“Faire crisser sa voix” : la saillance de la voix craquée en français

Justin T. Jacobs[1]*, *Maria* Candea[1]

[1] CLESTHIA, Université Sorbonne Nouvelle, Paris, France

Résumé. A la suite d’une pré-enquête exploratoire où un·e participant·e a mobilisé la voix craquée pour répondre à une tâche de “dégenrer sa voix” a été formulée l’hypothèse d’une possible émergence de sens social indexé à cette qualité de voix, en français, à l’instar de ce qui a été décrit pour l’anglais américain [1-2]. Il a été décidé de tester expérimentalement, dans un premier temps, la saillance perceptuelle de la voix craquée en français. Deux groupes sont comparés : des anglophones nord-américain·es et des francophones métropolitain·es. Le groupe d’anglophones a servi de contrôle, dans la mesure où la voix craquée est attestée en production et sa perception sociale est documentée en anglais américain. En analysant les résultats, les francophones ont répondu correctement aussi souvent que les anglophones, ce qui peut indiquer que les francophones sont capables de distinguer et d'apparier des extraits prononcés en voix modales et des extraits prononcés en voix craquées au moins autant que les anglophones américain·es. Ces résultats encouragent à approfondir les recherches en français, pour déterminer le degré de saillance de cette qualité de voix, étudier quelle serait l’indexation sociale de la voix craquée et dans quelle mesure elle serait mobilisable par les francophones.

1 Introduction

Une qualité de voix est un paramètre laryngal du comportement de la glotte pendant l’articulation et qui influence le timbre de la voix. Nous modifions notre qualité de voix de manière parfois incontrôlée mais souvent de manière contrôlée, lorsqu’on chante ou lorsqu’on parle de manière autoritaire, de manière tendre à un jeune enfant ou encore à un animal de compagnie, en contexte de séduction, etc., en obéissant à des conventions sociales. Notre démarche en sociophonétique accorde une large place à l’intérêt pour la dimension matérielle et corporelle de la parole ainsi que pour les enjeux symboliques et sociaux qui permettent d’interpréter une partie de ces pratiques langagières corporelles. Bien que la phonétique reconnaisse un large spectre de paramètres possibles par rapport à la qualité de voix, certaines catégories sont souvent citées pour les décrire : voix sourde ou chuchotée (la glotte est la plus ouverte ; sans aucun voisement), soufflée (la glotte est moins ouverte, ce qui donne un air de murmure avec une expiration audible), modale (plus proche

* Corresponding author: justin.jacobs@sorbonne-nouvelle.fr

de nos voix spontanées, en parole), craquée (un comportement d'oscillation avec peu de tension de la glotte quand elle est plus fermée) et un état de fermeture totale [3]. La production de parole mobilisant la qualité de voix craquée (*creaky voice* ou *vocal fry*) en anglais américain est décrite depuis au moins une quinzaine d'années selon l'angle de l'indexicalité de classe sociale et de genre. Après avoir été considérée comme une pratique masculine elle a commencé à être décrite comme une pratique adoptée par les jeunes femmes urbaines, mobiles et en ascension sociale ou en position sociale forte "Young Urban-Oriented Upwardly Mobile American Women" [1] et ces descriptions ont largement été relayées dans les discours publics, médiatiques ou spontanés. On compte par centaines les vidéos parodiques ou tutoriels "stop vocal fry", par dizaines de milliers les commentaires sur internet en anglais, et certaines vidéos sur la plateforme Youtube ont atteint des millions de vues : *Vocal fry: An attack on women?*[1] (mise en ligne en 2015 par CBC News, 1,7 millions de vues) ou bien *Why you can't stand vocal fry*[2] (mise en ligne en 2023 par Dr Geoff Lindsey, 3,4 millions de vues). Pour la littérature scientifique consacrée à cette pratique vocale en anglais américain il y a même des méta-analyses qui comparent les résultats et les méthodes, comme par exemple [2]. En ce qui concerne le français, en particulier en France qui est notre pays cible, la situation semble très différente : nous trouvons surtout des tutoriels plutôt de niche pour la voix chantée qui expliquent la technique du *vocal fry* ou de la voix craquée, sans aucune évaluation sociale explicite. Une exception : deux chroniques bien diffusées du journaliste David Castello Lopes, en 2021 sur la radio *Europe 1* et en 2023 sur la radio *France Inter* où il explique ce qu'est le vocal fry qu'il traduit par "friture vocale" mais il ne parle que de l'anglais et de la perception sociale différente selon que cette qualité de voix est produite par des femmes ou par des hommes. Autrement dit, on peut parler, sur un plan macroscopique, d'un énorme écart de notoriété de cette qualité de voix entre l'anglais notamment nord-américain et le français de France. En outre, à un niveau plus microscopique, une étude pionnière de Pillot-Loiseau et al. [4] a montré que des locutrices nord-américaines venues en France pour améliorer leur français ont été sensibles aux différences de pratiques vocales entre les deux cultures et, par effet d'accommodation, elles ont éliminé la voix craquée de leurs productions en français L2 tout en la conservant dans leurs productions en anglais L1.

Dans le cadre de la thèse de doctorat du premier auteur de cet article qui porte sur les corrélats acoustiques de la présentation et de l'identification du genre en français, une expérience pilote a été menée [5] pour observer quels sont les paramètres acoustiques que les personnes mobilisent pour répondre à une question peu commune : parler avec une voix dégenrée ou moins genrée. Cette pré-enquête est présentée en section 3.1. Une des réponses obtenues, qui mobilisait la voix craquée associée de manière explicite à l'expression d'une identité de genre non-binaire, a tout particulièrement retenu notre attention.

Cette observation intéressante a provoqué beaucoup de questions. La voix craquée pourrait-elle devenir saillante en français, à l'instar de ce qui s'est passé en anglais ? Pourrait-on s'attendre à observer cette pratique vocale en parole spontanée, la voir acquérir du sens social, peut-être même en lien avec l'expression du genre ? S'agirait-il d'une observation isolée et idiosyncrasique de la personne qui a répondu à notre enquête pilote ou bien cette indexation de la voix craquée serait en train d'émerger en français ? En l'absence de discours massifs et faciles à trouver en français, nous nous sommes d'abord demandé si l'"oreille" des personnes socialisées dans la culture française métropolitaine était capable de percevoir la différence entre un énoncé en voix modale et un énoncé en voix craquée et si une simple expérience d'appariement d'extraits sonores brefs, similaires ou dissemblants,

[1] https://www.youtube.com/watch?v=UuAQsnAVoMw
[2] https://www.youtube.com/watch?v=Q0yL2GezneU

donnait des résultats différents ou similaires en anglais (par des gens socialisés aux Etats-Unis) et en français (par des gens socialisés en France). Avant de pouvoir espérer aller plus loin et tenter de déceler des associations de sens social lié à cette qualité de voix nous avons donc voulu savoir si elle était tout simplement suffisamment saillante pour être perçue. C'est pour cela que nous avons monté une expérimentation dont les fondements théoriques sont présentés en section 2, la mise en place en section 3.2 et les résultats en section 4. La discussion nous permettra d'imaginer les suites possibles de cette étude exploratoire sur ce qui pourrait constituer un trait émergent.

2 Cadre théorique de la saillance

Afin de discuter le concept de saillance, nous nous placerons dans le domaine de la phonétique-phonologie dans une démarche qui réunit approche empirique et théorisation. Deux conceptualisations de la saillance, en phonétique et en phonologie, ont nourri le paradigme expérimental utilisé dans cette expérience. Selon Barzilai, la saillance phonétique se définit par des considérations acoustiques et la saillance phonologique se définit par la perception de variables comme des éléments contrastifs de la langue étudiée [6]. Barzilai explique le concept de saillance en général comme la perceptibilité relative d'un segment ou d'un processus. La perceptibilité serait relative parce que ce n'est pas une question binaire de présence ou absence de saillance, mais c'est plutôt une question de "à quel point est telle ou telle variable saillante dans cette langue" ? Du point de vue acoustique, la saillance se réfère à la manière de percevoir un son : les sons les plus saillants sont plus facilement traités ou plus correctement perçus acoustiquement que les sons moins saillants. Cela dit, dans le cas où un effet de saillance n'est pas observé, l'absence d'effet ne montre pas nécessairement qu'un élément n'est pas saillant ; parfois cela est dû au paradigme expérimental choisi, ce qui encourage la pratique d'une mixité de méthodes expérimentales pour explorer la saillance phonétique d'une variable. Il est possible d'exploiter une approche expérimentale qui analyse les variables dans leurs contextes sociaux. Pour vraiment théoriser une saillance perceptuelle il est nécessaire de prendre en compte toute sa complexité et donc identifier la signification sociale de la mesure choisie, développer une méthodologie qui permet de saisir la variation individuelle et accepter que la saillance n'est pas toujours simplement binaire en termes de saillant / pas saillant.

La saillance d'un trait peut être évaluée comme étant linguistiquement pertinente dès lors qu'on peut l'observer chez un individu qui prend en compte plusieurs types de contexte tels que le contexte linguistique, le contexte social et le contexte situationnel [7]. Pour sélectionner ce qui est saillant (dans un extrait de parole hors contexte ou en interaction dans un contexte écologique) tout individu opère une évaluation fondée sur toute son expérience linguistique et sa mémoire à long-terme. Dans le cas de la voix craquée en anglais, on peut évaluer ce trait de qualité de voix comme saillant si un individu le repère de manière consistante dans sa propre parole ou dans celle d'autrui, à travers des contextes différents. La saillance est bien individuelle [8] parce qu'un élément peut être saillant pour une personne et pas pour une autre ; un élément peut être saillant pour deux personnes mais pour des raisons variées. Cette définition est applicable à la sociolinguistique où elle se superpose en grande partie avec ce qu'on appelle l'indexicalité. La saillance différencie le *indicator* (des variables d'indexation sociale qui ne sont pas remarquées par les communautés analysées) et le *marker* (des variables d'indexation sociale qui sont remarquées par leurs communautés) [9]. Elle coïncide également avec le marquage parce que les variables marquées sont plus saillantes que les variables non-marquées.

Par exemple, on pourrait dire que la variable voix modale / voix craquée chez les femmes états-uniennes est une variable marquée, qui fait désormais l'objet de discours

explicites d'évaluation sociale. En revanche, rien ne permet pour le moment de dire qu'elle serait marquée en français. Nous pouvons nous demander si elle est en train de le devenir ou s'il s'agit d'un signal faible, pour le moment au stade d'indice (*indicator*).

La saillance est un concept qui accepte la gradation de sa capacité à indexer des affiliations sociales et de genre [10]. Dans le cadre qui nous intéresse, on parle de saillance si les personnes qui écoutent un extrait de parole enregistrée remarquent un trait particulier et lui attribuent un sens social (d'une manière consciente et explicite ou pas). Cela arrive à chaque fois qu'un élément (un trait de prononciation en phonétique) est inhabituel et n'est pas fréquent, mais cela devient un marqueur seulement quand le groupe s'accorde sur sa fonction.

Le concept de la saillance en soi, qui est un concept cognitif, se trouve lié au concept de l'indexation sociale et constitue par ce biais un sujet souvent discuté en sociophonétique. C'est grâce à leur saillance (à la suite d'un processus qui aboutit à une saillance) que des variables sociophonétiques, qui ne sont autre choses que des pratiques corporelles, de phonation, variables, acquièrent la capacité de porter des sens sociologiques. Plus précisément, les variables linguistiques ont la capacité de faire référence aux identités sociales [11] ou à d'autres affiliations d'un individu dans un contexte de communication. Le concept d'indexation sociale s'appuie sur la théorie de l'exemplarité où l'indexation a besoin d'exposition répétée à un input qui devient reconnaissable pour réussir à associer une forme avec un sens [12]. Après une quantité suffisante d'exposition à une variable phonétique comme trait d'une identité en particulier, l'individu commence à établir une association entre les deux éléments, entre la présence d'un trait et une identité. Ensuite, après le passage du temps, un ensemble de traits linguistiques qui sont indexés systématiquement avec un groupe social pourra faire l'objet d'un processus de "catégorisation sociale". Ce cycle entier est connu sous le nom d'"enregisterment" [13]. Plus explicitement, la catégorisation sociale de la situation résultante pendant l'*enregisterment* se manifeste par l'association du stimulus à une catégorie construite dans une structure de gradient où chaque association se base sur les expériences vécues de la personne qui fait la catégorisation [14]. La perception de la catégorie à laquelle un locuteur ou une locutrice appartient dépend à la fois de son articulation et des expériences de la personne qui perçoit ; ces expériences autorisent des associations entre l'articulation perçue et une catégorie d'identité [15].

Pour pouvoir répondre à la question de la saillance de la voix craquée en français, il faudra d'abord reconnaitre le côté gradient du concept de saillance et se demander dans quelle mesure elle est saillante, pour qui et dans quelles situations. Dans un premier temps nous avons tenté de répondre à la toute première partie de ce questionnement, qui nous permettra de formuler des hypothèses sur la possibilité d'un cycle de *enregisterment* de la voix craquée en français qui serait dans sa phase de début.

3 Démarche expérimentale

La récolte de nos données a suivi deux étapes : la première étape était une pré-enquête exploratoire, évoquée en introduction, qui nous a encouragés à formuler des hypothèses sur la voix craquée et la seconde étape était l'expérience elle-même pour tester nos hypothèses. L'expérience menée dans le cadre de l'interrogation de la saillance de la voix craquée en français s'est déroulée en deux versions : une version française (expérimentale) menée en France à Paris et une version anglaise (contrôle) menée aux Etats-Unis à Pittsburgh. La version contrôle vise à faire une comparaison avec la version expérimentale. Comme la voix craquée est bien documentée et discutée en anglais américain, la comparaison a semblé appropriée pour comprendre les résultats de l'expérience en français. Ce premier

volet expérimental doit permettre la poursuite de la comparaison par des tâches différentes, complémentaires à celle qui sera décrite dans cet article.

3.1 Pré-enquête exploratoire

Au début de la récolte de données de notre projet de recherche qui cherche à interroger les dynamiques en cours au sujet des corrélats acoustiques de l'identité de genre en français, on s'est posé les questions suivantes: "Et si on rejetait la binarité du marquage du genre à travers nos pratiques vocales ? Dans la mesure où on apprend, culturellement, à performer son genre y compris par ses pratiques vocales, peut-on envisager d'apprendre des pratiques différentes, inverses? A quel point peut-on dégenrer la voix humaine" ? Ces questions font déjà l'objet de conseils empiriques en anglais à l'usage des personnes trans et non-binaires dans une publication de 2017 par des orthophonistes [16] qui a connu une seconde édition en 2020.

Pour réfléchir à cette problématique, on a construit une pré-enquête exploratoire [5]. Le premier auteur a recruté des bénévoles d'un cours de master enseigné par la deuxième auteure de l'article. Ces bénévoles sont devenus des chercheur·euses pour participer à la construction de données observables.

Les chercheur·euses ont recruté des ami·es et des membres de leurs familles pour participer à une brève recherche. A cette étape, le genre du participant / de la participante n'était pas considéré comme pertinent et n'a pas été noté. L'âge très précis du participant / de la participante n'était non plus considéré comme pertinent mais la tranche d'âge a été notée car nous avions l'hypothèse de l'existence de différences générationnelles éventuelles. Trois tranches d'âge ont été considérés : 18-30, 31-50 et 51+. Pendant la présentation de la pré-enquête aux masterant·es pour solliciter leur coopération, le premier auteur a présenté un court texte d'après Arnold [17] et tous·tes présent·es ont collaboré en construisant un texte plus adapté aux fins de la pré-enquête. Le texte résultant a été le suivant : "Salut, ça va ? Ça te dit de boire un café cet après-midi ? Je connais un endroit sympa pas loin. Tiens-moi au courant. A plus !" La tâche pour les chercheur·euses du master était d'enregistrer ce texte produit à l'oral selon plusieurs consignes, avec n'importe quel instrument disponible, car le but de l'analyse de ces données n'était pas des analyses de détails fins, mais une compréhension plus globale des stratégies employées. Les chercheur·euses ont invité leurs participant·es de lire à haute voix le bref texte seize fois : quatre fois avec leurs voix spontanées, quatre fois avec une voix plus masculine, quatre fois avec une voix plus féminine et quatre fois avec une voix dégenrée ou moins genrée que leur voix spontanée. Si les trois premières tâches étaient faciles à comprendre (parler normalement, masculiniser ou féminiser sa voix), la dernière était plus surprenante, moins commune. Les instructions ne précisaient pas exactement ce que "dégenrée" veut dire, pour permettre que les participant·es développent leurs propres idées. A la fin, chaque participant·e se voyait poser la question : "C'est quoi pour toi, une voix dégenrée ?" pour essayer de comprendre un peu plus les enregistrements de chaque participant·e.

Les chercheur·euses ont enregistré un total de 62 personnes. Le but de cette pré-enquête était d'observer globalement les traits phonétiques mobilisés par les francophones pour dégenrer leurs voix. En résumé, il a été observé les cas de figure suivants: pas de stratégie (leurs voix spontanées ont été produites à nouveau pour répondre à la consigne de produire une voix dégenrée), ou bien des stratégies de modifications du débit, de l'intensité, de la labialisation, de la hauteur, de la qualité de voix et de l'intonation, mais aussi l'isolation de syllabes et d'autres stratégies utilisés auparavant pendant la production d'autres voix performées pendant l'expérience.

Parmi tous les enregistrements de la voix dégenrée, un·e participant·e, MTP11 (son numéro de participant anonymisé), a adopté une stratégie qui nous a semblé

particulièrement intéressante en raison de sa justification explicite. MTP11 a choisi de pratiquer une qualité de voix craquée pour lire la phrase demandée avec une voix dégenrée. Lors de la phase de discussion libre de la tâche de lecture en voix dégenrée, à la question "pourquoi tu as prononcé comme ça?" cette personne a répondu : "J'ai essayé de faire crisser ma voix". Pendant la discussion pour tenter de comprendre la volonté de faire "crisser" sa voix, cette personne a rapporté s'être laissé inspirer par des gens non-binaires de sa connaissance, car ces gens non-binaires, selon elle, ont l'habitude tellement remarquable d'adopter une voix craquée que notre participant·e a commencé à associer ce trait phonétique à cette identité.

Cette pré-enquête exploratoire montre que les francophones de France emploient une variété de stratégies lorsqu'on leur demande expérimentalement de dégenrer leurs voix ; de la modification de segments et modification suprasegmentale à la réutilisation de leurs voix spontanées. Il est possible que leur voix spontanée soit déjà considérée comme dégenrée par certaines personnes sollicitées, et que certaines personnes aient adopté des stratégies articulatoires pour construire une expression complexe de genre. La voix craquée a émergé comme étant un trait saillant explicite pour au moins un·e participant·e qui se réfère à un partage de référence avec des personnes non binaires et donc qui relie ce trait à une identité. C'est ce qui nous a donné l'idée de tester s'il s'agit d' un trait mobilisable en français. Est-il saillant perceptivement, au moins de façon basique ? Si c'est le cas, à quel point ?

3.2 Tâche de saillance

Pour mener une investigation de la saillance perceptive de la voix craquée, on a construit un protocole qui montre la capacité à traiter et à catégoriser la voix craquée en tant qu'auditeur ou auditrice. On a choisi un paradigme ABX, où les participant·es écoutent trois fichiers audio et leur tâche est de déterminer si le troisième fichier audio est plus similaire au premier ou au deuxième. C'est un paradigme très simple mais efficace[3].

Chaque essai se déroule de la façon suivante : d'abord, les participant·es écoutent deux enregistrements, A et B. Ensuite, les participant·es écoutent un troisième enregistrement, X. Après le troisième, il leur est demandé de déterminer si le troisième enregistrement (X) est plus similaire au premier (A) ou au deuxième (B). Certains extraits sont produits en voix modale et certains en voix craquée, mais cela ne leur est pas dit. La tâche vise à tester si les personnes qui écoutent le troisième enregistrement, qui peut être en voix modale ou en voix craquée, sont capables de l'apparier correctement avec l'extrait similaire, produit par une autre voix (modale ou craquée). En cas de réussite des appariements, cela conforte l'hypothèse sur la capacité à percevoir la voix craquée et éventuellement (par une opération d'évaluation) lui associer une indexation sociale. En cas d'échec, cela plaide pour l'hypothèse d'une impossibilité, à ce stade, à sélectionner le trait "creaky" comme saillant pour pouvoir catégoriser la voix craquée et donc une impossibilité à lui associer quelque sens que ce soit en français de France.

Lors du déroulement de la tâche, au début de chaque phase, l'écran affiche "Cliquez pour commencer". Chaque essai contient un écran identique : la progression *n*/30 où le *n* représente le numéro de l'essai actuel, les consignes "Déterminez si le troisième enregistrement est plus similaire au premier ou au deuxième", un bouton "premier", un bouton "deuxième", un bouton "Ecoutez à nouveau", un bouton "Continuez" qui apparait après la sélection du choix et un écran final qui affiche "La tâche est terminée". Le silence pré-stimulus est de 0,1 secondes, le silence inter-stimulus est de 0,3 secondes, avec un total de 30 stimuli. Le nombre de réécoutes possibles de chaque extrait n'était pas limité.. Une

[3] Le chercheur a programmé le protocole sur un script dans le logiciel Praat [18] basé sur le script experimentmfc.praat, proposé par le logiciel qui aide à construire des protocoles simples.

fois terminés, le chercheur télécharge les résultats dans un tableur en format .csv et les participant·es continuent par une autre tâche de production vocale. La version anglaise est techniquement identique à part le texte des boutons qui est en anglais : "Click here to start," "Determine if the third recording is more similar to the first or second one," "First," "Second," "Listen again," "Continue," "The task is finished."

Les stimuli sont composés d'enregistrements de locutrices qui ont lu des extraits de contes de fées. Chaque extrait dure de deux à trois secondes ; assez long pour que les participant·es puissent entendre une qualité de voix constante pour comprendre que ce n'est pas isolé sur une syllabe, mais que c'est la qualité de voix utilisée par la locutrice. Cette durée est également assez courte pour que ce soit possible de tout retenir avant d'oublier, donc dans les limites de la mémoire à court-terme. Les quatre locutrices (deux françaises, deux américaines) sont des femmes cisgenres, choisies parce que la voix craquée sera plus saillante chez les femmes à cause du stéréotype que les femmes ont des voix plus aigües et les hommes ont des voix plus basses. Si la voix craquée se situe vers les fréquences plus basses, la voix craquée utilisée par un homme sera moins inattendue. La majorité des participant·es dans la pré-enquête résumée en 3.1 ont élevé la hauteur de leurs voix dans la persona "féminine" et la majorité ont abaissé la hauteur de leurs voix dans la persona "masculine". Comme les locutrices choisies sont des femmes, cela leur laisse plus de plage pour baisser leur fréquence de la voix. Pour la version anglaise, la voix craquée est facilement perçue et reproduite, donc deux femmes naïves ont été sollicitées pour produire les stimuli. Pour la version française, deux professeures qui sont des expertes en phonétique, qui connaissent les qualités de voix et qui connaissent la voix craquée, ont été sollicitées pour produire les stimuli.

Les textes lus par les locutrices sont des citations bien connues de contes de fées bien connus. Le premier auteur a choisi 15 citations et les a envoyées à 10 personnes dans leurs L1 français ou anglais. Il leur a posé la question : connais-tu cette citation et le conte de fées d'où elle vient ? Il n'a sélectionné que les citations qui ont reçu 9 / 10 de réponses "oui, je connais la citation et je sais de quel conte de fées elle vient". Une fois les 10 citations choisies, les locutrices ont enregistré chaque citation cinq fois : une fois dans une voix modale (leurs voix spontanées), une fois dans une voix craquée, une fois dans une voix soufflée, une fois dans une voix de fausset et une fois dans une voix grave. Le choix de qualité de voix pour chaque extrait était un choix assez aléatoire, sauf les extraits expérimentaux. Nous avons choisi deux extraits pour la voix craquée parce que la citation est d'un vilain (dans la version française : "Tire la bobinette et la chevillette cherra" énoncé par le loup déguisé en grand-mère dans *Le petit chaperon rouge* et "Miroir, mon beau miroir, dis-moi qui est la plus belle ?" énoncé par la méchante reine dans *Blanche-Neige*. Dans la version anglaise : "Fee-fi-fo-fum, I smell the blood of an Englishman" énoncé par l'ogre dans *Jack et le haricot magique* et "Mirror, mirror, on the wall, who's the fairest of them all?" énoncé par la méchante reine dans *Blanche-Neige*). Un extrait a été choisi pour voir ce qui arrive quand la voix grave est posée juste avant ou après la voix craquée car ces deux qualités de voix pourraient être confondues (dans la version française : "C'est à Monsieur le Marquis de Carabas" énoncé par les moissonneurs dans *Le chat botté*. Dans la version anglaise : "It was a dark and stormy night" un cliché de n'importe quel conte de fées).

Nous avons effectué un petit pré-test avec 10 enregistrements dont 3 sont en voix craquée auprès de trois spécialistes en sciences du langage de son réseau professionnel. La tâche était d'indiquer quels enregistrements sont en voix craquée. Chaque spécialiste a reçu une version d'enregistrements dans sa langue de première socialisation (L1). Leur tâche était formulée ainsi : si ce que vous entendez est en voix craquée, indiquez 1 et si ce que vous entendez n'est pas en voix craquée, indiquez 0. Chaque enregistrement en voix craquée a reçu des jugements corrects 100% du temps.

En préparation de l'emploi des enregistrements dans des stimuli, le premier auteur a segmenté et prétraité les enregistrements dans Praat [18] en utilisant un script du professeur Jalal Al-Tamimi pour ajuster l'intensité, rééchantillonner, filtrer et réduire le bruit. Une fois organisés en stimuli, deux stimuli ont été mis de côté pour servir à des essais dans la phase d'entrainement. Par rapport aux stimuli qui restent, chaque stimulus est composé de deux qualités de voix. De nombreux stimuli sont des distracteurs produits en voix soufflée, en voix de fausset, etc. afin d'éviter de focaliser l'attention des personnes participantes sur l'objet de l'étude. Seulement les stimuli choisis pour la voix craquée sont les stimuli expérimentaux. Le protocole final comprend 18 stimuli (liste des phrases lues en annexe) constitués chacun de trois enregistrements : deux enregistrements de différentes qualités de voix, les deux enregistrés par la même personne, A et B, et un troisième enregistrement X d'une deuxième personne qui utilise une des deux qualités de voix utilisées dans le premier (A) ou le deuxième (B) enregistrement.

Tableau. 1. Exemple d'un stimulus.

Stimulus *n*	Enregistrement *n*A	Enregistrement *n*B	Enregistrement *n*X

Dans le tableau 1, le stimulus *n* est composé de trois enregistrements. Le premier (*n*A) enregistré par locutrice 1 dans une voix modale, le deuxième (*n*B) enregistré par locutrice 1 dans une voix de fausset et le troisième (*n*X) enregistré par locutrice 2 dans une voix de fausset. C'est donc le cas pour l'ensemble du test : les enregistrements A et B par la même personne, l'enregistrement X de chaque stimulus vient d'une autre personne. Chaque citation apparait quatre fois dans le protocole, dans un carré latin.

Tableau 2. Exemple du carré latin où L fait référence à une locutrice et Q fait référence à une qualité de voix.

	A	**B**	**X**
Stimulus	**L1, Q1**	*L1, Q2*	L2, Q1
Stimulus	*L1, Q2*	**L1, Q1**	L2, Q1
Stimulus	**L1, Q1**	*L1, Q2*	L2, Q2
Stimulus	*L1, Q2*	**L1, Q1**	L2, Q2

Dans le tableau 2, quatre stimuli sont présentés pour montrer le carré latin qui organise la distribution de chaque citation. De cette manière, chaque citation sera entendue dans chaque position (A, B ou X) pour éviter des biais de positionnement. Les stimuli sont présentés dans un ordre aléatoire où un stimulus donné ne peut pas suivre un stimulus de la même citation.

Les participant·es écoutent la citation lue dans deux qualités de voix différentes et ensuite la citation lue par une locutrice différente qui utilise une des deux qualités de voix utilisées par la première locutrice. Les participant·es doivent ainsi choisir lequel des deux premiers enregistrements est le plus similaire au troisième.

Au début de chaque passation, les participant·es signent un formulaire de consentement et ensuite le chercheur explique les consignes. La tâche se déroule en deux phases : la première phase et la phase d'entraînement pendant laquelle deux essais sont présentés aux participant·es. Le chercheur reste à côté des participant·es pour répondre aux questions, intervenir si quelque chose ne fonctionne pas dans le logiciel ou l'ordinateur, etc.

Le chercheur a présenté la tâche aux participant·es sur un laptop avec des écouteurs.

Durant la phase d'entrainement, le logiciel ne sauvegarde pas les choix des participant·es. Ensuite de la phase d'entrainement commence la phase de la vraie tâche pendant laquelle le chercheur laisse les participant·es tous·tes seul·es et tous leurs choix sont sauvegardés par le logiciel.

Pour cette tâche de saillance, la version anglaise de l'expérience s'est passée à l'University of Pittsburgh aux Etats-Unis et la version française de l'expérience s'est passée à l'Université Sorbonne Nouvelle en France, dans le cadre de la thèse de doctorat du premier auteur. La grande majorité des participant·es sont des étudiant·es à chaque université qui se situent dans une tranche d'âge 18-40 avec très peu de participants de l'âge de 40 ans ou plus. L'appel à participation a été diffusé dans de nombreux cours de linguistique de niveau débutant et au-delà, vers d'autres départements. Au final, la moitié des participant·es ne suivent pas un parcours en sciences du langage et donc ont une connaissance limitée de la linguistique. Les participant·es suivent des parcours de lettres, théâtre ou cinéma à tous les niveaux de la licence, de L1 à L3 en France et de *freshman* (L1) à *senior* (L3+1) aux Etats-Unis. Le total final des effectifs réunis comptait 45 participant·es américain·es et 60 participant·es français·es. Les participant·es américain·es ont été rémunéré·es avec un bon cadeau de $15. Les participant·es français·es d'autres cours ont été rémunéré·es avec un bon cadeau de 10€, sauf . Les participant·es français·es des cours enseignés par le chercheur qui ont été rémunéré·es avec deux points supplémentaires sur leurs notes finales dans le cours. Parmi les participant·es français·se des cours enseignés par le chercheur, plusieurs participant·es avaient un profil qui ne correspondait pas aux critères de sélection (avoir été socialisé depuis son enfance en France métropolitaine) et nous avons dû exclure leurs données. Après l'exclusion, le total final retenu a été de 45 participant·es du groupe américain et 52 participant·es du groupe français. Lors de la récolte de données, ni le genre ni l'âge précis de chaque participant·e n'ont été demandés car il n'y avait pas d'hypothèse précise liée à ces catégorisations.

4 Résultats

Le paradigme classique ABX correspond à un protocole où il s'agit de discriminer la présence ou l'absence d'un élément (souvent un segment, ici il s'agissait de la présence ou absence d'une qualité de voix). La version du paradigme exploité dans cette expérience s'élabore autour d'un protocole 2AFC *two alternative forced-choice* "choix forcé à deux alternatives". Au lieu de discriminer l'absence ou présence d'un élément, ce qui est demandé c'est de catégoriser par la similarité sans critère explicite un élément par rapport à deux autres ; la tâche de catégorisation est implicite et opérée par les chercheurs dans le protocole. La mesure souvent consultée pour expliquer les résultats de données ABX s'appelle *d-prime* (d'), une mesure qui considère les écarts types et en calcule la différence entre les moyennes de distributions pour exprimer un degré de différence entre les deux ensembles de données [19]. Dans la version ABX classique il faut trier les données en quatre catégories : *hits* (la détection du signal quand le signal est présent), *misses* (la manque de détection du signal quand le signal est présent), *false alarms* (la détection du signal quand le signal est absent) et *correct rejections* (la manque de détection du signal quand le signal est absent). Dans notre version 2AFC il n'y a pas de possibilité de triage dans cette manière, alors le calcul du z-score, qui est normalement fait avec les chiffres de chaque catégorie, se fait avec la proportion correcte multiplié par la racine 2 (qui prend en compte l'aspect d'avoir deux choix possibles) [20]. Ensuite, avec les valeurs *d'* il est nécessaire de comprendre si les résultats obtenus de la tâche sont dus au hasard ou si la différence observée est vraiment significative. Pour faire ce type de comparaison, on utilise un test-t. Un test-t appliqué à un échantillon analyse toutes les données ensemble pour

calculer le degré de certitude qui permet de rejeter (ou pas) l'hypothèse nulle (dans ce cas, l'hypothèse nulle serait le fait que les gens auraient répondu au hasard lors de la tâche d'apparier l'extrait X avec l'extrait A ou avec l'extrait B). Dans le cas de notre méthode, un test-t pour échantillons appariés nous permet de comparer les deux conditions (version anglaise américaine et version française métropolitaine) pour observer si les conditions sont différentes et à quelle mesure elles sont différentes.

Tout d'abord, le calcul du *d'* de tous·tes les participant·es suit l'équation de Sargiotis [20].

$$d' = \sqrt{2} \cdot \Phi^{-1}(P_C) \quad (1)$$

Dans l'équation 1, la valeur d-prime est calculé de la multiplication de racine 2 par le z-score Φ^{-1} et la proportion correcte (P_C). Il est simple de faire ce calcul dans Python avec un script partagé par Sargiotis [20].

Avant de commencer l'analyse des stimuli expérimentaux, il avait semblé que la version anglaise n'avait pas réussi. La moitié des participant·es ont raté la tâche, ce qui semblait curieux parce que la condition anglaise était censée être la condition de contrôle. Un réexamen post hoc des données a révélé que 37 des 41 participant·es de la tâche en anglais ont raté deux stimuli [j5m;j13c;m6c] et [j13c;j5m;m6c], toujours les mêmes. Il s'agit de deux présentations du stimulus où la locutrice J a lu la citation 5 en voix modale et la citation 13 en voix craquée, et la locutrice M a lu la citation 6 en voix craquée. En réécoutant les stimuli, nous avons trouvé que l'enregistrement donné à évaluer en position X de chaque stimulus, dit craqué, n'était pas lu en voix craquée pendant toute la durée de l'enregistrement, comme c'était le cas pour l'enregistrement de la condition "voix craquée" de la paire AB et BA. Il apparait vraisemblable que les participant·es ont jugé l'enregistrement X produit partiellement en voix modale et partiellement en voix craquée comme plus similaire à l'enregistrement en voix modale parce que la présence de la voix craquée en deuxième partie d'enregistrement a été moins saillante. Par conséquent, nous avons décidé d'écarter ces stimuli comme ayant été trop ambigus. L'analyse statistique suivante a été fait dans JASP [21] et s'est portée sur les stimuli restants.

Tableau 3. Test-t à un échantillon des données expérimentales des deux groupes ensemble.

Test-t à un échantillon			
	t	df	p
d'	23.86	96	< .001

Dans le tableau 3 sont présentés les résultats du test-t à un échantillon des données expérimentales. Après avoir trouvé le d' de chaque participant des deux groupes combinés, il est possible de calculer les valeurs et la valeur-p résultante est en dessous de .001, ce qui indique qu'il est fortement possible de rejeter l'hypothèse nulle qui montre que les réponses des participant·es ne sont pas aléatoires.

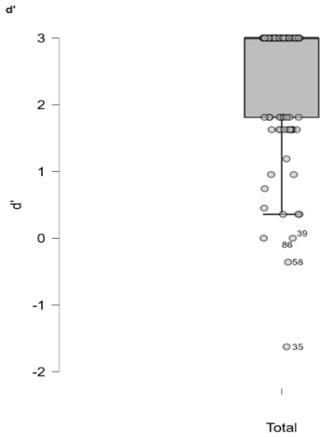

Fig. 1. Boite à moustaches des données expérimentales des deux groupes ensemble où l'axe Y décrit les valeurs d'.

Les deux premières propriétés qu'il faut vérifier : des variables dépendantes continues et deux conditions catégoriques. Or, la figure 1, laisse apparaitre un regroupement très dense de toutes les réponses recueillies à l'exception de quatre points aberrants visibles dans le diagramme. Le calcul lancé avec ces points statistiquement aberrants laissait apparaitre (tableau 4 ci-dessous), une asymétrie entre -2 et +2, mais aussi une acuité de 2.528, ce qui signifie une distribution anormale.

Tableau 4. (gauche) Statistiques descriptives des données expérimentales des deux groupes ensemble.
Tableau 5. (droite) Statistiques descriptives des données expérimentales corrigées des deux groupes ensemble.

	d'
Moyen	2.386
Asymétrie	-1.654
Acuité	2.528

	d'
Moyen	2.509
Asymétrie	-1.318
Acuité	0.555

Selon les pratiques habituelles pour le traitement statistique de ce type de réponses, dans la mesure où des réponses aberrantes, même en quantité très limitée, influencent la normalité de distribution, elles ont été écartées pour focaliser l'analyse sur la masse compacte des réponses groupées, qui se comportaient selon une distribution tout à fait normale [22]. Un relancement des statistiques descriptives pour vérifier les propriétés des données montre des valeurs corrigées par rapport au tableau 4 et acceptables (tableau 5).

Sur la base de cette distribution concentrée, rendue normale, où les valeurs d'asymétrie et d'acuité sont entre -2 et +2, il est possible de séparer les deux groupes pour faire le test-t pour les comparer.

Tableau 6. Test-t pour échantillons appariés des données expérimentales corrigées.

Test-t pour échantillons appariés

Mesure 1	Mesure 2	t	df	p	*d* de Cohen	TE *d* de Cohen
d' (anglais)	d' (français)	-0.305	44	.972	-0.005	0.202

Le tableau 6 présente une valeur-p haute avec un taille d'effet du *d* de Cohen bas. Visualisés, la différence n'est pas très apparente entre les deux (figure 2) ce qui indique que les participants du groupe américain ont donné des réponses un peu moins dispersées que le groupe français, mais avec une moyenne très proche.

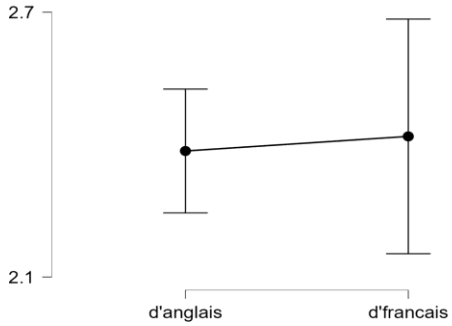

Fig. 2. Diagramme qui montre la relation entre les conditions où l'axe X décrit le groupement par langue et l'axe Y décrit les valeurs d'.

Dans le tableau 7 ci-dessous, il n'y a pas de différence frappante pour les participant·es anglais·es (M = 2.347, ET = 1.032) en comparaison avec les participant·es français·es (M = 2.419, ET = 0.951).

Tableau 7. Statistiques descriptives des données expérimentales corrigées.

	N	Moyen	ET	TE	Coefficient de variation
d' (anglais)	45	2.347	1.032	0.154	0.440
d' (français)	52	2.419	0.951	0.132	0.393

Ces résultats nous permettent de dire que, en ce qui concerne les stimuli expérimentaux focalisés sur la différence entre qualité de voix craquée et qualité de voix modale, le groupe testé en France a répondu de manière aussi performante que le groupe testé aux Etats-Unis.

5 Discussion

Le test-t à un échantillon nous permet de déterminer si la performance dans l'expérience est vraiment observable ou si c'est dû au hasard (hypothèse nulle) ; on mesure la probabilité que les réponses recueillies aient été aléatoires. Comme la valeur-p du test-t à un échantillon est inférieure à 0.001, nous pouvons rejeter l'hypothèse nulle avec une bonne certitude. Par contre, la valeur-p du test-t pour échantillons appariés est 0.972, ce qui est une valeur très haute avec un taille d'effet du *d* de Cohen également bas. C'est un résultat attendu, si les deux groupes discriminent pareillement les stimuli expérimentaux, il sera possible de dire que la voix craquée n'est pas imperceptible ou inconséquente en français. Elle est au moins aussi perceptible pour les francophones que pour les anglophones. Les

résultats globaux ont montré une meilleure performance des anglophones dans la tâche d'appariement ABX pour les distracteurs, ce que nous ne pouvons pas expliquer et qui ne rentre pas dans les objectifs de cette recherche. En raison de la différence des langues et des cultures, les choix des distracteurs ne pouvaient pas être les mêmes et il est possible qu'il y ait eu des effets culturels imprévus. En somme, la tâche a réussi à montrer que les francophones sont sensibles à la voix craquée au moins autant que les anglophones.

Pour aller un peu plus loin et préparer les étapes suivantes de cette recherche, le premier auteur a mené une nouvelle pré-enquête exploratoire en posant une question ouverte dans un cours de linguistique à des étudiant·es en première année, donc en tout début de leur cursus universitaire. L'enregistrement produit par MTP11 de la pré-enquête exploratoire de la partie 3.1, qui parle en voix craquée, a été diffusé en classe et il a été demandé de répondre sur un formulaire par les premiers adjectifs qui venaient à l'esprit, associés à l'audio écouté. L'enregistrement a été écouté trois fois et les réponses collectées visaient à explorer des discours épilinguistiques après écoute d'une audio produit en voix craquée sans aucune focalisation de l'attention sur la qualité de voix. Les 53 réponses obtenues de 18 personnes interrogées ont été codées selon la polarité (positive, négative, neutre) et le thème (description objective, jugement subjectif). Par exemple, <crispant> est codé avec [négatif, jugement] codage doublement vérifié par un locuteur de français L1. Toutes les catégories du codage ont reçu plus ou moins une vingtaine de réponses, sauf la catégorie "jugement", qui a reçu 39 des 53 réponses. Parmi les jugements et descriptions variées, nous avons observé des descriptions de pathologie (fatigué, malade, anxieux) et des jugements de plaisanterie (drôle, amical, gentil). Ces résultats incitent à approfondir la piste choisie car ils constituent autant d'indices concordants en faveur de l'hypothèse selon laquelle la voix craquée est déjà porteuse de sens en français, mais des sens différents que ceux attestés en anglais et pas encore stabilisés.

6 Conclusion

La question de la saillance de la voix craquée en français est intéressante parce qu'elle n'est pas attestée de manière courante pour le moment [4] sauf dans des marqueurs vocaliques d'hésitations [23]. Par contre, la voix craquée est bien documentée en anglais [24-26], surtout dans l'anglais américain [1-2, 27-29]. Nous nous sommes donc demandé si les francophones peuvent prêter attention à la voix craquée autant que les anglophones américain·es et si cela pourrait constituer un trait en passe de devenir saillant et mobilisable pour acquérir un sens social.

Un paradigme qui mesure la perception tel que le paradigme ABX 2AFC est une méthodologie expérimentale efficace pour révéler le côté perceptuel d'une problématique qui cherche à interroger la saillance, mais pour vraiment lier la phonétique et la sociolinguistique (autrement dit, la sociophonétique), il faut une approche mixte avec des mesures explicites et implicites, des mesures quantitatives et qualitatives. La limite de cette première approche est de ne pas donner assez d'information sur les motivations sociales qui sous-tendent les perceptions sociales complexes. Un élément devient plus saillant quand il est perçu plus explicitement et quand il commence à porter des sens sociaux par une stratégie d'indexation, mais si on ne pouvait montrer aucune saillance on ne pourrait pas espérer avoir de résultats intéressants par la suite. La puissance du paradigme n'éclaire donc que la moitié de notre questionnement : cela incite à poursuivre les enquêtes, à adopter une approche mixte pour combler les angles morts et pour mieux capter les pratiques vocales culturellement marquées.

Cette expérience ne permet pas de décider définitivement, mais au moins, nous pouvons faire l'hypothèse que la voix craquée est mobilisable en français, en France, et ne constitue pas un trait inaudible ou imperceptible. Cette expérience forme la base d'une interrogation

d'indexations sociophonétiques en mobilisant la voix craquée. Nous défendons l'hypothèse que les francophones font attention à la voix craquée (tâche de perception en 3.2), que c'est mobilisable (pré-enquête exploratoire en 3.1) et que c'est porteuse de sens (la petite enquête informelle rapportée ci-dessus). Les résultats aboutissent à de nouvelles questions sur les discours épilinguistiques qui concernent les limites de l'indexation sociophonétique éventuelle. Pour le moment, les adjectifs produits par les étudiant·es pendant la petite enquête de réaction à un extrait produit en voix craquée sont dispersés : on voit autant d'adjectifs positifs que d'adjectifs neutres et négatifs. Si la voix craquée a la capacité de communiquer des informations sociolinguistiques ou stylistiques nous ne pouvons pas, pour le moment, savoir si on peut émettre l'hypothèse d'une imitation en français d'un phénomène états-unien, ou bien l'hypothèse d'une indexation d'identité non-binaire qui émergerait ou d'un autre développement local qui servira à indexer d'autres informations. Le terme populaire "vocal fry" existe en anglais et il est présent culturellement pour désigner la mobilisation de la voix craquée ; une recherche très courte de la traduction "friture vocale" sur un moteur de recherche fournit déjà des résultats en français, mais il s'agit plutôt d'articles qui expliquent un phénomène américain ou de tutoriels pour la technique de chant en voix craquée. Il est possible que ce soit un paramètre dont le sens social va s'enrichir dans les années à venir, et qu'on assiste à un changement en cours et à une évolution de la saillance, c'est la raison pour laquelle nous proposons que ce paramètre fasse désormais objet d'intérêt en sociophonétique du français.

Références

1. I.P. Yuasa, Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? American Speech. **85**, 3 (2010). https://doi.org/10.1215/00031283-2010-018
2. P. Meier, Creaky voice in American English: how are American women who use creaky voice perceived? A literature review. **9** (2023). https://doi.org/10.7146/lev92023136544
3. M. Gordon, P. Ladefoged, Phonation types: a cross-linguistic overview. J. Phonetics. **29** (2001). https://doi:10.006/jpho.2001.0147
4. C. Pillot-Loiseau, C. Horgues, S. Scheuer, T. Kamiyama, The evolution of creaky voice use in read speech by native-French and native-English speakers in tandem: a pilot study. Anglophonia. 27 (2019). https://doi.org/10.4000/anglophonia.2005
5. J.T. Jacobs, M. Candea, Jusqu'où peut-on performer le genre avec notre voix ? Poursuivre les controverses 'anatomie versus socio-articulatoire'. 6ème Congrès du Réseau Francophone de Sociolinguistique, Louvain-la-Neuve, Belgique, 28 mai - 1 juin (2024). https://hal.science/hal-04967879v1
6. M.L. Barzilai, Phonetic and phonological salience effects in different speech processing tasks, dans Proceedings of the 2020 Annual Meeting on Phonology, Santa Cruz, Californie, Etats-Unis, 18-20 septembre (2020). https://doi.org/10.3765/amp.v9i0.4898
7. H.-J. Schmid, F. Günther, Toward a unified socio-cognitive framework for salience in language. Frontiers in Psych. **7** (2016). https://doi.org/10.3389/fpsyg.2016.01110
8. V. Boswijk, M. Coler, What is salience? Open Ling. **6**, 1 (2020). https://doi.org/10.1515/opli-2020-0042
9. W. Labov, Transmission and diffusion. Lang. **83**, 2 (2007). https://doi.org/10.1353/lan.2007.0082
10. C. Llamas, D. Watt, A.E. MacFarlane, Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. Frontiers in Psych. **7** (2016). https://doi.org/10.3389/fpsyg.2016.01163

11. E. Ochs, Indexing gender. Rethinking Context: Language as an Interactive Phenomenon, (Cambridge University Press, Cambridge, 1992)
12. P. Foulkes, G. Docherty, The social life of phonetics and phonology. J. Phonetics. **34**, 4 (2006). https://doi.org/10.1016/j.wocn.2005.08.002
13. A. Agha, Voicing, footing, enregisterment. Ling. Anthro. **15**, 1 (2005). https://doi.org/10.1525/jlin.2005.15.1.38
14. E. Rosch, Categorization. Handbook of Pragmatics. **4** (1998). https://doi.org/10.1075/hop.4.cat2
15. A.W. Coetzee, P.S. Beddor, W. Styler, S. Tobin, I. Bekker, D. Wissing, Producing and perceiving socially indexed coarticulation in Afrikaans. Lab. Phonology. **13**, 1 (2022). https://doi.org/10.16995/labphon.6450
16. M. Mills, G. Stoneham, The Voice Book for Trans and Non-Binary People: A Practical Guide to Creating and Sustaining Authentic Voice and Communication, (Jessica Kingsley Publishers, Philadelphia, 2017).
17. A. Arnold, L'abaissement de la fréquence fondamentale comme pratique de séduction, dans Actes des XXXIIe Journées d'Etudes sur la Parole, Aix-en-Provence, France, 4-8 juin (2018). https://doi.org/10.21437/JEP.2018-49
18. P. Boersma, D. Weenink, Praat: doing phonetics by computer. Glot International. **5**, 9/10 (2025). https://praat.org
19. R.E. Greenway, ABX discrimination task. Discrimination Testing in Sensory Science: A Practical Handbook, (Woodhead Publishing, Sawston, 2017). https://doi.org/10.1016/B978-0-08-101009-9.00013-7
20. D. Sargiotis, How to compute d-prime calculation for a 2-alternative force choice data?, Researchgate.net, 2023. https://www.researchgate.net/post/How_to_compute_d-prime_calculation_for_a_2-alternative_force_choice_data/67a453a3c7aa13fa870310ab/
21. JASP Team, JASP (version 0.95.3), (logiciel, 2025). https://jasp-stats.org/
22. K.E. Scott, H.J. Osborn, E.N. Prince, R. Walker, Running and interpreting a paired samples t test in JASP, Exploring Diversity with Statistics Using JASP: Step-by-Step Guides, (University of Tennessee at Chattanooga, Chattanooga, 2021). https://utc.pressbooks.pub/step-by-step-JASP-guides/
23. M. Candea, I. Vasilescu, M. Adda-Decker, Inter- and intra-language acoustic analysis of autonomous fillers. Dans Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop, Aix-en-Provence, France, 10-12 septembre (2005). https://shs.hal.science/halshs-00321914v1
24. J. Stuart-Smith, Glottals past and present: a study of t-glottaling in Glaswegian. Leeds Studies in Eng. **30** (1999).
25. J. Beck, F. Schaeffler, Voice quality variation in Scottish adolescents: gender versus geography. Dans Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, Scotland, 10-14 août (2015). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0737.pdf
26. J. Pearce, Identity, socialization and environment in transgender speakers: sociophonetic variation in creak and /s/. Dans Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 5-9 août (2019). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_78.pdf
27. R.J. Podesva, Phonation type as a stylistic variable: the use of falsetto in constructing a persona. J. of Socioling. **11**, 4 (2007). https://doi.org/10.1111/j.1467-9841.2007.00334.x

28. E. Pépiot, Gender identification from speech in Parisian French and American English speakers. Paisii Hilendarski University of Plovdiv–Bulgaria Research Papers. **55**, 1B (2017). https://hal.science/hal-03004709v1/file/NTF_2017_55_1_B_60_72.pdf
29. E. Pépiot. A. Arnold, Cross-gender differences in English/ French bilingual speakers: a multiparametric study. Perceptual and Motor Skills. **128**, 1 (2021). https://doi.org/10.1177/0031512520973514

Annexe

Les citations lues pour créer les stimuli :

Tableau 8. La version française

Citation	*Origine*	*Traduction*
Il était une fois une petite fille de village.	Le petit chaperon rouge	
Sœur Anne, Sœur Anne, ne vois-tu rien venir ?	La barbe-bleue	
C'est à Monsieur le Marquis de Carabas.	Le chat botté	
Un jour mon prince viendra.	Blanche-Neige et les sept nains	
Tire la bobinette et la chevillette cherra.	Le petit chaperon rouge	
S'il te plaît, dessine-moi un mouton.	Le petit prince	
Miroir, mon beau miroir, dis-moi qui est la plus belle ?	Blanche-Neige et les sept nains	
Ils vécurent longtemps et eurent de nombreux enfants.	Traditionnel	
Vous êtes le phénix des hôtes de ce bois.	Le corbeau et le renard	
To be or not to be; that is the question.	La tragique histoire d'Hamlet, prince de Danemark	Etre ou n'être pas, la question est là.

Tableau 9. La version anglaise

Citation	*Origine*	*Traduction*
Once upon a time in a land far, far away.	Traditionnel	Il était une fois dans un pays lointain.
It was a dark and stormy night.	Traditionnel	C'était une nuit orageuse et sombre.

They lived happily ever after.	Traditionnel	Ils vécurent longtemps et eurent de nombreux enfants.
Someday, my prince will come.	Blanche-Neige et les sept nains	Un jour mon prince viendra.
Someone's been sleeping in my bed!	Boucle d'or et les trois ours	Quelqu'un a dormi dans mon lit !
Mirror, mirror, on the wall, who's the fairest of them all?	Blanche-Neige et les sept nains	Miroir, mon beau miroir, dis-moi qui est la plus belle ?
Fee-fi-fo-fum, I smell the blood of an Englishman!	Jack et le haricot magique	Fi-faï-foe-fum, je sens le sang d'un Anglais !
Grandmother, what big ears you have!	Le petit chaperon rouge	Grand-mère, que tu as des grandes oreilles !
I'll huff and I'll puff, and I'll blow your house in!	Les trois petits cochons	Je soufflerai, je soufflerai, et ta maison s'envolera !
To be or not to be; that is the question.	La tragique histoire d'Hamlet, prince de Danemark	Etre ou n'être pas, la question est là.