

Relations entre évaluation acoustique et perceptive de la qualité de la voix : une approche dynamique

Nicolas Audibert¹, Claire Pillot-Loiseau¹, Lise Crevier-Buchman^{1,3}, Sophie Fagniard², Bernard Harmegnies², Kathy Huet², Amelia Pettirossi^{1,4}, Myriam Piccaluga², Reina Remman¹, Virginie Roland², Lola TERNY², Clémence Verhaegen², Véronique Delvaux²

¹ Laboratoire de Phonétique et Phonologie (LPP), UMR7018 CNRS/Université Sorbonne Nouvelle, Paris, France

² Service de Métrologie et Sciences du Langage, Institut de Recherche en Sciences et Technologies du Langage & Fond National de La Recherche Scientifique, Université de Mons, Belgique.

³Hôpital Foch : Service de Laryngologie Phoniatre, 92150 Suresnes, France

⁴Imagerie & Cerveau, UMR1253 INSERM/Université de Tours, 37000 Tours, France

prenom.nom@sorbonne-nouvelle.fr ; Prenom.NOM@umons.ac.be

Introduction

La notion de qualité vocale (QV) est complexe et multiparamétrique. En contexte clinique, évaluation typiquement fondée sur une diversité d'outils/indicateurs parmi lesquelles : évaluation perceptive par le clinicien à l'aide d'échelles (GRBASI, CAPE-V), auto-évaluation par le patient (VHI), indicateurs acoustiques simples (jitter, shimmer, HNR, CPPs, etc.) ou composites (DSI, AVQI), dont l'interprétation est modulée par l'évaluation de la fonction pneumo-phonatoire et l'imagerie des plis vocaux.

(Kreiman et Gerratt, 2010 ; Barsties et De Bodt, 2015 ; Brockmann-Bausser, 2023)

Recherche en évaluation de la QV marquée par deux problématiques non résolues :

- Question des relations, non linéaires et non transparentes, entre paramètres acoustiques et jugements perceptifs
- Lien entre évaluation de la QV et l'échantillon évalué, typiquement /a/ tenus vs. échantillons de parole continue

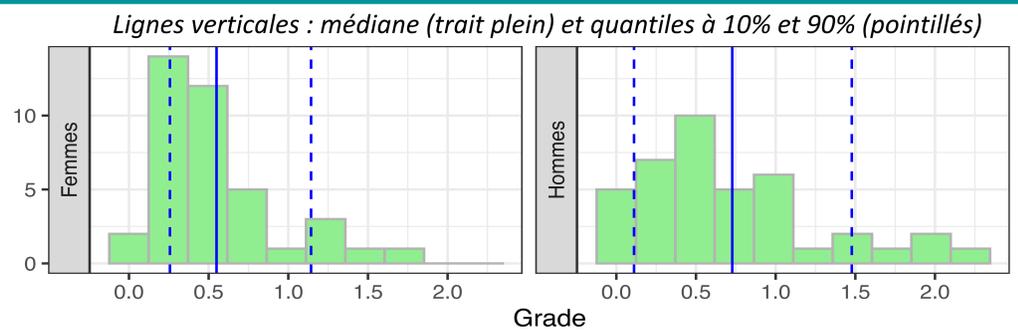
Question liée : dynamique temporelle de la QV dans la parole

Questions de recherche

Dans quelle mesure la prise en compte de la dynamique d'un ensemble de paramètres acoustiques permet de mieux prédire les jugements perceptifs sur les mêmes échantillons de parole ?

Données

- Productions issues d'enregistrements exploités par Delvaux et Pillot-Loiseau (2020) à l'aide de MonPaGe (Laganaro et al., 2021)
- Extraction de la phrase entièrement voisée « Mélanie vend du lilas » produite par 39 femmes (21-85 ans) et 40 hommes (20-88 ans) normophoniques de la région parisienne
- Jugements sur l'échelle GRBASI par 9 orthophonistes expérimentées, exploitation du grade pour cette 1^{ère} étude



Mesures acoustiques

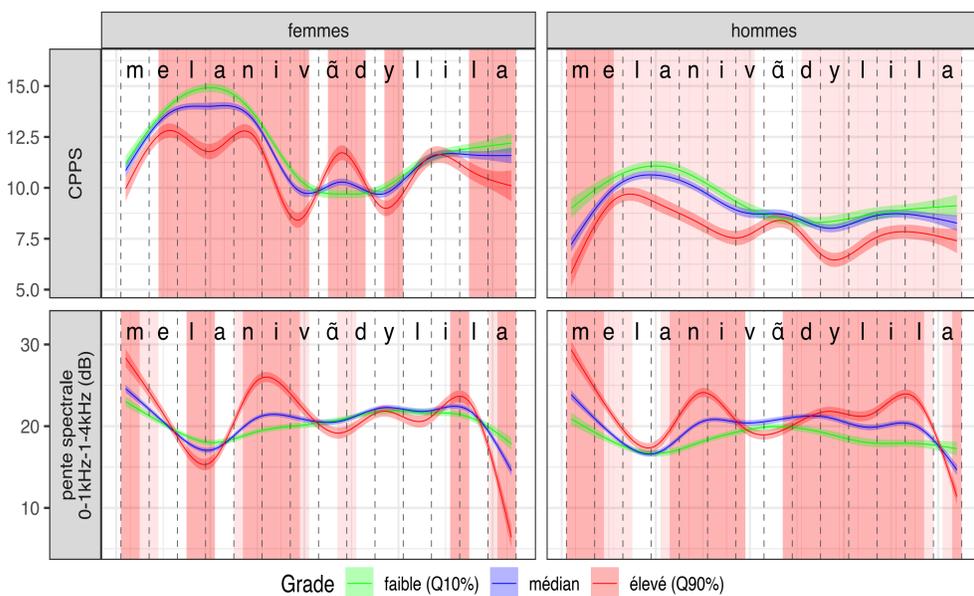
- Rapport harmonique sur bruit HNR
- Pic de proéminence cepstrale lissé CPPs (Maryn et al. 2009)
- Rapport énergie entre bandes 0-1kHz et 1-4kHz (Kitzing, 1986)

Après alignement forcé (WebMAUS + correction manuelle), mesure sur 3 points par segment

Modélisation statistique dynamique

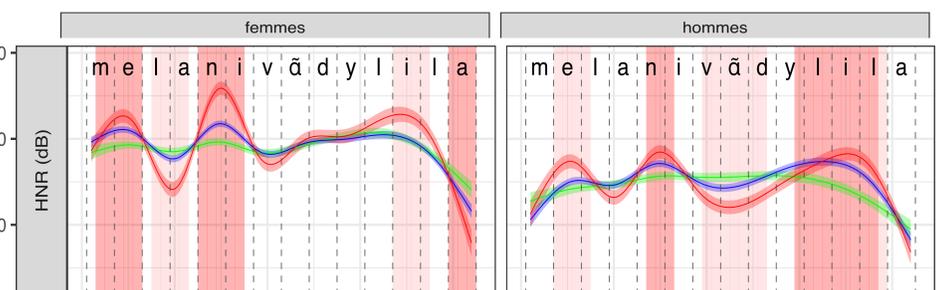
- GAM par variable et par sexe : évolution des mesures dans la phrase
- Correction de l'influence de f_0 sur les mesures via prise en compte comme lissage (*smooth*) dans les modèles
- Projection des valeurs acoustiques pour différentes valeurs de grade + estimation des points de mesures avec effet significatif du grade

Résultats et discussion



Prédictions issues de chacun des 6 modèles GAM pour 1^{er} décile (F : 0,14 ; H : 0,11), médiane (F : 0,43 ; H : 0,56) et 9^e décile (F : 1,14, H : 1,44) des valeurs de grade.

Fond rose pâle = diff. significative uniquement entre grade faible et élevé ; rose foncé = différence significative également entre grade faible et médian



- Pouvoir prédictif des indices acoustiques / altérations perçues évolue au cours du temps, i.e. en fonction du contenu segmental et/ou de la position prosodique
- Pattern différent pour H et F malgré correction f_0 , et variable selon l'indice acoustique concerné
- Effet plus fort sur nasales, contradiction avec résultats littérature sur classification dysphonie en fonction du segment (Pouchoulin et al., 2008)
- Perspectives : prise en compte du GRBASI au-delà du seul grade, évaluation de la réplicabilité sur autres données