisé], et non la [-voisé], montre cet effet de position ? L'absence d'effet de la position finale pour la fricative [-voisé] chuchotée pourrait s'expliquer par le fait qu'une fricative longue serait par défaut interprétée comme [-voisé] et que le biais perceptif vers le trait [-voisé] (relevé dans la littérature) jouerait pour une reconnaissance immédiate des fricatives [-voisé] allongées en finale de mot. Leur reconnaissance ne nécessiterait donc pas de traitement particulier dans cette position. Inversement, la reconnaissance du trait [+voisé] chuchoté engage un traitement spécifique et plus coûteux des détails phonétiques corrélatifs du voisement (Dufour & Meynadier 2019). La durée faisant partie de ces traces acoustiques du voisement, en position finale, du fait de l'allongement final prosodique, la fricative [+voisé] chuchotée serait potentiellement plus ambiguë et nécessiterait un traitement supplémentaire impliquant le recrutement d'informations prosodiques (top-down), permettant la compensation perceptive relative à sa durée segmentale.

Remerciements

Financement de l'ANR-16-CONV-0002 (ILCB), ANR-11-LABX_0036 (BLRI) et A*MIDEX. Et à Thierry Legou (LPL) pour son aide technique dans l'analyse des données.

Références

- ANDRÉ C., GHIO A., CAVÉ C., TESTON B. (2003). PERCEVAL: a computer-driven system for experimentation on auditory and visual perception. *Proceedings of the 15th ICPHS*, 1421-1424.
- BAAYEN R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- BAAYEN R. H., DAVIDSON D. J., BATES, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- BARR D. J., LEVY R., SCHEEPERS C., TILY H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- BATES D., BOLKER B., WALKER, S. (2016). Package 'lme4', version 1.1- 12, linear mixed effects models using 'Eigen' and S4. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>.
- BOERSMA P., WEENINK D. (2019). *Praat: doing phonetics by computer*. Version 6.0.52. <http://www.fon.hum.ua.nl/praat>.
- CHO T. (2016). Prosodic boundary strengthening in the phonetics–Prosody interface. *Lang Ling Compass*, 10(3), 120-141.
- CORRETGE R. (2019). Praat Vocal Toolkit. <http://www.praatvocaltoolkit.com>.
- DANNENBRING G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Percept Motor Skill*, 51(3), 979-985.
- DENES P. (1955). Effect of duration on the perception of voicing. *JASA*, 27(4), 761-764.
- DUFOUR S., MEYNADIER Y. (2019). Temporary ambiguity in whispered word recognition: a semantic priming study. *J Cogn Psychol*, 31(2): 157-174. DOI: 10.1080/20445911.2019.1573243.
- GILICHINSKAYA Y., STRANGE W. (2011). Perception of final consonant “voicing” in whispered speech. *JASA*, 129, 2420.
- MEYNADIER Y., DUFOUR S. (2016). Accès lexical et reconnaissance du voisement en voix chuchotée. *Actes des XXXI^e JEP*, 19-27. Paris.
- MEYNADIER Y., DUFOUR S. (2018). Ambiguïté temporaire des obstruantes voisées en parole chuchotée. *Actes des XXXII^e JEP*, 125-133. Aix-en-Provence.
- MEYNADIER Y., GAYDINA Y. (2013). Aerodynamic and durational cues of phonological voicing in whisper. *Proceedings of the 14th Interspeech*, 335-339, Lyon.
- MITTERER H., CHO T., KIM, S. (2016). How does prosody influence speech categorization? *J Phon*, 54, 68–79.
- MUNRO M. J. (1990). Perception of “voicing” in whispered stops. *Phonetica*, 47, 173–181.
- STEFFMAN J. (2019). Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing. *JASA*, 154, EL560.
- TARTTER V. C. (1989). What's in a whisper? *JASA*, 86, 1678–1683.
- VERCHERAND G. (2010). *Production et perception de la parole chuchotée en français : analyse segmentale et prosodique*. Thèse de doctorat. Université Paris VII.