



HAL
open science

Frontières entre la perception de la voix normophonique et pathologique chez des auditeurs naïfs

Amelia Pettirossi, Nicolas Audibert, Lise Crevier-Buchman

► To cite this version:

Amelia Pettirossi, Nicolas Audibert, Lise Crevier-Buchman. Frontières entre la perception de la voix normophonique et pathologique chez des auditeurs naïfs. 35èmes Journées d'Études sur la Parole (JEP 2024) 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2024) 26ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2024), Jul 2024, Toulouse, France. pp.401-411. hal-04623091

HAL Id: hal-04623091

<https://inria.hal.science/hal-04623091>

Submitted on 1 Jul 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Frontières entre la perception de la voix normophonique et pathologique chez des auditeurs naïfs

Amelia Pettirossi¹ Nicolas Audibert¹ Lise Crevier-Buchman^{1,2}

(1) Laboratoire de Phonétique et Phonologie (CNRS & Sorbonne Nouvelle), Paris, France

(2) Hôpital Foch : Service de Laryngologie Phoniatre, Suresnes, France

amelia.pettirossi@sorbonne-nouvelle.fr, nicolas.audibert@sorbonne-nouvelle.fr, lise.buchman1@gmail.com

RÉSUMÉ

Nous étudions l'hétéroévaluation de la qualité de voix chez une population de femmes francophones comprenant des professionnelles de la voix et des patientes recrutées lors d'une consultation de phoniatrie. Des extraits de parole ont été évalués par un expert à l'aide du GRBAS et par deux groupes de 30 et 40 auditeurs naïfs avec des protocoles perceptifs destinés à évaluer la qualité de voix globale. Ces évaluations ont été mises en correspondance afin d'observer de potentielles corrélations entre la perception de la pathologie vocale chez les auditeurs naïfs et les paramètres du GRBAS expert. Les voix perçues comme plus pathologiques par les auditeurs naïfs sont bien associées à un grade global plus élevé, avec une influence prépondérante de la sévérité du serrage vocal. Les indices de raucité et de souffle sont moins liés à la perception par les naïfs des voix comme pathologiques.

ABSTRACT

Boundaries between the perception of normophonic and pathological voice by naive listeners

We study hetero evaluation of voice quality in a population of French female speakers including voice professionals and patients recruited during a phoniatic consultation. Speech samples were evaluated by an expert using the GRBAS and by two groups of 30 and 40 naïve listeners with perceptual protocols to assess overall voice quality. These assessments were matched to observe potential correlations between perception of vocal pathology in naive listeners and expert GRBAS parameters. Voices perceived as more pathological by naive listeners are well associated with a higher overall grade, with a predominant influence of vocal strain severity. The roughness and breathiness dimensions are less related to naive listeners' perception of voices as pathological.

MOTS-CLÉS : voix, dysphonie, hétéro-perception, auditeurs experts, auditeurs naïfs

KEYWORDS : voice, dysphonia, hetero-perception, expert listeners, naive listeners

1 Introduction

Les normes culturelles peuvent faire varier différents aspects de l'évaluation de la qualité de voix, même ceux supposés s'appuyer sur des références solides comme l'évaluation perceptive experte des dysphonies. Il a été observé que des professionnels italiens et français, évaluant des locuteurs dysphoniques de ces deux nationalités, sont en accord pour ce qui est du grade de sévérité général de la dysphonie, mais pas de la raucité, sous-cotée par les experts italophones en comparaison avec

les francophones (Ghio et al., 2011 ; 2015). Cet écart semble lié au fait que l'auditeur, bien qu'expert, est directement influencé par ses représentations sociales, c'est-à-dire à ce qui est localement et momentanément validé au sein d'une culture donnée (Rouquette et al., 1998). Ces représentations sociales viennent s'ajouter à une expérience personnelle et une éventuelle formation pour venir créer un référent interne propre à chaque auditeur.

Afin de quantifier les altérations de la qualité de voix, il existe des outils acoustiques composites tel que l'AVQI (Acoustic Voice Quality Index) mais également des outils perceptifs qui sont en général considérés comme la référence (*gold standard*) du diagnostic de la dysphonie. Bien que des revues de littérature mettent en évidence qu'il existe une grande variété de protocoles perceptifs pour l'hétéroévaluation de la qualité de voix pathologique, le GRBAS de Hirano (1981) reste l'un des protocoles les plus utilisés en recherche et en clinique (Kreiman et al., 1993 ; De Bodt et al., 1996 ; Suhail et al., 2016 ; Schuering et al., 2021). Le GRBAS permet de quantifier le grade général de dysphonie (G), la raucité (R), le souffle (B), l'asthénie (A) et le serrage vocal (S). Chaque critère pouvant être coté de 0 (aucun trouble) à 3 (trouble sévère).

Si nous savons que de manière générale les voix dysphoniques induisent des jugements plus négatifs que les voix saines sur différents aspects comme l'attractivité physique (Blood et al., 1979) ou la personnalité (Amir et al., 2013) il est aussi vrai que tout ce qui est catégorisé comme pathologique pour un expert n'est pas forcément associé à une représentation négative chez les auditeurs naïfs. Il a déjà été démontré que dans une population francophone, une légère raucité chez l'homme peut être considérée comme séduisante pour des auditrices alors que cette même raucité est considérée comme pathologique par un expert (Barkat-Defradas & al, 2012). Cette conclusion a également été vérifiée pour les femmes francophones, les locutrices ayant une raucité légère sont évaluées de manière plus positive par des auditeurs des deux sexes (Pettirossi et al., 2020). Ces observations sont facilement explicables par l'idée selon laquelle « *What sounds beautiful is good.* », largement mise en évidence dans les études sur le jugement vocal (Zuckerman & al, 1988).

Il semble nécessaire de comprendre où se place la limite entre une voix qui sera perçue comme pathologique par un auditeur naïf et qui engendrera une image détériorée du locuteur, d'une voix normophonique. Il pourrait être pertinent de prendre en compte qu'en l'absence de plainte physique ou fonctionnel, tous les paramètres perceptifs utilisés par les experts du domaine médical ou paramédical n'évoquent pas nécessairement la pathologie chez les naïfs.

Pour autant, comme susmentionné, l'évaluation des auditeurs naïfs peut être sévère et provoquer certains handicaps sociaux. Il a déjà été mis en évidence que des auditeurs ayant reçu des fiches explicatives visant à les sensibiliser aux troubles vocaux évaluent aussi sévèrement, malgré ce dispositif préalable, des locutrices atteintes de dysphonie et d'hyper-nasalité qu'un deuxième groupe d'auditeurs naïfs maintenu volontairement dans la méconnaissance de la pathologie vocale (Lallh et al., 2000). Les locutrices atteintes de troubles vocaux sont catégorisées comme peu amicales, ennuyeuses et peu attractives, ce qui peut évidemment être pénalisant dans divers contextes sociaux comme les entretiens d'embauche ou même les relations amicales. L'une des principales conclusions de ces travaux est donc que les cliniciens pourraient avoir besoin de faire de la prévention auprès de leurs patients quant aux attitudes sociales négatives des interlocuteurs.

Nous savons également que les auditeurs naïfs sont peu sensibles aux dysphonies légères (G1, G2), et n'obtiennent respectivement que 49% et 38% de bonnes catégorisations lors de l'évaluation de ces voix sur le paramètre G du GRBAS en comparaison avec une évaluation experte (Ghio et al.,

2011). A l'inverse, les dysphonies sévères (G3) obtiennent 68% de bonnes réponses et 86% pour les voix normophoniques (G0), ces résultats sont vérifiables même après un entraînement des naïfs.

D'autre part, certaines études traitent de la différence de perception de l'altération de la qualité de voix entre auditeurs experts (orthophonistes et oto-rhino-laryngologistes) et naïfs en termes d'accord inter-juges. Sans étonnement nous observons généralement de meilleurs degrés d'accord chez les experts (Helou et al., 2010 ; Misono et al., 2012) mais parfois un accord inter-juges aussi élevé pour les experts que les naïfs (Eadie et al., 2010). Ces approches mettent finalement en évidence qu'il y aurait un référent interne plus similaire chez les experts. Évidemment, il est légitime de penser que les naïfs, qui ont une expérience personnelle n'incluant pas de formation commune en santé, auront un référent interne plus variable que les experts.

Dans cet article, nous tenterons de répondre à la question suivante : où se trouve la frontière, aussi ténue soit-elle, entre ce qui est perçu comme une qualité de voix normophonique ou pathologique par un panel d'auditeurs naïfs ? Nous nous baserons sur la perception générale de la qualité de voix par des naïfs par rapport à la cotation des paramètres perceptifs du GRBAS réalisée par une médecin phoniatre à travers deux groupes de locutrices, l'un ayant été recruté dans le cadre d'une consultation de phoniatry et l'autre étant composé de locutrices professionnelles de la voix volontaires.

2 Méthodologie

2.1 Locutrices, corpus et prises de données

Nous avons ici deux panels distincts. Nous nous référerons à notre premier panel comme « groupe patientes », il s'agit ici de 10 femmes francophones (29-78 ans) recrutées dans le cadre d'une consultation phoniatry avec laryngostroboscopie. Leurs métiers peuvent être variés, bien que nous retrouvions 9 professionnelles de la voix dont 5 professeuses des écoles. Les critères d'inclusion de ce panel comprenaient le fait d'être diagnostiquées d'une dysphonie lors de la consultation et de ne pas avoir été opéré du larynx précédemment, ni avoir subi de traitement radiochimiothérapique. Les atteintes du groupe patientes sont multiples : kyste, nodule, laryngite, granulome, œdème de Reinke, reflux, ou encore paralysie récurrentielle.

Nous appellerons le deuxième panel « groupe PE » car il est constitué de 61 femmes francophones, professeuses des écoles (PE) en activité (23-61 ans). Il s'agit ici uniquement de volontaires tout-venant. Pour autant, nous nous attendons à recruter des locutrices avec des dysphonies dysfonctionnelles car d'après une étude menée pour l'INSERM la prévalence dans cette population, majoritairement féminine, est très élevée (Autesserre et al., 2006). En 2017, une enquête menée sur 709 femmes professeuses des écoles travaillant en France a mis en évidence que 80% de la population auto-déclare des troubles vocaux alors que seulement 18% des PE ont déjà consulté un orthophoniste ou un ORL (Pettirossi, 2021). Les critères d'inclusion de ce panel comprenaient également de ne pas avoir été opéré du larynx précédemment, ni avoir subi de traitement radiochimiothérapique.

Toutes les participantes ont été enregistrées à partir de la station Computerized Speech Lab 4500 de KayPENTAX avec un micro-casque AKG C 410 positionné à environ 5 cm des lèvres. Les

différentes évaluations perceptives de cette étude ont été réalisées à partir de phrases lues ou de la lecture du texte « La bise et le soleil ».

2.2 Hétéroévaluation de la qualité de voix

Dans le cadre de notre hétéroévaluation experte de la qualité de voix, une même phoniatre a réalisé une cotation GRBAS en direct pour le panel « patientes » et en différé avec une interface sur ordinateur réalisée avec Praat (Boersma et al., 2023) pour le panel « PE », dans les deux cas à partir d'un extrait de parole lue. Nous observons, des dysphonies beaucoup plus sévères dans notre groupe patientes et de nombreuses dysphonies légères dans notre groupe PE (Table 1).

| Cotation experte | Groupe « patientes » (n = 10) | | | | Groupe « PE » (n=61) | | | |
|------------------|-------------------------------|---|---|---|----------------------|----|---|---|
| | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| G (n) | 0 | 2 | 3 | 5 | 37 | 22 | 2 | 0 |
| R (n) | 3 | 3 | 1 | 3 | 41 | 19 | 1 | 0 |
| B (n) | 2 | 5 | 3 | 0 | 52 | 9 | 0 | 0 |
| A (n) | 9 | 0 | 1 | 0 | 61 | 0 | 0 | 0 |
| S (n) | 4 | 2 | 3 | 1 | 57 | 4 | 0 | 0 |

TABLE 1 : Récapitulatif du nombre de locutrices dans chaque panel cotée 0, 1, 2 ou 3 sur les différents paramètres perceptifs du GRBAS par une experte phoniatre

Quant à eux, les auditeurs naïfs ont évalué la qualité de voix des locutrices sur des extraits de lecture de « La bise et le soleil » à l'aide d'interfaces générées avec le logiciel Praat (Boersma et al., 2023) avec des variations mineures dans les protocoles d'évaluation appliqués aux deux panels de locutrices. Nous comptons 30 auditeurs pour l'évaluation du panel patientes et 40 pour le panel PE parmi lesquels 15 auditeurs sont communs aux deux tâches, avec environ deux ans d'écart entre les deux évaluations.

Pour le panel patientes, les auditeurs ont dû répondre à la question « Pensez-vous que cette personne a des problèmes de voix ? » à l'aide de 4 boutons « Absolument pas », « Probablement pas », « Probablement », « Absolument ». À la fin de cette expérimentation les réponses ont été cotées 0 pour « Absolument pas », jusqu'à 3 pour « Absolument ».

Enfin, pour le panel PE, les évaluations ont été réalisées sur une échelle sémantique différentielle comprenant 5 échelons avec les mentions « Aucun trouble vocal » et « trouble vocal sévère » à chaque extrémité. Les réponses ont été cotées de 0 pour « Aucun trouble vocal » à 4 pour « trouble vocal sévère ».

3 Résultats

Nous réalisons ici des corrélations de Spearman pour comparer les évaluations de la qualité de voix globales réalisées par les auditeurs naïfs avec la cotation experte du GRBAS mais également entre le grade de dysphonie (G) et les autres dimensions (R, B, A et S) de l'évaluation experte (Table 2).

| Corrélation avec les paramètres GRBAS | G expert : Coefficient ρ (Valeur-p) | Naïfs « patientes » : Coefficient ρ (Valeur-p) | Naïf « PE » : Coefficient ρ (Valeur-p) |
|---------------------------------------|--|---|---|
| G | | 0.822 (0.003) | 0.630 (< 0.0001) |
| R | 0.786 (< 0.0001) | 0.227 (0.528) | 0.579 (< 0.0001) |
| B | 0.653 (< 0.0001) | 0.508 (0.134) | 0.391 (0.002) |
| A | 0.180 (0.132) | -0.175 (0.630) | n/a |
| S | 0.605 (< 0.0001) | 0.892 (0.0003) | 0.304 (0.02) |

TABLE 2 : Corrélation et valeur-p entre le Grade expert et les autres dimensions du GRBAS

Nous effectuons également des comparaisons entre ces corrélations expertes et naïves pour les dimensions perceptives R, B, A et S (Figure 1). Ces comparaisons sont effectuées à l'aide du package *cocor* (Diedenhofen et al., 2015) implémenté dans R (R Core Team, 2024), avec la méthode de Meng et al. (1992).

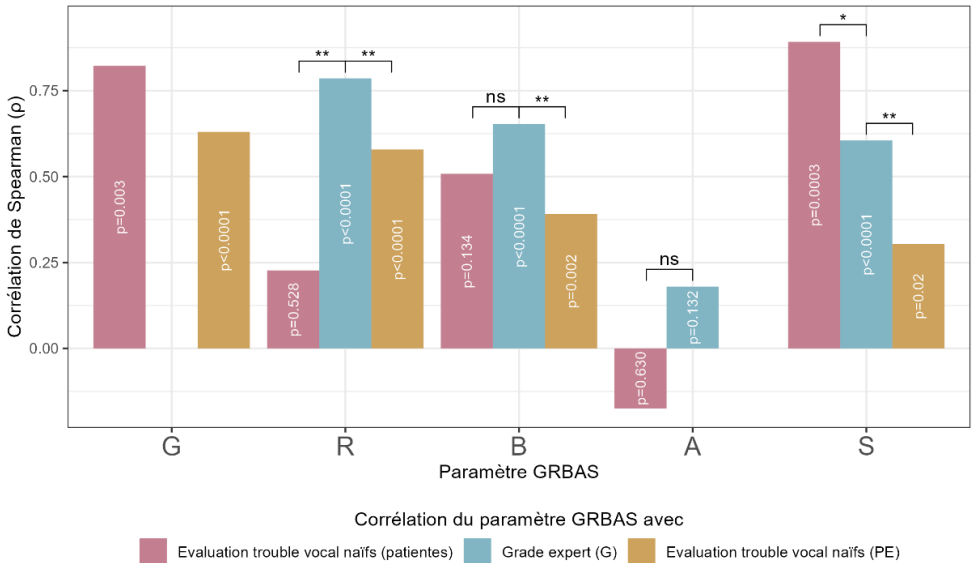


FIGURE 1 : Corrélation et valeur-p entre le Grade expert et les autres dimensions du GRBAS toutes locutrices confondues (bleu) et entre le GRBAS et le score moyen de l'évaluation globale de la qualité de voix naïve pour chaque groupe de locutrices (rose et ocre). Les accolades indiquent la significativité de la différence entre corrélations au sens de Meng et al. (1992).

Tout d'abord, nous obtenons une corrélation élevée entre le grade général (G) coté par l'expert et la raucité (R), des corrélations moyennes entre G et le souffle (B) et le serrage vocal (S) mais aussi une absence de corrélation entre G et l'asthénie (A). Si les auditeurs naïfs sont sensibles aux mêmes dimensions perceptives que les experts, ils devraient donc être plus influencés par la raucité, et dans une moindre mesure par le souffle, le serrage vocal et enfin l'asthénie.

Les corrélations entre le GRBAS expert et l'évaluation globale de la qualité de voix par les naïfs sont toutes plus modérées pour notre groupe PE. Ce résultat n'est pas étonnant car les dysphonies sont beaucoup plus légères que dans le groupe patientes. En revanche, il semble y avoir une sensibilité des auditeurs naïfs quant à « l'impression générale » d'altération des voix car le grade de sévérité coté par l'expert est fortement corrélé avec l'évaluation globale de la qualité de voix faite par les naïfs pour les patientes et modérément pour le groupe PE. Notons également des divergences presque systématiques entre les corrélations des dimensions R, B, A et S expert avec le G expert et l'évaluation globale de la qualité de voix par les auditeurs naïfs, la plupart des différences entre naïfs et expert étant significatives. Cela suggère que les dimensions perceptives prises en compte par les naïfs et les experts pour évaluer la pathologie vocale seraient foncièrement différentes.

Tout d'abord, pour ce qui est de la raucité, la corrélation avec l'évaluation globale de la qualité de voix réalisée par les naïfs est faible pour le groupe patiente et très modérée pour le groupe PE. Ainsi, la raucité semble plus impacter la perception de la qualité de voix globale dans notre population tout-venant comprenant des locutrices témoins comme des dysphonies légères plutôt que dans notre panel de patientes aux dysphonies plus sévères. Nous observons également une différence significative entre les corrélations experte et naïves pour le groupe patiente ($z=-3.071$, $p=0.002$), comme le groupe PE ($z=-2.809$, $p=0.005$). Ce résultat suggère que la raucité est plus prise en compte par les experts que par les naïfs pour catégoriser une voix comme pathologique.

La présence de souffle ne semble que très peu corrélée avec la perception de la pathologie vocale par les naïfs dans nos deux groupes de locutrices. Pour ce qui est de la comparaison entre les corrélations experte et l'évaluation naïve, nous observons une différence significative chez les PE ($z=-2.852$, $p=0.004$), mais pas chez les patientes ($z=-0.812$, $p=0.417$). Bien que ce résultat soit plus mitigé, l'expert semble plus influencé par le souffle que les naïfs pour évaluer la dysphonie.

Nous observons une absence totale de corrélation entre la perception d'une asthénie par l'expert et de la qualité de voix par les auditeurs naïfs. Rappelons tout de même que l'asthénie n'est présente chez aucune de nos locutrices du groupe PE et qu'une seule du groupe patiente en est atteinte. Cela limite grandement la portée de ce résultat, d'autant plus que l'asthénie est souvent le paramètre du GRBAS pour lequel l'accord inter-juges est le plus faible même chez des experts (Hidaka et al., 2022; Yamaguchi et al., 2003; Sellars et al., 2009). La comparaison entre les corrélations experte et naïves ne laisse apparaître aucune différence significative chez les patientes ($z=-1.568$, $p=0.117$).

Quant au serrage vocal, il semble toujours assez fortement corrélé à la perception de la pathologie vocale chez les auditeurs naïfs. En effet, dans notre groupe patiente la corrélation de cette dimension avec l'évaluation de la qualité de voix globale des naïfs est élevée et dépasse même celle du grade de sévérité général défini par l'expert. De plus, la différence entre les corrélations experte et naïve est significative pour le groupe patientes ($z=2.240$, $p=0.025$) laissant entendre que le serrage vocal est plus considéré comme marqueur de voix pathologique par les naïfs qu'il ne l'est par l'expert. Étonnamment, une corrélation plus faible de l'évaluation naïve du trouble vocal avec le serrage vocal est observée pour le groupe PE, à un niveau significativement inférieur à la corrélation experte ($z=3.095$, $p=0.002$). Ce résultat peut néanmoins s'expliquer par le fait que nous n'avons que quatre locutrices avec un serrage vocal léger dans ce panel.

Enfin, nous notons un accord inter-juges modéré pour l'évaluation naïve de la qualité de voix globale des locutrices patientes (corrélation intraclasse ICC de 0.633) alors que ce même accord est faible pour le groupe PE (ICC de 0.201). Évidemment, ce résultat est attendu car des locutrices avec des

troubles dysphoniques plus sévères se trouvent dans le groupe patientes contrairement au groupe PE qui comprend des dysphoniques légères ainsi que des locutrices normophoniques.

4 Discussion et conclusion

Nos résultats confirment que la perception de la dysphonie et ce qui la caractérise n'est pas uniforme. En effet, la délicate limite entre la perception de la voix normophonique et de la voix pathologique a déjà été régulièrement mise en avant et commence évidemment par une simple question : qu'est-ce que la normophonie, qu'est-ce qu'une voix « normale » ? Nous ne sommes pas en mesure d'évaluer ce qui constitue une « déviation » sans une norme supposée, ce qui est loin d'être aisé puisqu'il n'y a pas de consensus sur la définition même de la qualité de voix (Kreiman et al., 2011). Une étude récente a mis en lumière que des auditeurs invités à classer 20 voix de femmes de la plus « normale » à la plus « non-normale » montrent un degré d'accord limité pour ce qui est de classer les voix « anormales » et un degré d'accord encore plus faible pour les voix « normales » (Kreiman et al., 2020). Dans cette étude, parmi les mesures acoustiques prises en compte, la fréquence fondamentale (f_0) ainsi que le premier et second formant sont celles qui prédisent le mieux l'évaluation des auditeurs et non des paramètres plus spécifiques à la qualité de voix comme les mesures d'apériodicités ou de différence d'amplitude entre les premières harmoniques.

Il est important de prendre en compte qu'une voix peut être altérée par de nombreux biais comme un souffle, une raucité, une hyper ou encore hypo-fonctionnalité et que ces dimensions ne sont pas nécessairement indépendantes les unes des autres (Ghio et al., 2011). Encore une fois, la raucité semble être une dimension problématique puisqu'il a déjà été mis en lumière que son évaluation influence grandement la cotation du souffle, de manière générale, une forte interdépendance perceptive semble lier ces deux dimensions vocales (Kreiman et al., 1994). La littérature indique aussi que la perception de la pathologie n'est pas nécessairement constante, même lorsqu'elle est réalisée par des experts. En effet, l'évaluation de la raucité peut être catégorisée avec toutes les cotations possibles pour une même voix (Kreiman et al., 1993). La difficulté à isoler les dimensions perceptuelles est une des raisons pour lesquelles nous ne retrouvons pas un très bon accord entre les auditeurs sur la perception de la qualité de voix. A l'inverse, il existe un meilleur consensus sur la qualité de voix dans un corpus composé de voix de synthèse ne variant que par la f_0 par rapport à des échantillons de voix naturelles (Kreiman et al., 2000). Il semble même que parmi diverses mesures acoustiques objectives, seule la f_0 soit un indice robuste et presque unanimement utilisé par les auditeurs lors de l'évaluation de la qualité de voix d'un locuteur (Kreiman et al., 2020, 1992).

Cette grande variabilité dans l'évaluation perceptive de la qualité de voix semble être une conséquence de l'image que chacun se fait de ce qui est pathologique selon son expérience et sa culture. L'accord inter-juge modéré mis en évidence dans nos résultats même sur un panel de locutrices avec des dysphonies sévères en est une preuve. De plus, cette représentation n'est pas non plus uniforme parmi les experts. En effet, l'idée même d'un « espace perceptuel commun » a été remis en question par les divergences observées entre auditeurs et entre études (Kreiman et al., 1996).

Selon nos résultats, pour catégoriser une voix comme pathologique, les auditeurs naïfs sont particulièrement sensibles au-delà du grade général de dysphonie (G) à la perception du serrage vocal (S). Nous avons par ailleurs mis en évidence dans le cadre de travaux non-publiés sur le groupe patientes qu'il y a une forte corrélation positive entre la sévérité du serrage vocal d'une voix et le fait que les auditeurs naïfs la considèrent comme étant plus dérangement à écouter. Les voix avec un

fort serrage vocal sont également connues pour entraîner un plus gros effort d'écoute chez des auditeurs naïfs (Farahani et al., 2020). Ce retour d'impression « dérangent » ou « fatigant » pourrait provenir du fait que le serrage vocal donne l'impression de forcer sur la voix, une altération du comportement vocal qui serait perçue plus facilement et plus négativement par les naïfs.

En revanche, nous obtenons des corrélations très modérées à faibles entre la catégorisation d'une voix comme étant pathologique et la raucité (R) mais aussi le souffle (B). L'expert semble plus influencé par ces deux dimensions vocales que les naïfs pour évaluer une voix comme dysphonique. Pour le souffle, ces résultats pourraient nous mener à mettre en rapport l'évaluation par les naïfs et les représentations sociales liées à la voix soufflée. En effet, cet indice pourrait également faire partie de ceux qui sont perceptivement appréciés pour les femmes francophones comme c'est le cas pour les femmes anglophones aux États-Unis (Babel et al., 2014). Pour ce qui est de la raucité, ces résultats sont probablement en lien direct avec les représentations sociales et les référents internes des auditeurs naïfs qui semblent apprécier une raucité légère pour les femmes francophones. Nos résultats, ainsi que ceux de la littérature mise en avant dans le cadre théorique, tendent à montrer que la raucité n'est pas considérée comme pathologique par les naïfs francophones. Beaucoup moins de résultats existant sur le souffle, il pourrait être intéressant de réaliser une étude concernant plus précisément cette dimension vocale.

Enfin, au vu des résultats de la littérature et de ceux que nous apportons, il pourrait être intéressant de reconsidérer certains traits perceptifs comme relevant nécessairement de la pathologie, si toutefois ils ne font pas l'objet d'une plainte ou d'une atteinte organique. L'exemple de la raucité est particulièrement frappant puisqu'elle est visiblement dans la culture française, considérée comme vocalement attractive chez les hommes comme chez les femmes, du moins dans le cas d'une raucité légère et n'est que peu prise en compte par les auditeurs naïfs lors de l'évaluation d'une voix comme pathologique. Doit-on alors nécessairement considérer cette dimension comme un indice de pathologie ? L'idée selon laquelle une dimension perceptive pourrait être reconsidérée selon la norme sociale en cours au moment et à l'endroit de l'évaluation vocale pourrait être appuyé par certains arguments, comme le fait qu'en Angleterre, les femmes peuvent moduler leur voix chantée pour augmenter leur attractivité vocale en utilisant une qualité de voix plus rauque (Barkat-Defradas et al., 2013). Dans ce contexte, la raucité n'est pas la conséquence directe d'une altération de la qualité de voix mais d'une volonté de se rapprocher artificiellement d'une image vocale socialement validée.

Pour conclure, les voix perçues comme plus pathologiques par les auditeurs naïfs semblent être celles avec un grade de sévérité global plus élevé mais également celles avec un serrage vocal plus sévère. Cet indice est, selon nos résultats, plus important dans l'identification d'une pathologie vocale pour le jury naïf comparé au jury expert. Certaines autres dimensions comme la raucité ou le souffle semblent être plus à la frontière entre la perception de la voix pathologique et normophonique. Ces écarts entre experts et naïfs proviennent probablement d'un référent interne différent ainsi que de représentations sociales ancrées et hautement variable d'un individu à l'autre. Ce référent interne reste connu pour être plus commun entre experts qu'entre naïfs car bâti sur le nombre d'année d'exercice, ce qui permet aux experts de rattacher ce qu'ils perçoivent à ce qu'ils imaginent du mode de fonctionnement des plis vocaux.

Remerciements

Ce travail a été soutenu par le Laboratoire d'Excellence Empirical Foundations of Linguistics (LabEx EFL, ANR-10-LABX-0083). Il contribue à l'IdEx Université de Paris (ANR-18-IDEX-0001).

Références

- AMIR, O., & LEVINE-YUNDOF, R. (2013). Listeners' Attitude Toward People With Dysphonia. *Journal of Voice*, 27(4), 524. <https://doi.org/10.1016/j.jvoice.2013.01.015>
- AUTESSERRE D, CHARPY N, CREVIER-BUCHMAN L, et al. La Voix: Ses Troubles Chez Les Enseignants. Paris: Les éditions INSERM; 2006. <https://hal-lara.archives-ouvertes.fr/hal-01570681v1>
- BABEL, M., MCGUIRE, G., & KING, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS one*, 9(2), e88616. <https://doi.org/10.1371/journal.pone.0088616>
- BARKAT-DEFRADAS, M., BUSSEUIL, C., CHAUVY, O., HIRSCH, F., FAUTH, C., & RÉVIS, J. (2012). Dimension esthétique des voix normales et pathologiques : Approches perceptives et acoustiques. TIPA. *Travaux interdisciplinaires sur la parole et le langage*, (28), 2-15. <https://shs.hal.science/halshs-00778795>
- BARKAT-DEFRADAS, M., FAUTH, C., HIRSCH, F., AMY DE LA BRETÈQUE, B., SAUVAGE, J., & DODANE, C. (2013). Rauque « n » Roll : La raucité, entre symptôme pathologique & expression artistique. *Présenté à 5^e Journées de Phonétique Clinique*, Liège, Belgium. <https://hal.univ-lorraine.fr/hal-00918332>
- BLOOD, G. W., MAHAN, B. W., & HYMAN, M. (1979). Judging personality and appearance from voice disorders. *Journal of Communication Disorders*, 12(1), 63-67. [https://doi.org/10.1016/0021-9924\(79\)90022-4](https://doi.org/10.1016/0021-9924(79)90022-4)
- BOERSMA, P., & WEENINK, D. (2023). Praat: Doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 15 December 2023 from <http://www.praat.org/>.
- DE BODT, M. S., VAN DE HEYNING, P. H., WUYTS, F. L., & LAMBRECHTS, L. (1996). The perceptual evaluation of voice disorders. *Acta Oto-Rhino-Laryngologica Belgica*, 50(4), 283-291.
- DIEDENHOFEN, B. & MUSCH, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE*, 10(4): e0121945. doi:10.1371/journal.pone.0121945
- EADIE, T. L., KAPSNER, M., ROSENZWEIG, J., WAUGH, P., HILLEL, A., & MERATI, A. (2010). The Role of Experience on Judgments of Dysphonia. *Journal of Voice*, 24(5), 564-573. <https://doi.org/10.1016/j.jvoice.2008.12.005>
- FARAHANI, M., PARSA, V., HERRMANN, B., KADEM, M., JOHNSRUDE, I., & DOYLE, P. C. (2020). An auditory-perceptual and pupillometric study of vocal strain and listening effort in adductor spasmodic dysphonia. *Applied Sciences*, 10(17), 5907. <https://doi.org/10.3390/app10175907>
- GHIO, A., WEISZ, F., BARACCA, G., CANTARELLA, G., ROBERT, D., WOISARD, V., FUSSI, F., et al. (2011). Is the Perception of Voice Quality Language-Dependent? A Comparison of French and Italian Listeners and Dysphonic Speakers. *Présenté à INTERSPEECH 2011*, Florence, Italy. <https://hal.science/hal-01514687>
- GHIO, A., DUFOUR, S., ROUAZE, M., BOKANOWSKI, V., POUCHOULIN, G., RÉVIS, J., & GIOVANNI, A. (2011). Mise au point et évaluation d'un protocole d'apprentissage de jugement perceptif de la sévérité de dysphonies sur de la parole naturelle. *Revue de Laryngologie Otologie Rhinologie*, 132(1), 19-27. <https://hal.science/hal-01491737v1>

- GHIO, A., CANTARELLA, G., WEISZ, F., ROBERT, D., WOISARD, V., FUSSI, F., GIOVANNI, A., et al. (2015). Is the perception of dysphonia severity language-dependent? A comparison of French and Italian voice assessments. *Logopedics Phoniatrics Vocology*, 40(1), 36-43. <https://doi.org/10.3109/14015439.2013.837503>
- HELOU, L. B., SOLOMON, N. P., HENRY, L. R., COPPIT, G. L., HOWARD, R. S., & STOJADINOVIC, A. (2010). The Role of Listener Experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) Ratings of Postthyroidectomy Voice. *American Journal of Speech-Language Pathology*, 19(3), 248-258. [https://doi.org/10.1044/1058-0360\(2010/09-0012\)](https://doi.org/10.1044/1058-0360(2010/09-0012))
- HIDAKA, S., LEE, Y., NAKANISHI, M., WAKAMIYA, K., NAKAGAWA, T., & KABURAGI, T. (2022). Automatic GRBAS Scoring of Pathological Voices using Deep Learning and a Small Set of Labeled Voice Data. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2022.10.020>
- HIRANO, M. (1981). Clinical examination of voice. Wien; New York: Springer-Verlag. <https://doi.org/10.1121/1.393788>
- KREIMAN, J., GERRATT, B. R., KEMPSTER, G. B., ERMAN, A., & BERKE, G. S. (1993). Perceptual Evaluation of Voice Quality: Review, Tutorial, and a Framework for Future Research. *Journal of Speech, Language, and Hearing Research*, 36(1), 21-40. <https://doi.org/10.1044/jslr.3601.21>
- KREIMAN, J., GERRATT, B. R., & BERKE, G. S. (1994). The multidimensional nature of pathologic vocal quality. *The Journal of the Acoustical Society of America*, 96(3), 1291-1302. <https://doi.org/10.1121/1.410277>
- KREIMAN, J., & GERRATT, B. R. (1996). The perceptual structure of pathologic voice quality. *The Journal of the Acoustical Society of America*, 100(3), 1787-1795. <https://doi.org/10.1121/1.416074>
- KREIMAN, J., & GERRATT, B. R. (2000). Sources of listener disagreement in voice quality assessment. *The Journal of the Acoustical Society of America*, 108(4), 1867-1876. <https://doi.org/10.1121/1.1289362>
- KREIMAN, J., & SIDTIS, D. (2011). Foundations of voice studies: An interdisciplinary approach to voice production and perception. John Wiley & Sons.
- KREIMAN, J., AUSZMANN, A., & GERRATT, B. R. (2020). What Does It Mean for a Voice to Sound “Normal”? Voice Attractiveness (p. 83-99). Springer.
- LALLH, A. K., & ROCHET, A. P. (2000). The Effect of Information on Listeners' Attitudes Toward Speakers With Voice or Resonance Disorders. *Journal of Speech, Language, and Hearing Research*, 43(3), 782-795. <https://doi.org/10.1044/jslhr.4303.782>
- MENG, X. L., ROSENTHAL, R., & RUBIN, D. B. (1992). Comparing correlated correlation coefficients. *Psychological bulletin*, 111(1), 172.
- MISONO, S., MERATI, A. L., & EADIE, T. L. (2012). Developing Auditory-Perceptual Judgment Reliability in Otolaryngology Residents. *Journal of Voice*, 26(3), 358-364. <https://doi.org/10.1016/j.jvoice.2011.07.006>
- PETTIROSSI, A., AUDIBERT, N., & CREVIER-BUCHMAN, L. (2020). Corrélats acoustiques et perceptifs de la personnalité perçue à travers la voix dans une population de dysphoniques légères. In *6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique*

pour le Traitement Automatique des Langues (RÉCITAL, 22e édition) (pp. 489-497). ATALA. <https://hal.science/hal-02798576/>

PETTIROSSI, A. (2021). La dysphonie chez les professeurs des écoles: perception et représentations (Doctoral dissertation, Université Sorbonne Nouvelle). <https://theses.hal.science/tel-03152574>

R CORE TEAM. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Consulté à l'adresse <https://www.R-project.org/>.

ROUQUETTE, M.-L., & RATEAU, P. (1998). Introduction à l'étude des représentations sociales. Presses Universitaires de Grenoble.

SCHUERING, J. H. C., VAN HOF, K. S., HEIJNEN, B. J., VAN BENTHEM, P. P. G., SJÖGREN, E. V., & LANGEVELD, A. P. M. (2021). Proposal for a Core Outcome Set of Measurement Instruments to Assess Quality of Voice in Adductor Spasmodic Dysphonia Based on a Literature Review. *Journal of Voice*, 35(6), 933.e7-933.e21. <https://doi.org/10.1016/j.jvoice.2020.02.010>

SELLARS, C., STANTON, A. E., MCCONNACHIE, A., DUNNET, C. P., CHAPMAN, L. M., BUCKNALL, C. E., & MACKENZIE, K. (2009). Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. *The Journal of Laryngology & Otology*, 123(7), 755-763. <https://doi.org/10.1017/S0022215109004605>

SUHAIL, I., KAZI, R., & JAGADE, M. (2016). Perceptual evaluation of tracheoesophageal speech: Is it a reliable tool? *Indian Journal of Cancer*, 53(1), 127-131. <https://doi.org/10.4103/0019-509X.180814>

YAMAGUCHI, H., SHRIVASTAV, R., ANDREWS, M. L., & NIIMI, S. (2003). A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatria et Logopaedica*, 55(3), 147-157. <https://doi.org/10.1159/000070726>

ZUCKERMAN, M., & DRIVER, R. E. (1988). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13(2), 67-82. <https://doi.org/10.1017/S0022215109004605>