Perception auditive des effets de différents masques anti-COVID sur la parole conversationnelle, déclamée et le chant : étude de cas

Claire Pillot-Loiseau, Bernard Harmegnies

Perception auditive des effets de différents masques anti-COVID sur la parole conversationnelle, déclamée et le chant : étude de cas

Claire PILLOT-LOISEAU
Université Sorbonne Nouvelle, Laboratoire de Phonétique et Phonologie UMR 7018 (FRANCE)

Bernard HARMEGNIES
Institut de Recherche en Sciences et Technologies du Langage, Université de Mons (BELGIQUE)

Résumé

Quels sont les effets perceptifs, en environnement calme, du port de six types de masques anti-COVID sur l'identification et la discrimination de consonnes, celle de phrases parlées, déclamées et chantées en français produits par un homme et une femme ? 21 auditeurs ont identifié puis discriminé les consonnes /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/ en intervocalique produites par la locutrice, puis 39 auditeurs ont discriminé une phrase parlée, déclamée (par un locuteur et une locutrice) et chantée (par la locutrice, chanteuse) avec et sans masque : l'identification consonantique est peu affectée, excepté pour /b/ avec le masque à fenêtre transparente, mais la présence des masques à fenêtre transparente et FFP2 est fortement discriminée et avec certitude par rapport à la condition sans masque, à l'exception du chant qui semble peu sensible à l'atténuation de timbre entendue et verbalisée par les auditeurs pour les autres tâches.

Mots clé : Masques faciaux, perception auditive, consonnes, voix parlée, voix chantée.

Abstract

What are the perceptual effects, in a quiet environment, of wearing six types of anti-COVID masks on the identification and discrimination of consonants, and on discrimination in spoken, declaimed and sung French sentences produced by one man and one woman? 21 listeners identified then discriminated the consonants /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/ in intervocalic position, produced by the female speaker. 39 listeners discriminated a spoken, declaimed (produced by a male and a female speaker) and sung (produced by the female speaker, a singer) sentence with and without mask: consonant identification was little affected, except for /b/ with the transparent window mask, but the presence of the transparent window and FFP2 masks was strongly discriminated with certainty compared to the no-mask condition, with the exception of singing, which appeared to be insensitive to the timbre attenuation heard and verbalized by the listeners for the other tasks.

Keywords: Facial masks, auditory perception, consonants, speaking voice, singing voice.

Resumen

¿Cuáles son los efectos perceptivos, en un ambiente no ruidoso, de llevar seis tipos de mascarilla anti-COVID en la identificación y la discriminación de consonantes, de enunciados orales, declamados y cantados en francés producidos por un hombre y una mujer? 21 informantes identificaron y luego discriminaron las consonantes /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/ en posición intervocálica producidas por la locutora, luego 39 informantes discriminaron un enunciado hablado, declamado (por un locutor y una locutora) y cantado (por la locutora, que era cantante) con mascarilla y sin ella: la identificación consonántica apenas se vio afectada, salvo para /b/ con la mascarilla transparente, pero la utilización de mascarillas transparentes y FFP2 ciertamente está muy discriminada con respecto a la ausencia de mascarillas, salvo en el caso del canto que parece poco sensible a la atenuación del timbre percibida y verbalizada por los informantes para las demás tareas.

Palabras clave: mascarillas faciales, percepción auditiva, consonantes, voz hablada, voz cantada.

Resum

Quins són els efectes perceptius, en un ambient no sorollós, de portar sis tipus de mascareta anti-COVID en la identificació i la discriminació de consonants, d'enunciats orals, declamats i cantats en francès produïts per un home i una dona? 21 informants van identificar i després discriminar les consonants /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/ en posició intervocàlica produïdes per la locutora, després 39 informants van discriminar un enunciat parlat, declamat (per un locutor i una locutora) i cantat (per la locutora, que era cantant) amb mascareta i sense: la identificació consonàntica gairebé no es va veure afectada, tret del cas de /b/ amb la mascareta transparent, però la utilització de mascaretes transparents i FFP2 certament està molt discriminada respecte a l'absència de mascaretes, lleva del cas del cant que sembla poc sensible a l'atenuació del timbre percebuda i verbalitzada pels informants per a les altres tasques.

Paraules clau: mascaretes facials, percepció auditiva, consonants, veu parlada, veu cantada.

1. Introduction

La pandémie a obligé tout individu à se protéger de la COVID-19 avec un masque : il est probable que le recours aux dispositifs de ce type s'impose comme un comportement banal dans les usages sociaux partout à la surface de la planète, de prévalence évoluant au rythme de l'évolution des variants de la Covid ou d'autres épidémies ou pandémies faisant désormais leur apparition.

Les effets acoustiques des masques faciaux sur la parole montrent des atténuations au-dessus de 1 kHz et pour des fréquences plus élevées dans des proportions différentes en fonction de leur type, au détriment de masques à fenêtre transparente et en faveur des masques chirurgicaux (Corey *et al.*, 2020, Magee *et al.*, 2020). Sur un plan perceptif, l'atténuation acoustique de la pression sonore par le matériau du masque réduit la capacité de perception de la parole, en particulier dans les situations bruyantes ; par exemple, comme le rapportent Rahne *et al.* (2021), un masque chirurgical réduit significativement le seuil de perception de la parole dans le bruit. De plus, la plupart des études dans le domaine ont adopté une approche holistique pour apprécier l'impact du port d'un masque facial (entre autres : Fecher, Watt, 2013, Palmiero *et al.*, 2016, Magee *et al.*, 2020, Rahne *et al.*, 2021), et éventuellement du style de parole (Cohn *et al.*, 2021) sur l'intelligibilité de celle-ci, la plupart du temps en langue anglaise. L'impact perceptif des masques faciaux antiviraux sur tout l'énoncé et au niveau consonantique en parole, déclamation et chant, n'a donc pas encore été quantifié en français pour les masques courants, ni pour ceux adaptés aux chanteurs.

Cette recherche exploratoire, et donc dénuée de prétentions nomothétiques, vise à aborder une vaste problématique encore incomplètement investiguée et à identifier les questions de recherche à y développer, notamment concernant les relations entre perception par l'interlocuteur et production par le locuteur sous diverses conditions masquées dans une variété de tâches. Le masque est ici considéré comme élément perturbateur externe (au sens de la théorie de la perturbation : Sock, 2001) de la production et de la diffusion (articulation, boucle de régulation audio-phonatoire, transmission du son rayonné aux lèvres) et, en conséquence, comme potentiel inducteur d'altérations perceptives aux plans local (identification des consonnes occlusives et constrictives) et global (qualité vocale et/ou intelligibilité modifiées), variables selon le type de masque, le corpus et la tâche de production. Nous cherchons donc à interroger les effets perceptifs, en environnement calme, du port de six types de masques anti-COVID sur l'identification et la discrimination de consonnes, et sur la discrimination de phrases parlées, déclamées et chantées en français produites par un homme et une femme, avec et sans masque.

2. Méthode

2.1. Corpus enregistré

Un test d'identification consonantique et cinq tests de discrimination ont été élaborés avec le logiciel Praat (Boersma et Weenink, 2021). A une fréquence d'échantillonnage de 44 100 Hz, et à l'aide d'un microphone serre-tête AKGC520 placé à 3 cm de la bouche, d'une carte-son Edirol UA-25 USB Audio Capture et du logiciel Audacity, une locutrice francophone de 50 ans a enregistré trois répétitions des logatomes /aCa/, où C= consonne /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/, une lecture parlée et déclamée du texte *La Bise et le Soleil*, et la mélodie chantée *Clair de Lune* de Fauré sans et avec les masques faciaux suivants : chirurgical, tissu, FFP2, transparent de marque à fenêtre transparente, et deux masques conçus pour les chanteurs (grand masque et masque chanteur Aertec). À l'aide d'un microphone Sennheiser e840S placé à 35 cm de la bouche, d'une carte-son A-D Focusrite Scarlett 2i4 USB et du logiciel Praat (Boersma et Weenink, 2021), un locuteur francophone de 64 ans a enregistré une lecture parlée de *La Bise et le Soleil* sans et avec les masques chirurgical, tissu, FFP2, et à fenêtre transparente. La voyelle /a/ a été choisie car sa réalisation correspond à la zone fréquentielle de meilleure audibilité. À partir de ces enregistrements, la deuxième répétition des logatomes /aCa/ (où C= consonne /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/), l'extrait *La bise et le soleil se disputaient* de *La Bise et le Soleil* parlé et déclamé, et l'extrait *Au calme clair de lune, triste et beau* de la mélodie chantée, ont été sélectionnés pour constituer les stimuli de ces tests de perception. Les signaux et spectrogrammes en filtrage large superposés à la courbe de fréquence fondamentale F0 montrant la mélodie sont illustrés pour ces phrases parlées par le locuteur sans masque (figure 1, en haut à gauche), avec le masque à fenêtre transparente (figure 1, en haut à droite), et ces mêmes phrases parlées dans les mêmes conditions par la locutrice, à savoir sans masque (figure 1, en bas à gauche) et avec le masque à fenêtre transparente (figure 1, en bas à droite). La figure 2 illustre les mêmes tracés pour les phrases déclamées et chantées par la locutrice dans les mêmes conditions.

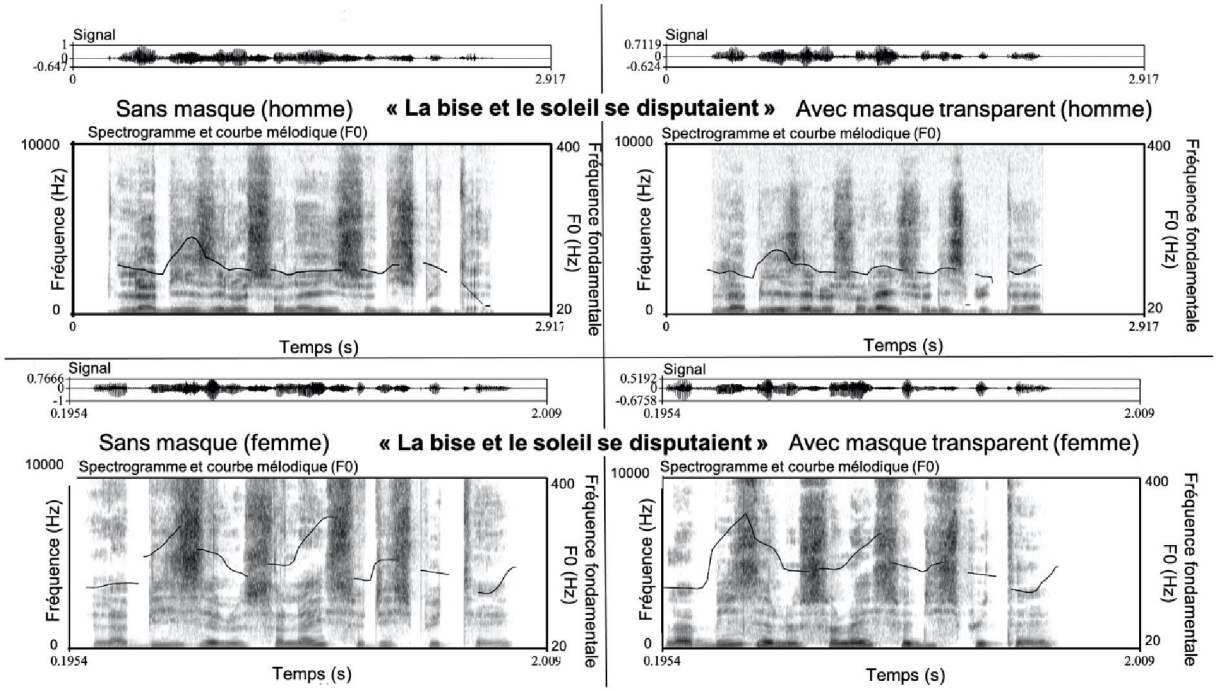

Figure 1. Signaux et spectrogramme en filtrage large superposé à la courbe de fréquence fondamentale F0 de la phrase « La bise et le soleil se disputaient » parlée par le locuteur (haut) et la locutrice (bas) sans masque (à gauche) et avec le masque à fenêtre transparente (droite)

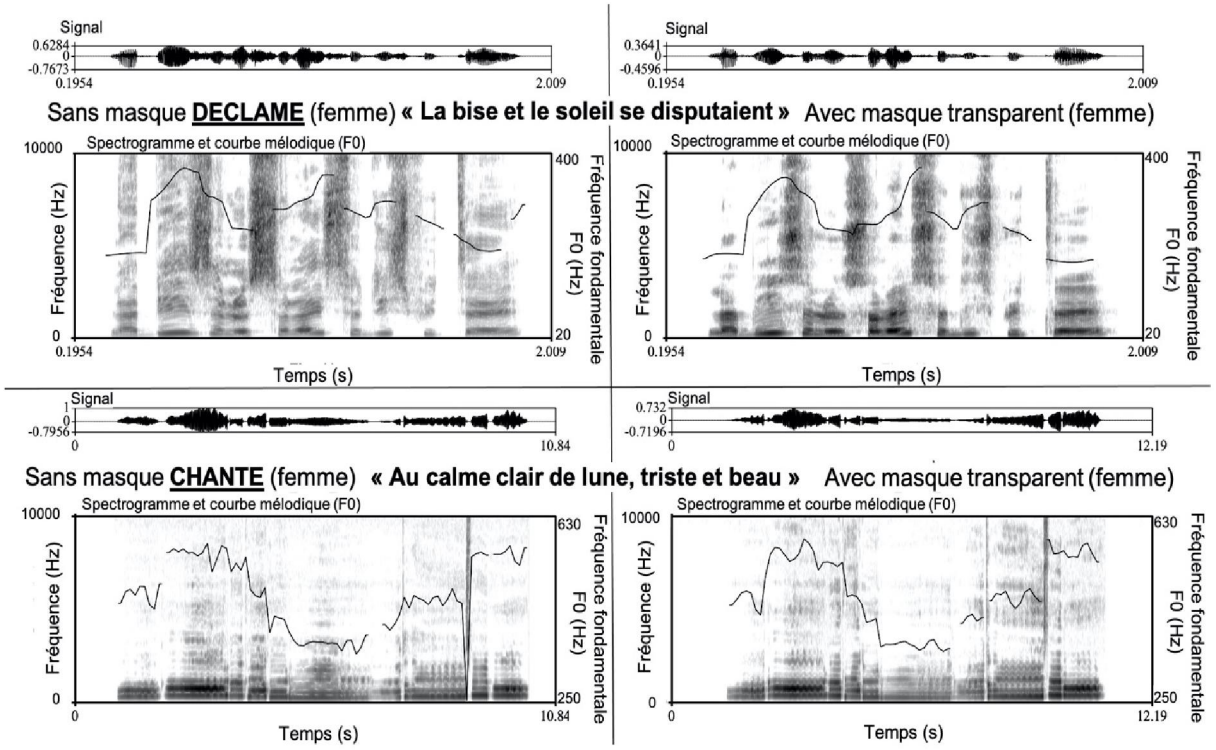

Figure 2. Signaux et spectrogramme en filtrage large superposé à la courbe de fréquence fondamentale F0 de la phrase « La bise et le soleil se disputaient » déclamée par la locutrice (haut) et de la phrase « Au calme clair de lune, triste et beau » chantée par la locutrice (bas) sans masque (à gauche) et avec le masque à fenêtre transparente (droite)

2.2. Tests perceptifs

Chaque type de test était précédé d'un test d'entraînement sur quatre autres stimuli. Puis, pour le test d'identification, les auditeurs ont entendu en ordre semi-randomisé 70 logatomes /aCa/. Ces 70 logatomes se composent d'abord de 60 stimuli /aCa/ avec 12 consonnes (/p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/) x 5 conditions (sans masque, chirurgical, tissu, FFP2, à fenêtre transparente) : on compte, par exemple, le logatome /apa/ produit sans le masque, puis avec le masque chirurgical, puis avec les masques en tissu, FFP2 et à fenêtre transparente, soit 5 stimuli ; puis le logatome /ata/ répété dans les mêmes conditions, et ensuite les logatomes /aka/, /aba/, etc. À ces 60 logatomes utilisés comme stimuli pour nos tests de perception ont été ajoutés 10 logatomes : à nouveau /apa/ produit avec chaque type de masque (soit 5 stimuli) et à nouveau /ava/ identiquement répété (5 autres stimuli). Ces 10 logatomes supplémentaires permettent d'évaluer la cohérence intra-auditeur dans les réponses obtenues, pour deux consonnes dont nous supposons que l'action du masque peut être plus importante en raison de leur lieu d'articulation (consonnes labiales et labio-dentales) pouvant interférer avec le port d'un masque. Pour chaque stimulus écoutable une deuxième fois, les auditeurs devaient identifier quelle consonne ils entendaient, puis leur degré de certitude (de 1 *pas du tout sûr.e* à 4 *totalement sûr.e*). À la fin de ce test d'identification consonantique, les auditeurs étaient invités à commenter librement cette tâche.

Ces auditeurs devaient ensuite réaliser un premier test de discrimination basé sur 70 paires composées de ces logatomes parlés. L'espace temporel entre deux stimuli d'une même paire était de 600 ms, et la présentation des paires était commandée par le sujet. Une paire était composée d'un premier logatome /aCa/ produit sans masque, et du deuxième identique sans masque, ou bien de ce deuxième logatome produit avec les masques chirurgicaux ou en tissu, ou FFP2, ou à fenêtre transparente. Cela donnait, par exemple, à écouter aux auditeurs les paires suivantes : /afa/ produit sans masque-/afa/ produit sans masque, /afa/ produit sans masque-/afa/ avec le masque chirurgical, /afa/ produit sans masque-/afa/ avec le masque en tissu, etc. On comptait donc dans ce test 1 de discrimination 5 paires par consonne (une paire par type de masque), 12 consonnes testées auxquelles se sont ajoutées 10 duplications (5 avec /b/ et 5 dans l'ordre inverse pour les autres occlusives entre *sans masque*

et *à fenêtre transparente*). Pour chaque paire écoutable une deuxième fois, les auditeurs devaient répondre si les paires leur semblaient identiques ou différentes, puis leur degré de certitude (de 1 *pas du tout sûr.e* à 4 *totalement sûr.e*).

Quatre autres tests de discrimination avec les mêmes consignes, le même choix de conditions des stimuli et portant sur les phrases parlées par une femme (test 2), par un homme (test 3, séparé du précédent), déclamées (test 4) et chantées par une femme (test 5) ont également été proposés aux auditeurs : après quatre stimuli d'entraînement, pour chaque stimulus écoutable une deuxième fois, les auditeurs devaient répondre si les paires leur semblaient *identiques*, *un peu différentes* ou *très différentes*, puis exprimer leur degré de certitude de 1 *pas du tout sûr.e* à 4 *totalement sûr.e*. 17 paires de phrases parlées par le locuteur masculin ont constitué le test 3 (8 paires sans masque et avec un type de masque, dans les deux ordres, 5 paires avec le même stimulus, issu de la production sans masque ou avec un masque donné, et 4 paires dupliquées avec l'ordre sans masque / chaque type de masque). 24 phrases parlées (test 2), déclamées (test 4) et chantées (test 5) par la locutrice ont constitué la liste de stimuli se décomposant comme suit : 12 paires sans masque et avec un type de masque, dans les deux ordres, 6 paires avec le même stimulus, issu de la production sans masque ou avec un masque donné, et 6 paires dupliquées avec l'ordre sans masque / chaque type de masque. Pour ces 5 tests, les stimuli ont été classés en ordre semi-randomisé, c'est-à-dire sans alternance régulière pouvant influencer les résultats obtenus, mais avec le même ordre pour tous les auditeurs. Bien que tout le soin désirable ait été apporté par les locuteurs à éviter les variations prosodiques, il était demandé aux auditeurs de ne pas porter leur attention sur l'intonation des phrases écoutées, mais à tout autre aspect pouvant expliquer une différence perçue par les auditeurs, même infime, comme la qualité d'enregistrement, la consonne, la voyelle, le timbre ou tout autre aspect jugé pertinent par les auditeurs. Ceux-ci ignoraient que l'utilisation de différents types de masque était la source de variation des stimuli entendus.

Les auditeurs devaient, par ailleurs, répondre à un questionnaire précisant leur profil (âge, langue maternelle, déficience auditive auto-estimée, activités vocale et musicale professionnelle ou de loisirs), et leurs

conditions de passation du test (au casque ou non, durée passée à chaque test). À ces questions concernant leur profil s'ajoutaient deux interrogations : les sujets devaient d'abord répondre à la question suivante : *Sur quel(s) indice(s) vous êtes-vous appuyé.e pour différencier les deux phrases d'une paire écoutée au test de comparaison (discrimination) ? Écrivez ces indices ci-dessous*. Il y avait une question par type de phrases perçues. Finalement, les sujets devaient librement verbaliser leurs impressions sur les quatre derniers tests de discrimination. La verbalisation libre se définit comme « la description des caractéristiques des objets perceptifs ». La méthodologie consiste alors à effectuer une analyse de contenu du protocole verbal librement généré par les auditeurs, issu d'un questionnement totalement ouvert. Les consignes données sont de verbaliser les stimuli présentés individuellement, comme dans le cas de cette recherche. Cette méthode a pour avantages de ne pas contraindre ni limiter l'auditeur aux choix de champs sémantiques restreints comme peuvent l'être les méthodes à choix forcé (Nosulenko, Samoylenko, 1997).

2.3. Auditeurs

Tous les auditeurs, excepté un, ont passé ces tests de perception avec un casque. 22 auditeurs francophones d'âge moyen 30,6 ans (σ=12, de 20 à 65 ans) dont un ayant déclaré d'importants problèmes auditifs, ont répondu aux tests d'identification et de discrimination des logatomes parlés, parmi lesquels 14 étudiants en orthophonie, 3 en linguistique, 3 ayant une activité vocale de loisirs et 10 une activité musicale extraprofessionnelle. Les résultats de l'auditeur ayant auto-déclaré une déficience auditive ont été exclus de l'étude.

39 auditeurs (38 pour le test de discrimination chanté) sans problèmes auditifs auto-déclarés, dont 8 ayant déjà passé les tests précédents, ont répondu aux 4 tests de discrimination des phrase chantées, parlées (femme puis homme), et déclamées dans cet ordre. Ils sont tous francophones, excepté un anglophone, et d'âge moyen 36,8 ans (σ=13, de 20 à 64 ans). 6 sont chanteurs professionnels, 4 sont musiciens, 11 sont orthophonistes spécialisés en rééducation vocale, 4 sont étudiants en orthophonie, 2 le sont en linguistique et le reste de ces sujets est enseignant. 18 d'entre eux exercent une activité vocale de loisirs, et 6 une activité musicale dans cette condition.

3. Résultats

3.1. Tests d'identification consonantique

3.1.1. Pourcentages d'identification

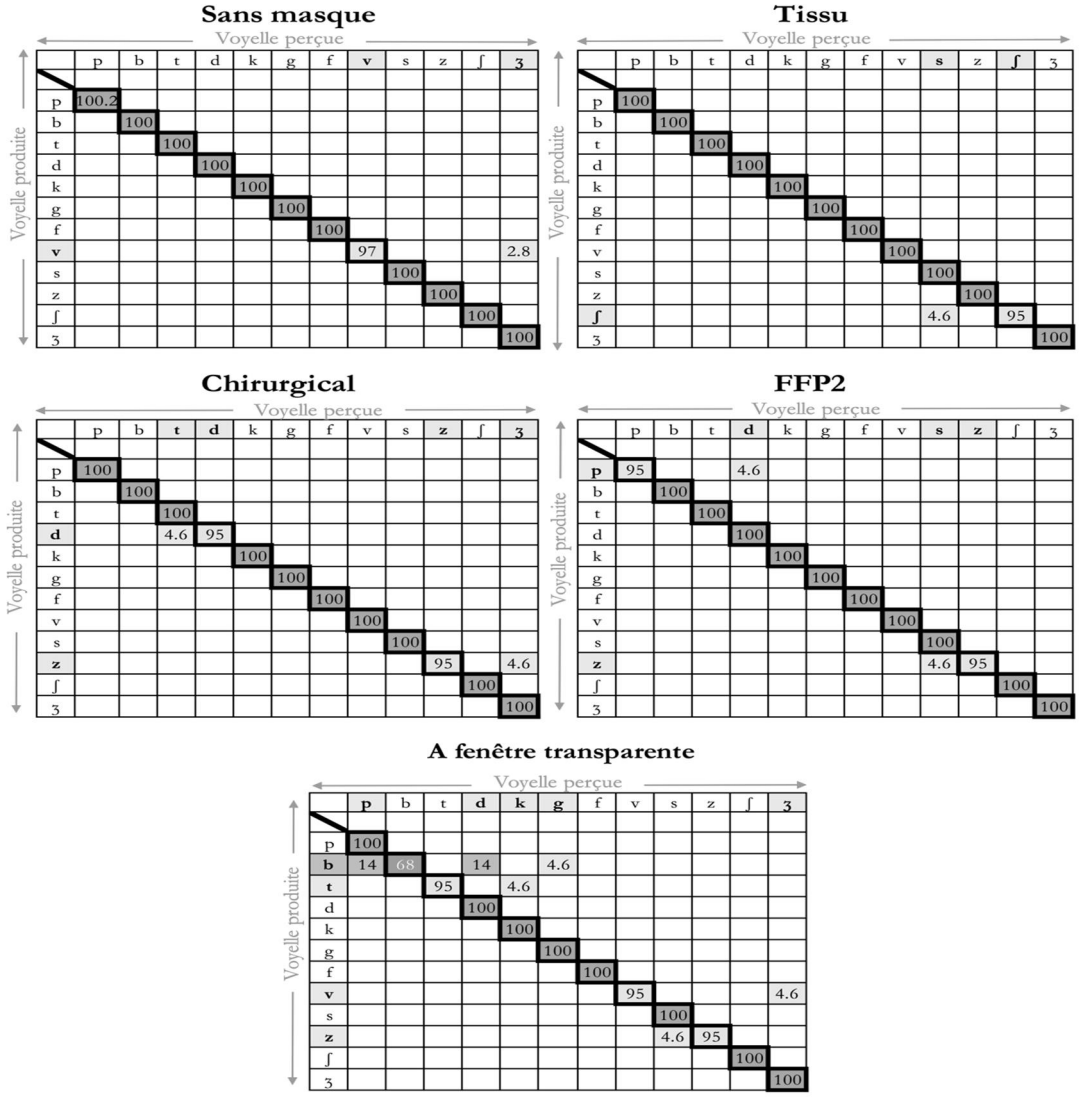

Figure 3. Matrices de confusion des tests d'identification consonantique sans et avec chaque type de masque, tous auditeurs confondus (N=21)

La figure 3 montre les matrices de confusion des tests d'identification consonantique sans et avec chaque type de masque, tous auditeurs confondus : il existe peu de changements perçus dans l'identification consonantique en fonction du masque porté : seul le masque à fenêtre transparente affecte l'identification de /b/, identifié chez le tiers des auditeurs comme /p/, /d/ ou /g/. /z/ est assimilé à /s/ chez un sujet pour ce masque et le masque FFP2, mais à /ʒ/ avec le masque chirurgical. /ʃ/ est identifié comme /s/ par un auditeur pour le masque en tissu, et /v/

comme /ʒ/ avec le masque à fenêtre transparente. La perception altérée n'entraîne donc pas de changement de mode d'articulation, mais plutôt des changements de voisement et de lieu d'articulation : dans la plupart des cas de perception consonantique altérée, la consonne perçue est soit dévoisée (exemple : /b/ perçue comme /p/ avec le masque à fenêtre transparente), soit perçue comme plus postérieure (exemple : /b/ perçue comme /d/ ou /g/ ; /v/ perçue comme /ʒ/ avec ce même masque).

L'âge de nos auditeurs ne semble pas être un facteur influençant l'identification consonantique : les altérations dans l'identification perçue de /b/ produit avec un masque à fenêtre transparente concernent 2 auditrices de 21 ans (une erreur de voisement et une erreur de lieu d'articulation perçus), une de 23 ans (erreur de voisement), une de 35 ans, une de 38 ans (erreur de lieu d'articulation pour ces deux dernières), une de 52 ans (erreur de lieu d'articulation) et une de 64 ans (erreur de voisement).

3.1.2. Degré de certitude

À partir de l'échelle proposée aux auditeurs de 1 *pas du tout sûr.e* à 4 *totalement sûr.e*, l'analyse des degrés de certitude, tous auditeurs confondus, a été effectuée à partir de la somme des réponses obtenues par stimulus, convertie en pourcentage. Ainsi, 100 % signifie que tous les auditeurs étaient totalement sûrs de leurs réponses pour un stimulus donné.

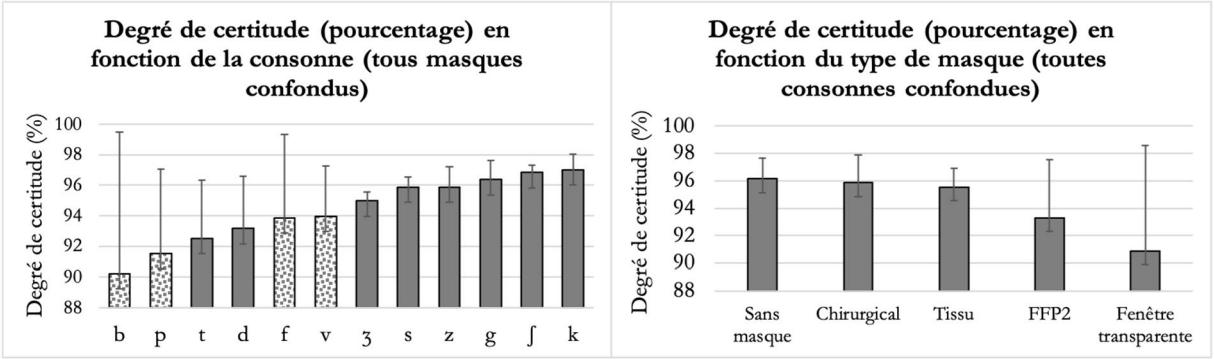

Figure 4. Pourcentage des degrés de certitude en fonction de la consonne (gauche) et du type de masque (à droite), tous auditeurs confondus. Les consonnes bilabiales /p b/ et labio-dentales /f v/, aux pourcentages de degrés de certitude plus variables, sont illustrées avec un motif différent

La figure 4 montre des pourcentages de degrés de certitude élevés. Toutefois, les auditeurs répondent avec moins de certitude et de façon plus variable pour /p b f v/ et /t d/, les autres consonnes totalisant des pourcentages élevés compris entre 95 % et 97 %. En outre, les degrés de certitude sont moins élevés et plus variables pour les masques à fenêtre

transparente (91 %) puis FFP2 (93,3 %). Les identifications des consonnes sans masque (96,1 % de certitude), avec le masque chirurgical (95,8 %) et en tissu (95,5 %) sont effectuées avec des pourcentages de degrés de certitude plus élevés et semblables.

3.1.3. Commentaires des auditeurs

Les auditeurs ont laissé très peu de commentaires relatifs à cette tâche d'identification. La majorité a affirmé que cette tâche leur était facile et beaucoup de commentaires tournaient autour d'affirmations du type « Je ne sais pas quel est le but de la recherche, mais, mis à part deux ou trois stimuli, je n'ai pas du tout hésité quant à l'identification des consonnes ».

3.2. Tests de discrimination

3.2.1. Accords inter-auditeurs

D'après Kreiman, Gerratt (1990), l'accord (ou concordance) inter auditeurs existe lorsque des auditeurs portent des jugements strictement identiques (accord absolu), ou à plus ou moins un point près (accord relatif) sur des échelles suffisamment larges ; les juges assignent alors une même impression perceptive à l'écoute d'une même production de parole. Nous avons utilisé ces deux types d'accord pour évaluer la concordance entre nos 39 auditeurs sur tous les stimuli des tests de discrimination n°2 à 5, portant sur les phrases parlées, déclamées et chantées.

La figure 5 illustre les accords inter-auditeurs absolu et relatif (à un point près) pour les tests 2 à 5 de discrimination, en fonction du type de masque. Elle montre que les accords absolus sont modérés (entre 40 et 60 %) pour la perception des paires de phrases produites avec les masques FFP2, chirurgical, tissu, grand masque et masque chanteur. Cependant, il faut remarquer qu'un grand nombre d'auditeurs (39) a fait l'objet de ce calcul. Ces taux absolus atteignent toutefois des valeurs supérieures à 90 % pour les paires identiques, sauf pour ces paires de phrases chantées. Puis, ce sont les phrases produites avec des masques à fenêtre transparente qui entraînent des valeurs d'accord inter-auditeurs absolus forts (situés entre 60 et 80 % sauf pour les phrases chantées).

On observe également que les accords relatifs avoisinent les 100 % pour l'ensemble des tests de discrimination, surtout concernant les paires de phrases identiques, sans masque / masque à fenêtre transparente et FFP2.

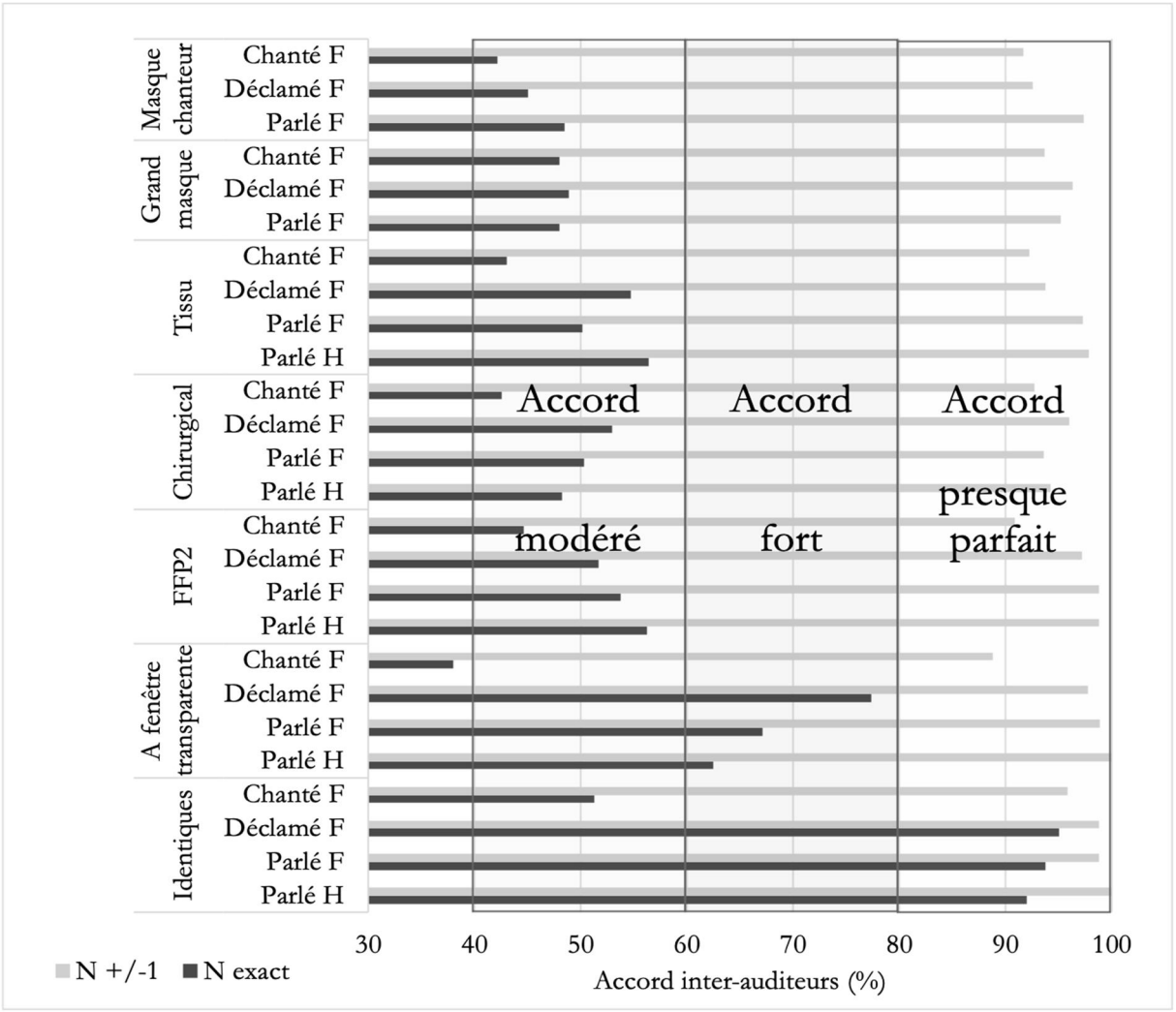

Figure 5. Accords inter-auditeurs absolu (gris) et relatif (à un point près, noir) pour les tests 2 à 5 de discrimination, en fonction du type de masque

3.2.2. Pourcentage de discrimination

Pour rappel, les sujets ont effectué 5 tests de discrimination de paires sans et avec chaque type de masque : à partir de 70 paires de logatomes /aCa/ (où C= /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/, test 1), et de 17 paires de phrases « La bise et le soleil se disputaient » parlées par un locuteur masculin (test 3), 24 phrases parlées (test 2, même phrase), déclamées (test 4, même phrase) et chantées (test 5 : « Au calme clair de lune, triste et beau ») par une locutrice.

La figure 6 illustre les pourcentages de discrimination obtenus en fonction de l'individu ayant produit les stimuli, du type de masque, et de la tâche produite, tous auditeurs confondus : plus ce pourcentage est élevé, plus les auditeurs ont discriminé comme différents les stimuli d'une même paire. Nous pouvons observer des taux maximaux de ce pourcentage pour

les paires sans masque/masque FFP2 (de 68,7 % pour les consonnes à 89,3 % pour les phrases parlées par la locutrice) et surtout pour les paires sans masque / masque à fenêtre transparente (de 65,1 % pour les phrases chantées à 95,3 % pour les phrases déclamées). Les paires identiques n'aboutissent pas à des pourcentages de discrimination nuls (de 34,6 % pour les phrases parlées par la locutrice, à 50,2 % pour les consonnes) : rappelons que les auditeurs ignoraient la source de différences entre deux stimuli d'une même paire. Les pourcentages de discrimination concernant les masques en tissu (de 55,1 % pour les consonnes à 72,6 % pour la phrase parlée par le locuteur) et chirurgicaux (59,3 % pour les consonnes à 71,1 % pour les phases chantées par la locutrice) sont de valeurs intermédiaires entre les paires identiques et celles contenant les productions utilisant les masques FFP2 et à fenêtre transparente.

En d'autres termes, les masques FFP2 et à fenêtre transparente font l'objet des plus grands différences perçues avec les productions sans masque, mais il existe une variation en fonction du type de production : la discrimination perçue est croissante dans l'ordre suivant pour la voix parlée des consonnes et de la phrase par la locutrice : sans masque < Masque chanteur < Tissu < Chirurgical < Grand masque < FFP2 < Masque à fenêtre transparente. L'ordre des phrases déclamées est analogue, mais le taux de discrimination entre les masques en tissu et chirurgical, et entre le grand masque et le masque FFP2, varient peu. Concernant la voix parlée masculine, on observe l'ordre Chirurgical < Tissu < FFP2 < Masque à fenêtre transparente. Enfin, les phrases chantées par la locutrice révèlent moins de différences perçues, selon l'ordre suivant : Grand masque < Masque à fenêtre transparente < Masque Chanteur < Tissu < Chirurgical < FFP2.

Les stimuli ne sont pas identiquement perçus en fonction de l'individu qui les a produits pour les paires utilisant les masques en tissu, probablement à cause d'une différence de matériau entre ces types de masques portés par le locuteur et la locutrice.

En outre, la nature de la tâche influence la perception des stimuli par les auditeurs : à type de masque égal, les paires de phrases sont plus discriminées que les consonnes, avec toujours des pourcentages de discrimination plus élevés pour les masques FFP2 et à fenêtre transparente. De plus, les paires de phrases déclamées sont plus discriminées entre les

conditions *sans masque* et *masque chanteur*, *sans masque* et tissu, et *sans masque* et *chirurgical*, à l'inverse de la condition *sans masque* et *FFP2*. Enfin, la perception des paires contenant les phrases chantées par la locutrice semble obéir à un autre *pattern* que celui des autres tâches, vers des taux de discrimination moins variables en fonction du type de masque : par exemple, les paires *sans masque* et *masque à fenêtre transparente* ne sont pas les plus différenciées (65,1 % contre 71,1 % pour les paires de phrases *sans masque* et *masque chirurgical*).

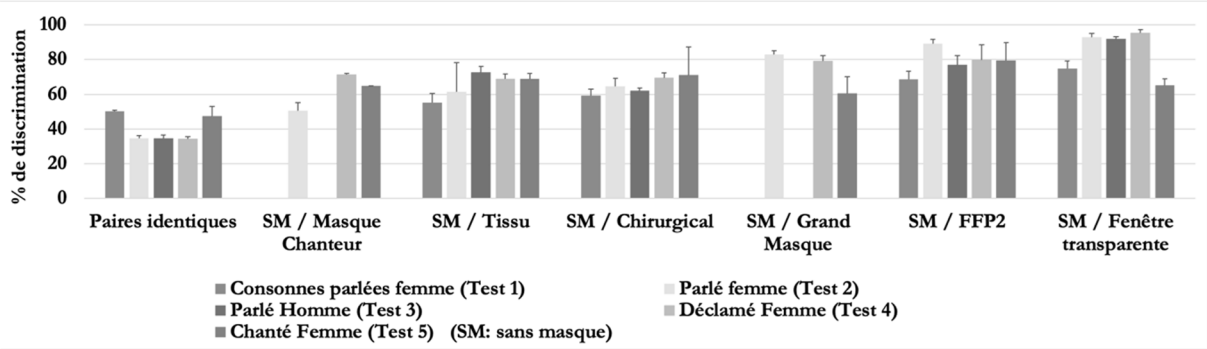

Figure 6. Pourcentage de discrimination en fonction de l'individu ayant produit les stimuli, du type de masque, et de la tâche produite, tous auditeurs confondus. Dans la légende, « Parlé », « déclamé », « chanté » signifie « phrase parlée », « phrase déclamée », et « phrase chantée ». Les masques chanteur et grand masque n'ont pas été utilisés pour la production des consonnes chez la locutrice, et de la phrase parlée chez le locuteur, ce qui explique l'absence de données dans ces rubriques

L'âge de nos auditeurs ne semble pas être un facteur influençant la discrimination des phrases : par exemple, le taux de discrimination des phrases parlées sans et avec le masque à fenêtre transparente n'augmente pas avec l'âge : la corrélation de Pearson entre l'âge des auditeurs et le pourcentage de discrimination, non significative, est presque nulle (r_{39}=0,07). Cela signifie qu'aussi bien des auditeurs plus âgés que des personnes jeunes ont discriminé identiquement les phrases parlées dans ces conditions.

Il est cependant à noter que ces pourcentages de discrimination varient en fonction d'autres éléments du profil des auditeurs. La figure 7, illustrant la comparaison de ces pourcentages pour les phrases chantées entre 6 auditeurs-chanteurs professionnels et 7 auditeurs-non chanteurs, révèle que les chanteurs professionnels perçoivent, davantage que les non chanteurs, plus de différences entre la voix et la parole produites sans et avec masque, sauf pour le « masque chanteur ». En outre, ils perçoivent moins de différences que les non chanteurs pour les paires de stimuli identiques.

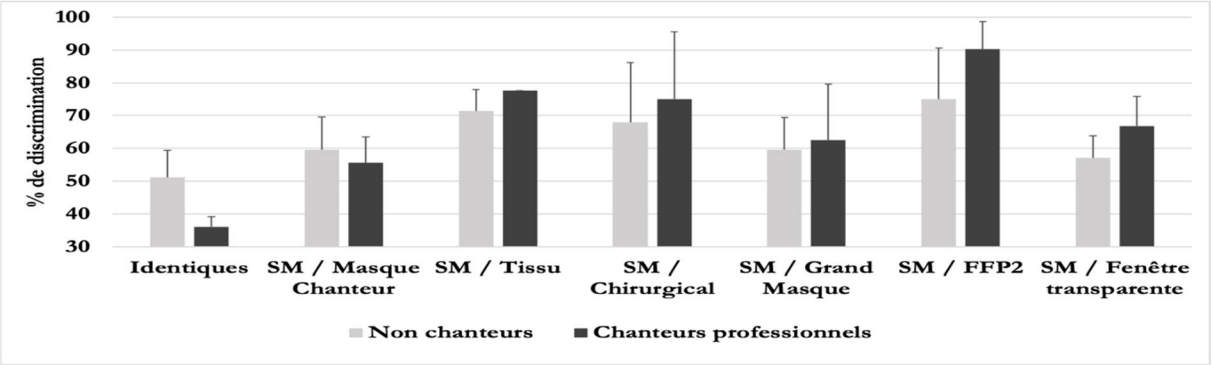

Figure 7. Pourcentage de discrimination des phrases chantées par la locutrice en fonction du type de masque : comparaison de la perception de 6 auditeurs-chanteurs professionnels, et 7 auditeurs-non chanteurs. SM : Sans masque

3.2.3. Degré de certitude

La figure 8 illustre les degrés de certitude convertis en pourcentage, obtenus en fonction de l'individu ayant produit les stimuli, du type de masque et de la tâche produite, tous auditeurs confondus : plus ce pourcentage est élevé, plus les auditeurs sont certains de leur discrimination des stimuli d'une même paire. Nous pouvons observer des taux maximaux de ce pourcentage pour les paires de phrases déclamées par la locutrice, et des taux minimaux pour les degrés de certitude concernant la discrimination des paires de logatomes, puis des phrases chantées. Toutefois, ces degrés de certitude varient peu selon le type de masque porté : en moyenne de 77,4 % pour les paires identiques, de 76 % pour les paires sans masque / masque en tissu, et 78 % pour les paires sans masque / masque chirurgicaux, ils atteignent 84,1 % pour les paires sans masque / masque chanteur, 84,8 % pour les paires sans masque / masque FFP2, et 85,5 % pour les paires sans masque / masque avec fenêtre transparente.

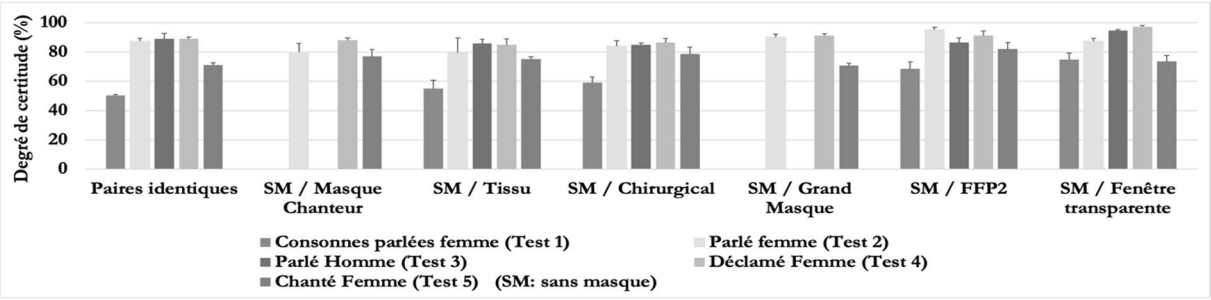

Figure 8. Degrés de certitude en pourcentage, en fonction de l'individu ayant produit les stimuli, du type de masque, et de la tâche produite, tous auditeurs confondus. Dans la légende, « Parlé », « déclamé », « chanté » signifie « phrase parlée », « phrase déclamée », et « phrase chantée ». Les masques chanteur et grand masque n'ont pas été utilisés pour la production des consonnes chez la locutrice, et de la phrase parlée chez le locuteur, ce qui explique l'absence de données dans ces rubriques

Il est à noter que ces degrés de certitude varient en fonction du profil des auditeurs. La figure 9, illustrant la comparaison de ces pourcentages pour les phrases chantées entre 6 auditeurs-chanteurs professionnels et 7 auditeurs-non chanteurs, révèle que les chanteurs professionnels répondent, davantage que les non chanteurs, avec plus de certitude et de façon moins variable, pour les paires identiques, sans masque/chirurgical, sans masque/grand masque, sans masque/FFP2 et sans masque/fenêtre transparente. Les deux profils d'auditeurs recueillent des pourcentages de degrés de certitude analogues pour les paires contenant les masques *chanteur* et en tissu, avec plus de variabilité dans les réponses obtenues pour les paires composées de productions avec le masque *chanteur*.

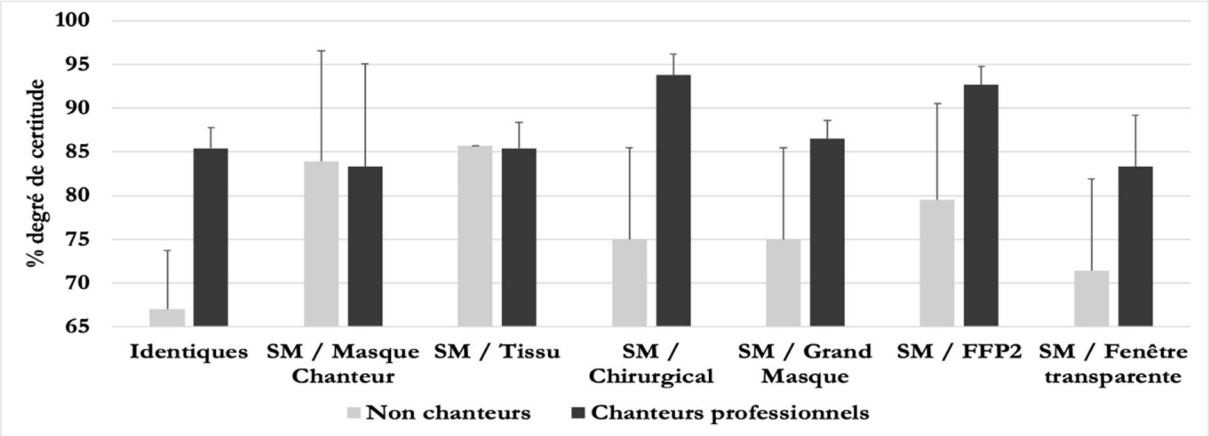

Figure 9. Pourcentage des degrés de certitude des phrases chantées par la locutrice en fonction du type de masque : comparaison de la perception de 6 auditeurs-chanteurs professionnels, et 7 auditeurs-non chanteurs. SM : Sans masque

3.2.4. Commentaires des auditeurs

Contrairement aux tests d'identification consonantique, tous les auditeurs ont émis un grand nombre de commentaires libres, et de commentaires propres aux indices de discrimination qu'ils disent avoir utilisés.

Sur 25 commentaires libres obtenus en tout, onze auditeurs ont mentionné une plus grande difficulté à percevoir des différences entre deux phrases d'une même paire de phrases chantées, par rapport aux phrases parlées. En outre, cinq commentaires relatent la difficulté des auditeurs à s'extraire de l'intonation pour discriminer les phrases parlées par le locuteur masculin, dont la figure 1 montre en effet des intonations finales différentes entre la production sans masque et celle avec le masque à fenêtre transparente. Quelques commentaires mentionnaient que les

différences à percevoir était subtiles et fines. Enfin, rappelant que les auditeurs ignoraient la source des différences entre les productions de deux phrases d'une même paire, nous citons ce commentaire d'un enseignant de 56 ans ayant une activité vocale et musicale comme loisir, qui nous paraît particulièrement pertinent en regard de la problématique qui nous intéresse :

> *J'entends essentiellement des différences de qualité non liée à la voix elle-même mais à un traitement après coup. En d'autres termes, j'ai eu l'impression d'évaluer un traitement (type filtrage) qui aurait pu avoir comme support n'importe quel son ayant environ la même bande de fréquences que la voix. Cela ne s'applique pas aux paires où le son original n'est pas le même (homme), où dans ce cas c'est la prosodie (et notamment la mélodie) qui permet de distinguer les sons.*

Au préalable, à la fin de chaque test de discrimination des phrases (tests 2 à 5), les sujets ont librement verbalisé par écrit la nature des différences perçues entre deux sons d'une paire sans et avec masque. À partir de ces verbalisations libres, le tableau 1 en illustre le contenu pour les phrases parlées par la locutrice ; des catégories ont été repérées (première ligne du tableau) pour classifier ces expressions.

Qualité ENR	**Intensité**	**Timbre**	**Parole**	**Distance au micro**	**Bruits parasites**
Clarté de l'ENR	Intensité globale	Parfois plus étouffé	Débit, rythme, attaques	Impression de distance / micro	Bruit de salive
Qualité de l'ENR (9 occurrences)	Modulation d'intensité	Voix plus nette ou chuchotée	Vitesse de la parole, F0	Voix plus éloignée	Bruits d'ENR
Filtrage de moyennes /hautes fréquences sur l'ensemble du son	Volume sonore (2)	Résonance	Couleurs des voyelles, durée et netteté des consonnes, débit, intonation	Voix plus ou moins proche	Bruits parasites
Conditions d'ENR	Intensité globale	Assourdissement, brillance	La qualité des voyelles	Aspect plus lointain	
Traitement numérique qui assombrit +/- la voix (filtre)	Volume global	Parfois clair et parfois plutôt sombre	Longueur des phonèmes (surtout final) et contours que je percevais	Distance au micro	

Son de l'ENR	Puissance et force de la voix	Timbre sombre ou clair	Hauteur et longueur des syllabes	Impression proximité ou éloignement du micro	
Son saturé ou non	Volume forte ou piano	Assourdissement du son	Clarté de projection des consonnes		
Écrêtement	Volume (2)	Timbre (4 occurrences)	Modification durée ?		
	Intensité (6)	Clarté du son	Couleur des voyelles		
	Amplitude globale	Brillance ou étouffement du son	ARTICULA-TION		
	Atténuation de force de voix	Éclat (brillance), timbre	Intonation		
	Volume perçu	Clarté (voix "assourdie")	'z' et 's' plus ou moins mous		
	Niveau sonore vocal	Clarté, richesse en harmoniques	Prononciation		
	ENR +/- fort	Aspect feutré ou + chuchoté	Fréquence, mélodie		
		Clarté + ou moindre du timbre			
		Timbre reverb ou sec, sourd environnement			
		La brillance de la voix			
		Sourd ou avec trop d'harmoniques aigües			
		Harmonique clarté			
		La clarté du timbre			
		Voix + sourde ou + claire			
16	21	37	24	7	3

Tableau 1. Catégorisation des qualificatifs librement employés par les auditeurs après le test de discrimination des phrases parlées par la locutrice, en réponse à la question : « Sur quel(s) indice(s) vous êtes-vous appuyé.e pour différencier les deux phrases d'une paire écoutée au test de comparaison (discrimination) ? » ENR : enregistrement. Les verbalisations des auditeurs sont écrites telles quelles, ce qui peut expliquer plusieurs redondances. Les numéros entre parenthèses renvoient au nombre de verbalisations identiques employées par plusieurs locuteurs. Les nombres dans la dernière ligne renvoient au nombre de qualificatifs de chaque catégorie

Toutes tâches confondues, 35 % de leurs verbalisations évoque des différences perçues de timbre (*plus ou moins clair, brillant, ou sourd, étouffé*), 23 % révèle des différences relatives à la prosodie, l'articulation ou la netteté des consonnes, 18 % l'intensité (*puissance, force ou atténuation de la voix*), 15 % évoque la qualité d'enregistrement perçue, 5 % la présence d'une production perçue comme plus ou moins lointaine, et 4 % pour la

présence de bruits parasites. Pour le chant uniquement, 12,3 % des verbalisations mentionnent aussi un vibrato perçu comme un autre indice de différences entre les deux productions d'une même paire, et moins d'occurrences relatives à l'intensité. La figure 10 détaille, tous types de masques confondus, la ventilation du pourcentage des catégorisations obtenues à partir de l'analyse des verbalisations libres des auditeurs répondant à la question *Sur quel(s) indice(s) vous êtes-vous appuyé.e pour différencier les deux phrases d'une paire écoutée au test de comparaison (discrimination) ?* Elle confirme que les indices en termes de timbre sont prépondérants pour différencier la voix avec et sans masque pour toutes les tâches de production, excepté pour la phrase produite par le locuteur masculin, pour laquelle l'intonation (catégorie *parole*) a davantage été inscrite comme verbalisation libre par les auditeurs, faisant écho à leurs commentaires spontanés évoqués plus haut.

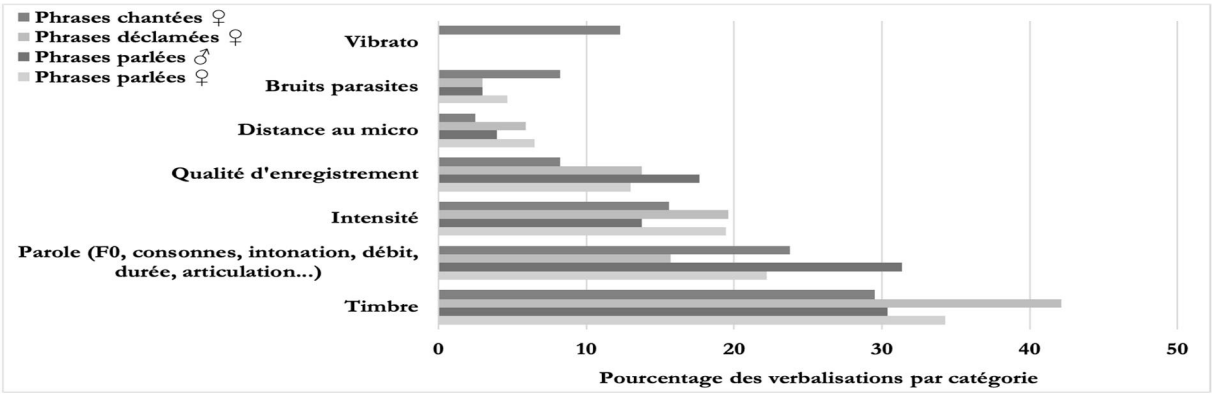

Figure 10. Pourcentage des catégories obtenues suite à la classification des verbalisations libres des auditeurs ayant répondu, pour chaque test de discrimination 2 à 5, à la question « Sur quel(s) indice(s) vous êtes-vous appuyé.e pour différencier les deux phrases d'une paire écoutée au test de comparaison (discrimination) ? »

4. Discussion

Cette étude exploratoire à partir de la production de deux locuteurs montre des effets différents du port de différents types de masques faciaux antiviraux sur la perception des voix et paroles produites avec ces dispositifs.

D'un côté, l'identification consonantique n'est pas perturbée pour les sujets (qui ont jugé ce tout premier test facile) ; elle est peu affectée par le port des différents types de masque, excepté le masque à fenêtre transparente dont la littérature a montré la médiocre performance acoustique, contrairement aux masques chirurgicaux en modalités auditive et audiovisuelle sans bruit de fond (Fecher et Watt, 2013, Keerstock *et al.*, 2020). Avec les masques à fenêtre transparente, les auditeurs étant moins

certains de leurs jugements, des consonnes voisées bilabiales (/b/), et, à un moindre degré, labiodentales (/v/) et dentales (/z/), présentent quelques difficultés d'identification, principalement des erreurs de lieu d'articulation, liées au filtrage des fréquences à partir de 2 000 Hz. Dans une recherche exploratoire analysant justement acoustiquement ces mêmes données (Pillot-Loiseau, Harmegnies, 2021), nous avions trouvé que le masque transparent présentait les distorsions acoustiques les plus significatives par rapport à la condition non masquée, notamment sous la forme d'une uniformité des valeurs du centre de gravité spectral[1] (plus élevé pour /p b/ et plus faible pour /t d k g f v/) ; la tendance est inversée pour ces consonnes concernant le coefficient d'asymétrie[2], moins positif (équivalent à une concentration d'énergie dans des bandes de fréquences légèrement plus élevées) pour les bilabiales. Ceci pourrait expliquer pourquoi les stimuli contenant /b/ sont identifiés à 14 % comme /d/ (figure 3). Nos résultats concernant le test 1 d'identification consonantique mentionnent aussi des degrés de certitude inférieurs, principalement, pour les consonnes bilabiales et labiodentales : il est probable que les consonnes mobilisant des mouvements labiaux occasionnent plus de distorsions acoustiques et perceptives que d'autres consonnes à cause du contact entre les lèvres et le masque facial plus fréquent pour ces segments. Ainsi, une consonne comme /b/ serait moins résistante à la perturbation externe occasionnée par le masque. Cohn *et al.* (2021) affirment d'ailleurs que les masques faciaux pourraient également modifier certains types d'articulations plus que d'autres, par exemple en empêchant la protrusion complète des lèvres pour les labiales.

[1] Le centre de gravité spectral, mesure exprimée en Hz, indique le lieu de l'axe des fréquences où le spectre fait apparaître la plus grande concentration énergétique. Il exprime ainsi la fréquence à laquelle l'énergie spectrale est prédominante : sa valeur sera par exemple plus élevée pour une fricative à timbre clair produite avec une petite cavité, comme le /s/, que pour une fricative à timbre plus sombre produite avec une plus grande cavité. Indice de position sur l'axe des Hertz, il est cependant insensible à la forme de la courbe de répartition des énergies sur l'axe fréquentiel.

[2] Le coefficient d'asymétrie est un nombre sans dimension exprimant le caractère plus ou moins symétrique de la courbe de répartition des énergies sur l'axe fréquentiel. Une valeur nulle exprime la parfaite symétrie de l'enveloppe (c'est-à-dire une concentration au centre du domaine spectral) alors que l'augmentation de l'indice, en valeur absolue, traduit une concentration graduellement plus marquée vers l'une des deux extrémités du domaine fréquentiel. Il est négatif pour une concentration vers les hautes fréquences au détriment des basses, et positif dans le cas contraire. Il peut donc être considéré, en première approche, comme une expression numérique globale, mais certes grossière, du timbre.

Il conviendrait cependant d'effectuer ces tests d'identification et d'autres tests d'intelligibilité à partir de phrases en présence d'un bruit de fond et en modalité audio-visuelle : le bruit en situation écologique diminue davantage l'intelligibilité de la parole, lorsque des indices visuels complémentaires bloqués par l'utilisation de masques faciaux jouent un rôle dans la communication (Magee *et al.*, 2020). Toutefois, comme l'indiquent Cohn *et al.* (2021), d'un point de vue théorique, l'information visuelle peut exercer des effets indépendants sur l'intelligibilité de la parole (par exemple, Babel, Russell, 2015) : il est donc important d'isoler d'abord les effets qui proviennent uniquement de l'acoustique, comme nous le faisons ici. D'un point de vue pratique, la communication orale en contexte de risque viral présente plusieurs situations d'écoute à visage masqué qui sont principalement auditives, même si, dans la communication ordinaire, la perception recourt également aux indices visuels.

De l'autre côté, il semble que, même si des travaux antérieurs ont montré que des mesures acoustiques de qualité vocale ne sont pas affectées dans les signaux acoustiques modifiés par plusieurs types de masques faciaux (Magee *et al.*, 2020), la perception de cette qualité de voix et de parole de signaux avec masque soit plus ou moins différenciée en fonction du type d'équipement facial porté, de l'individu produisant ces stimuli et du signal produit sans masque : les taux de discrimination sans masque / FFP2 et sans masque / masque à fenêtre transparente sont plus importants que pour les autres types de masques sauf pour le chant. Ces résultats relatifs aux pourcentages de discrimination sont renforcés par le fait que les degrés de certitude de nos auditeurs et leurs accords sont maximaux pour la discrimination entre la condition sans masque, et, respectivement celle avec les masques FFP2 et à fenêtre transparente, sauf en chant où les accords inter-auditeurs sont minimaux.

Les productions en voix chantée présentent un *pattern* particulier des réponses obtenues par rapport à la perception des productions parlées et déclamées, avec des différences entre la condition sans masque et celle avec masque perçues avec moins de certitude, comme l'ont confirmé les auditeurs dans leurs commentaires additionnels. Notre étude portant sur l'analyse acoustique de ces données (Pillot-Loiseau, Harmegnies, 2021) a en effet montré que les masques FFP2 et surtout les masques à fenêtre

transparente sont les plus filtrants, quelle que soit la tâche et le locuteur, sauf pour la première session de productions chantées par la locutrice : d'après cette étude, le masque transparent montre des antiformants autour de 3 000 Hz pour le locuteur masculin et autour de 3 800 Hz pour la locutrice féminine, surtout en voix déclamée. Cette étude montre aussi que le masque chanteur est le moins filtrant et que le masque en tissu porté par le locuteur masculin est plus filtrant que celui porté par le locuteur féminin. Dans les productions vocales chantées par la locutrice, les fréquences correspondant à celles du formant du chanteur, entre 2 000 Hz et 4 000 Hz (entre autres, Sundberg, 1999), sont atténuées par le grand masque et le masque transparent, puis par le masque FFP2 et, dans une moindre mesure, par le masque chanteur. Cette atténuation à ces fréquences explique pourquoi les sujets ont essentiellement qualifié les différences entendues en termes de timbre, spécialement de clarté, résonance et volume, qui sont des corrélats perceptifs connus d'une présence d'énergie importante au niveau des fréquences aigues, en particulier entre 2000Hz et 4000Hz, là où se localise le formant du chanteur : celui-ci est connu pour être un corrélat acoustique de la portée de la voix (Sundberg, 1999). Ces verbalisations de nos sujets en termes de timbre et résonance (tableau 1) sont en accord avec les résultats des tests de discrimination en aveugle de paires de phrases chantées enregistrées par une soprano sans et avec masque, dans l'étude de Oren *et al.* (2021) : sur une échelle visuelle analogique représentant la résonance, délimitée par les deux phrases chantées d'une même paire (avec et sans masque donc), quatre auditeurs expérimentés en pédagogie ou rééducation vocale devaient indiquer le degré de résonance en déplaçant le curseur vers l'échantillon de parole chantée le plus résonant pour eux. Les résultats ont montré que les phrases chantées avec les masques chirurgicaux puis FFP1 étaient perçues comme plus préjudiciables à la perception de la résonance et de la qualité générale du chant. Le masque FFP1 a eu le plus grand impact sur la réduction de la qualité de résonance du chant.

D'ailleurs, l'étude exploratoire susmentionnée (Pillot-Loiseau, Harmegnies, 2021) mentionne, dans les résultats d'une enquête menée auprès de 303 utilisateurs de masques faciaux, le fait que ces utilisateurs aient insisté dans leurs ressentis sur la perte de puissance de la voix et la nécessité de répéter plus souvent. Karagkouni (2021), dans une étude

transversale d'observation, menée en distribuant un questionnaire en ligne auquel ont répondu 143 personnes portant un masque facial, a en effet observé que ses sujets enquêtés ont en particulier déclaré qu'ils devaient parler plus fort et qu'ils avaient remarqué des altérations des caractéristiques perceptives de leur voix, l'enrouement et le volume étant les plus fréquemment affectés.

Les productions en voix chantée, tant sur les plans perceptif qu'acoustique, semblent donc plus résistantes à la perturbation occasionnée par le masque que la parole conversationnelle ou déclamée par la locutrice. Cela suggère un possible réajustement de la production vocale chantée en cas de perturbation plus importante par ces équipements faciaux : à l'image de ce que Cohn *et al.* (2021) ont constaté comme différences dans leurs résultats relatifs aux paroles conversationnelle, claire et émotionnelle, les changements de style de parole (ici parlée *vs* chantée) peuvent produire plus ou moins d'impact sur l'intelligibilité ou même sur la nature de la production avec masque par rapport à celle sans masque, par les auditeurs. Notre étude démontre ici que, dans la voix chantée, c'est la locutrice elle-même, et les adaptations en temps réel qu'elle réalise avec les masques pour les auditeurs, qui contribue, dans sa production chantée, de manière essentielle à la robustesse de la parole humaine dans ce contexte. Cela est d'ailleurs en accord avec les conclusions d'Oren *et al.* (2021), qui reconnaissent que les chanteurs professionnels pourraient adapter leur chant au masque, permettant ainsi à leur voix d'être perçue comme étant de même qualité par un auditeur, indépendamment du type de masque utilisé. Cependant, d'autres études ont souligné les potentiels risques vocaux pouvant résulter de la tentative du locuteur portant un masque d'ajuster ses caractéristiques de phonation pour rendre sa voix plus claire (Gama *et al.*, 2021).

5. Conclusion

Les travaux présentés dans cette contribution sont la première étude perceptive sur les effets de six types de masques sur les voix parlée, déclamée et chantée (énoncés entiers) et la parole (consonnes constrictives et occlusives) en français.

Ces données perceptives sont en lien avec certains effets acoustiques constatés dans des travaux précédents sur les mêmes données (Pillot-

Loiseau, Harmegnies, 2021) ; ils montrent en particulier qu'en modalité auditive calme à partir d'enregistrements originaux, les masques à fenêtre transparente et FFP2 sont les mieux discriminés en parole par rapport à celle-ci sans masque, que les masques à fenêtre transparente altèrent l'identification de certaines consonnes, notamment bilabiales, et que ces deux types de masques provoquent une diminution du timbre, du volume et de la résonance, surtout en parole. Les masques chirurgical et chanteur présentent moins d'impact perceptif, tout comme leurs moindres impacts acoustiques sur la parole et le chant, précédemment démontrés (Pillot-Loiseau, Harmegnies, 2021). De la sorte, le port de ces deux derniers types de dispositifs pourra être prioritairement recommandé pour toute situation de communication, mais aussi pour les chanteurs.

Cette étude a également montré la variation de ces effets perceptifs en fonction de la tâche (chanté vs parlé), tout comme les effets acoustiques (Pillot-Loiseau, Harmegnies, 2021), mais aussi en fonction du profil des auditeurs concernant leur pratique vocale. À notre connaissance, cette dernière dimension n'a pas été à ce jour considérée dans d'autres travaux.

Les masques, et en particulier ceux à fenêtre transparente et les masques FFP2, constituent une source externe de perturbation du système de production de la parole (Sock, 2001), mais dans une moindre mesure en chant.

Bien entendu, notre étude est limitée de par son caractère exploratoire : elle présente un petit nombre de locuteurs car il s'agissait d'une recherche menée sous confinement sans accès aux laboratoires de recherche. La variabilité des résultats obtenus auprès d'individus plus nombreux à produire ces stimuli n'a pas été explorée même si elle se retrouve dans la perception (cf. accords inter-auditeurs selon les types de masques, les individus et les tâches). Enfin la perception ici investiguée est unimodale (auditive) en environnement calme, contrairement à plusieurs situations communicationnelles écologiques (cours, commerce, etc.).

Remerciements

Ce travail a été soutenu par le Partenariat Hubert Curien (PHC) TOURNESOL n°46241WH et par le Laboratoire d'Excellence (LabEx) *Empirical Foundations of Linguistics* (EFL) n°ANR-10-LABX-0083. Il contribue à l'IdEx Université de Paris (ANR-18-IDEX-0001).

Références bibliographiques

BABEL, M., RUSSELL, J., Expectations and speech intelligibility, *THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA,* 2015, **137(5),** 2823–2833. https://doi.org/10.1121/1.4919317

BOERSMA, P., WEENINK. D., Praat: doing phonetics by computer, 1992-2021, V. 6.1.16 ret. 2 Feb. 2021 from http://www.praat.org/

COHN, M., PYCHA, A., ZELLOU, G., Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech, *COGNITION*, 2021, **210,** 104570. https://doi.org/10.1016/j.cognition.2020.104570

COREY, R. M., JONES, U., SINGER, A. C., Acoustic effects of medical, cloth, and transparent face masks on speech signals, *THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, 2020, **148(4),** 2371-2375. https://doi.org/10.1121/10.0002279

FECHER, N., WATT, D., Effects of forensically-realistic facial concealment on auditory-visual consonant recognition in quiet and noise conditions, in Actes de *Auditory-Visual Speech Processing (AVSP) 2013*. 29 août au 2 septembre 2013, 1-6, https://www.isca-speech.org/archive_v0/avsp13/papers/av13_081.pdf

GAMA, R., CASTRO, M. E., VAN LITH-BIJL, J. T., DESUTER, G., Does the wearing of masks change voice and speech parameters? *EUROPEAN ARCHIVES OF OTO-RHINO-LARYNGOLOGY*, 2022, **279**, 1701–1708. https://doi.org/10.1007/s00405-021-07086-9

KARAGKOUNI, O., The effects of the use of protective face mask on the voice and its relation to self-perceived voice changes, *JOURNAL OF VOICE*, 2021, in Press, publié en ligne à : https://doi.org/10.1016/j.jvoice.2021.04.014

KEERSTOCK, S., MEEMANN, K., RANSOM, S. M., SMILJANIC, R., Effects of face masks and speaking style on audio-visual speech perception and memory, *THE 179TH MEETING OF THE ACOUSTICAL SOCIETY OF AMERICA*, Virtual, 2020, **148(4),** 2747. https://doi.org/10.1121/1.5147635

KREIMAN, J., GERRATT, B.R., Listener experience and perception of voice quality, *JOURNAL OF SPEECH AND HEARING RESEARCH,* 1990, **33(1),** 103-115. https://doi.org/10.1044/jshr.3301.103

MAGEE, M., LEWIS, C., NOFFS, G., REECE, H., CHAN, J. C., ZAGA, C. J., PAYNTER, C., BIRCHALL, O., ROJAS AZOCAR, S., EDIRIWEERA, A., KENYON, K., CAVERLÉ, M. W., SCHULTZ, B. G., VOGEL, A. P., Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols, *THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, 2020, **148(6),** 3562-3568. https://doi.org/10.1121/10.0002873

NOSULENKO, V., SAMOYLENKO, E.S., Approche systémique de l'analyse des verbalisations dans le cadre de l'étude des processus perceptifs et cognitifs, *SOCIAL SCIENCE INFORMATION,* 1997, **36(2),** 223-261. https://doi.org/10.1177/053901897036002002

OREN, L., ROLLINS, M., GUTMARK, E., HOWELL, R., How face masks affect acoustic and auditory perceptual characteristics of the singing voice, *JOURNAL OF VOICE,* 2021, in Press, publié en ligne à : https://doi.org/10.1016 /j.jvoice.2021.02.028

PILLOT-LOISEAU, C., HARMEGNIES, B., Effect of Different Face Masks on Speech and Singing: Self-Perception and Acoustic Analysis. *ANNALI DI CA'FOSCARI. SERIE OCCIDENTALE*, 2021, **55**, 9-28. http://doi.org/10.30687/AnnOc/2499-1562/2021/09/012

RAHNE, T., FRÖHLICH, L., PLONTKE, S., WAGNER, L., Influence of surgical and N95 face masks on speech perception and listening effort in noise. *PLOS ONE*, 2021, **16(7),** e0253874. https://doi.org/10.1371/journal.pone.0253874

SOCK, R., La Théorie de la Viabilité en production-perception de la parole, in KELLER, D., DURAFOUR, J.-P., BONNOT, J.-F, SOCK, R. (éds.), *Psychologie et Sciences Humaines,* Liège, Mardaga, 2001, 285-316.

SUNDBERG, J. The perception of singing, in DEUTSCH, D (éds.) *The Psychology of Music,* San Diego, Academic Press 1999, 171-214). https://doi.org/10.1016/B978-012213564-4/50007-X.

Claire PILLOT-LOISEAU est Maître de Conférences HDR en phonétique à l'Université de la Sorbonne Nouvelle (Paris, France), membre du *Laboratoire de Phonétique et Phonologie*, et orthophoniste. Ses recherches caractérisent les variations de la parole atypique et étudient le lien entre la voix et la parole. Elles portent sur la phonétique et la pédagogie de la prononciation du français langue étrangère, sur la phonétique clinique, et la typologie de la voix et de la parole en chant dans des techniques vocales spécifiques. Depuis son doctorat en phonétique sur le chant lyrique, elle a développé plusieurs méthodes d'analyse acoustique de la voix parlée et chantée de sujets parlant plusieurs langues.

Bernard HARMEGNIES a dirigé durant plusieurs décennies le Laboratoire de phonétique de l'Université de l'UMONS. Ses recherches ciblent les phénomènes de communication à l'oral relevant des sciences de la parole. Il a consacré de nombreuses publications à la qualité vocale et aux styles de parole. Il a fondé, à l'UMONS, l'Institut de Recherche en Sciences et Technologies du Langage. Il a assuré diverses responsabilités dans des institutions de recherche. Aujourd'hui Professeur émérite, il poursuit diverses activités académiques (UMONS, ULB) et préside le Centre International de Phonétique Appliquée. Il préside le Comité d'avis du Conseil Supérieur d'Intégrité Scientifique de Belgique francophone.