Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

Étude de la qualité vocale dans la parole professionnelle des aides-soignants français

Jean-Luc Rouas[1] Yaru Wu[2] Takaaki Shochi[1, 3]
(1) Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
(2) CRISCO/UR4255, Université de Caen Normandie, 14000 Caen, France
(3) CLLE CNRS UMR 5263, Bordeaux, France
rouas@labri.fr, yaru.wu@unicaen.fr, takaaki.shochi@labri.fr

RÉSUMÉ

Cet article présente une méthodologie complète pour étudier les attributs vocaux des aides-soignants travaillant dans des maisons de retraite en France. L'objectif était d'analyser les modèles de parole de 20 aides-soignants dans deux établissements distincts. Les aides-soignants ont été équipés de microphones-casque connectés à des smartphones pour garantir une qualité audio optimale. Les données enregistrées comprenaient la lecture de texte, des entretiens informels et des jeux de rôle professionnels avec des patients fictifs. Le traitement des données a été effectué à l'aide d'un système de reconnaissance automatique de la parole de pointe, permettant de générer des séquences de mots ou de phonèmes avec leurs frontières. L'analyse s'est concentrée sur la détection des variations de la qualité vocale dans divers contextes de parole spontanée. L'objectif final est le développement d'outils de formation automatisés pour les aides-soignants, afin de capturer et reproduire leurs caractéristiques vocales uniques, améliorant ainsi leurs capacités professionnelles.

ABSTRACT

Voice quality in French Caregivers' Professional Speech.

This paper presents a comprehensive methodology for investigating the vocal attributes of caregivers in retirement homes in France. The aim was to analyze the speech patterns of 20 proficient caregivers across two facilities. Caregivers were equipped with headset microphones connected to smartphones for high-quality audio. Data included reading text, informal interviews, and professional role-play scenarios with fictitious patients. Data processing involved a state-of-the-art automatic speech recognition system, generating word or phone sequences with timestamps. Analysis focused on detecting nuanced variations in voice quality in diverse spontaneous speech contexts. The ultimate goal is to develop cutting-edge automated training tools tuned to capture and replicate caregivers' unique vocal characteristics, thereby enhancing their caregiving capabilities.

MOTS-CLÉS : Qualité de voix, aides-soignants, styles de parole, parole spontanée.

KEYWORDS: Voice quality, caregivers voice, speaking styles, spontaneous speech.

1 Introduction

Une communication verbale efficace est essentielle lors de la prise en charge de personnes âgées dépendantes. Cependant, dans les milieux hospitaliers, les aides-soignants sont souvent accaparés par leurs tâches, ce qui entraîne des interactions brèves avec les patients. Un rapport alarmant a révélé que

Actes des 35èmes Journées d'Études sur la Parole, pages 51–60.

la communication verbale avec les patients atteints de démence alités dans les établissements de soins de longue durée durait seulement deux minutes par jour (Gineste *et al.*, 2008). Les aides-soignants peuvent trouver décourageant que les patients ne réagissent pas ou de manière non pertinente.

Des recherches ont montré que la communication volontaire et positive des aides-soignants professionnels peut avoir un fort impact sur les patients âgés atteints de démence (Gineste & Pellissier, 2007). La méthode "Humanitude" est conçue pour promouvoir des interactions positives et continues grâce au développement de compétences de communication efficaces, basées sur le contact visuel, la communication verbale et l'interaction tactile. De nombreuses études ont démontré que cette méthode entraîne une réduction significative (88,5 %) des comportements agressifs des patients et une diminution de la nécessité de médicaments neuroleptiques (Honda *et al.*, 2013, 2016; Ito & Honda, 2015).

En ce qui concerne les compétences en communication verbale, la méthode "Humanitude" s'appuie sur des éléments phonétiques et lexicologiques, ainsi que sur une technique appelée "Auto-Feedback". Dans cette technique, les aides-soignants se doivent de parler sans interruption, même lorsque les bénéficiaires de soins fournissent des réponses inadéquates. Par conséquent, il existe deux principales catégories de paramètres qui nécessitent une investigation : les paramètres prosodiques (tels que l'intensité, le débit et la mélodie) qui devraient être en accord avec la voix douce, calme et mélodieuse recommandée, et les éléments lexicaux destinés à véhiculer des émotions positives.

L'étude des paramètres prosodiques de la parole professionnelle des aides-soignants a été menée dans Rouas *et al.* (2023b) tandis que les attributs affectifs ont été examinés dans Rouas *et al.* (2023a). L'analyse acoustique réalisée dans Rouas *et al.* (2023b) a montré que des valeurs plus élevées pour $F0$ sont observées pour les enregistrements professionnels des aides-soignants. Bien que cela semble contraire à ce qui peut être attendu en ce qui concerne la voix "douce" recommandée par "Humanitude", nous formulons ici l'hypothèse que les changements de qualité vocale, comme les modifications de la quantité de souffle, peuvent avoir un impact sur la perception de la douceur ou de la proximité de la voix. L'objectif de cet article est donc l'étude des attributs de qualité vocale de la parole des aides-soignants.

Nous résumons brièvement dans la Section 2 les informations qui peuvent être véhiculées à l'aide des attributs de qualité vocale. Le protocole d'enregistrement du corpus produit par les aides-soignants professionnels français avec la description de l'équipement spécifique utilisé pour permettre la liberté de mouvement, les tâches effectuées et les paramètres d'enregistrement sont décrits dans la Section 3. Ensuite, la Section 4 décrit comment nous avons prétraité les fichiers pour obtenir la transcription phonétique qui est ensuite utilisée pour extraire les caractéristiques de qualité vocale sur les Unités Inter-Pausales. L'analyse des caractéristiques est réalisée dans la Section 5 et nous discutons des résultats dans la Section 6.

2 Qualité de la voix

La variabilité de la qualité de la voix laryngée, telle que le souffle (« breathiness ») et le craquement (« creakiness »), est observée de manière étendue dans la parole. Ces variations peuvent servir à diverses fins linguistiques, comme l'utilisation contrastée de la qualité de la voix dans les langues ou en tant que caractéristiques prosodiques comme le craquement phrastique final dans plusieurs langues. Plus important encore pour notre étude actuelle, il existe des preuves de distinctions sociolinguistiques

dans la qualité de la voix (Stuart-Smith, 1999).

Les travaux de Gobl & Ní Chasaide (2003) et de Yanushevskaya *et al.* (2006) montrent que la qualité de la voix seule peut évoquer des associations affectives, même si celles-ci n'existent pas sur une base univoque ; une qualité de voix particulière, par exemple le craquement, peut être associée à plusieurs états affectifs. Cependant, la perception des affects exprimés par les qualités vocales varie selon les langues (Yanushevskaya *et al.*, 2018). En anglais, une voix chuchotée peut être associée à la peur (de Mareüil *et al.*, 2002), tandis que dans d'autres langues, elle peut véhiculer d'autres émotions. De même, en anglais, la voix soufflée est traditionnellement associée à l'intimité (Laver, 1980), alors qu'en japonais, elle est plus souvent liée à la formalité et à la politesse (Ito, 2004; Ishi *et al.*, 2008). De manière similaire, la voix craquée tend à évoquer différents états émotionnels en fonction du contexte linguistique. En français, Grichkovtsova *et al.* (2012) montrent que la perception des attitudes repose principalement sur le contour prosodique, tandis que les émotions reposent à la fois sur la qualité de la voix et sur le contour prosodique.

En plus de la relation entre qualité de voix et émotions, certains chercheurs ont également essayé d'établir des liens entre différentes qualités de voix et la personnalité perçue d'un locuteur en anglais américain (Pearsell & Pape, 2023). Alors que la voix craquée tend généralement à être perçue de manière négative, des recherches suggèrent que la voix soufflée influence principalement la perception de la voix d'une locutrice en termes de sexualité et de sensualité (Laver, 1980). De plus, la voix soufflée peut être liée à la perception de la solidarité (Pittman, 1985).

3 Protocole d'enregistrement

3.1 Équipement

Pour garantir une mobilité totale lors des enregistrements, nous avons créé un dispositif entièrement autonome. Nous avons équipé nos sujets d'un microphone directionnel haut de gamme, le casque DPA 4288 CORE, connecté à un préamplificateur iRIG PRO. Ce préamplificateur et un smartphone Samsung Galaxy A51 étaient placés dans un sac banane, offrant une liberté de mouvement totale, tout en maintenant un enregistrement de haute qualité.

3.2 Tâches

En utilisant cet équipement, nous avons enregistré nos sujets sur trois tâches différentes :

Lecture de texte : lecture à voix haute de *La bise et le soleil*. Le but de cette tâche est d'enregistrer un échantillon vocal contrôlé dans un contexte très structuré.

Entretien informel : réponses à des questions ouvertes axées sur le travail du soignant et sa routine quotidienne. Le but de cette tâche est de collecter des données spontanées dans une interaction "naturelle". Cet exercice aide également à renforcer la confiance du locuteur.

Tâche de soin professionnelle : réaliser une tâche de soin sur un patient fictif non réactif. La tâche de soin sélectionnée était l'habillage, ce qui comprenait le boutonnage d'une chemise et des techniques de réveil du corps le matin. Après l'habillage, le soignant aide le patient (simulé) à se lever et à marcher. Pour évaluer la technique de « Auto-Feedback », le patient reste totalement silencieux tout

en permettant au soignant d'administrer les soins. Le cadre a été intentionnellement conçu pour être familier au soignant, ressemblant à une chambre typique, avec des cloisons pour un sentiment d'intimité avec le patient simulé.

3.3 Données collectées

Les enregistrements audio ont été collectés dans deux établissements d'hébergement pour personnes âgées dépendantes (EHPAD) dans le sud-ouest de la France : « Les Balcons du Lot » à Prayssac et « Les Résidences du Quercy Blanc » à Castelnau-Montratier. Trois sessions d'enregistrement ont eu lieu : deux à l'établissement de Prayssac le 24 septembre 2021 et le 25 mars 2022, et une à Castelnau-Montratier le 24 novembre 2021.

Au total, 25 participants ont été enregistrés lors de ces sessions, dont 21 femmes et 4 hommes. Pour l'analyse, nous avons exclu les enregistrements des 4 participants masculins et 1 enregistrement d'une participante féminine en raison de la mauvaise qualité d'enregistrement. Nous avons donc conservé 20 participants pour une durée combinée de 2 heures et 30 minutes. Les durées spécifiques des tâches et les durées moyennes par locuteur par tâche sont spécifiées dans le Tableau 1.

Tâche	durée moyenne	durée totale (s)
Lecture de texte	45,0 s.	15 min 03 s.
Entretien	134,8 s.	44 min 56 s.
Tâche de soin	188,9 s.	62 min 58 s.

TABLE 1 – Durée moyenne par locuteur et par tâche et durée totale par tâche. L'ensemble des 20 sujets enregistrés a participé à chaque tâche.

4 Paramètres

4.1 Transcription orthographique et phonétique automatique

Nous avons utilisé un système automatisé, basé sur le framework Kaldi (Povey *et al.*, 2011), pour générer des transcriptions orthographiques et phonétiques. Ce système a été entraîné en utilisant la base de données ESTER (Galliano *et al.*, 2009) et utilise un réseau neuronal à retard temporel (Time Delayed Neural Network - TDNN) couplé à un modèle de Markov caché. Le TDNN est composé de 7 couches, chacune avec 1024 unités. Le modèle acoustique prend en entrée un vecteur MFCC haute résolution de 40 dimensions concaténé avec un i-vecteur de 100 dimensions (Gupta *et al.*, 2014). Ce système atteint un taux d'erreur de 13,7% sur l'ensemble de test du corpus ESTER (Boyer, 2021), ce qui est proche des performances de l'état de l'art sur le même corpus (légèrement inférieur à 12% WER (Heba, 2021)). Les symboles phonétiques et leur alignement sont obtenus à l'aide de la commande *lattice-align-phones*, ce qui permet de segmenter et d'annoter 35 phonèmes. La transcription phonétique de sortie est utilisée pour calculer les caractéristiques moyennes sur chaque unité phonétique.

4.2 Mesures de la qualité de la voix

Au cours des deux dernières décennies, la qualité de la voix laryngée a suscité un intérêt croissant dans les domaines de l'articulation, de l'acoustique et de la perception. Les phonéticiens ont activement recherché des attributs acoustiques capables de distinguer entre les voix modales, craquées et soufflées. Les dimensions articulatoires fondamentales qui contribuent à la qualité de la voix laryngée sont le degré de constriction des plis vocaux et la présence de bruit d'aspiration. Ces dimensions peuvent être quantifiées à l'aide de mesures telles que la pente spectrale et le rapport harmonique-sur/bruit (Harmoics to Noise Ratio - HNR) (Kane & Gobl, 2011).

La pente spectrale peut être mesurée de différentes manières, la plus courante étant le calcul de H1–H2 (Bickley, 1982) qui représente la différence d'amplitude entre le premier et le deuxième harmonique. De plus, des mesures alternatives de la pente spectrale comprennent la mesure de la différence d'amplitude entre les formants et le premier harmonique (H1–A1, H1–A2 et H1–A3) qui peuvent être liées à des types spécifiques de phonation (Kane & Gobl, 2011). Dans un travail récent, Chai & Garellek (2022) a proposé de remplacer le H1-H2 par un calcul corrigé de H1. Ici, en plus des différences, nous calculons donc également H1 et H1*, la correction du calcul de H1 avec la méthode décrite dans Iseli & Alwan (2004) pour éliminer l'influence des résonances du conduit vocal.

En plus des mesures d'inclinaison spectrale et de HNR, plusieurs autres indicateurs acoustiques de la qualité de la voix laryngée ont été proposés dans la littérature. Le plus prévalent est le Pic Cepstral Prominent (Cepstral Prominence Peak - CPP) (Hillenbrand *et al.*, 1994) . Les développements récents dans l'analyse acoustique de la voix ont constamment renforcé l'importance du CPP comme indicateur objectif de l'existence de souffle et de la dysphonie globale (Patel *et al.*, 2018). Des valeurs moins importantes pour la mesure du CPP suggèrent plus de souffle dans la voix.

La qualité de la voix soufflée est corrélée avec l'énergie du bruit, principalement à des fréquences élevées, et avec la hauteur relative du premier harmonique par rapport au reste (Hillenbrand *et al.*, 1994). Cela suggère que des valeurs plus faibles pour le HNR à haute fréquence et des valeurs plus élevées pour H1, H1-H2 et H1-Ax sont attendues pour des voix plus soufflées. De manière analogue, les corrélats acoustiques les plus saillants pour la voix craquée sont un H2 plus élevé et un HNR plus élevé en dessous de 500 Hz (Xu *et al.*, 2023).

Afin d'extraire automatiquement les paramètres de qualité de voix, nous avons utilisé la boîte à outils "Snack" (Sjölander, 2004) pour obtenir les valeurs de fréquence fondamentale et les fréquences des formants. A partir de ces valeurs, nous avons calculé le CPP, le HNR ainsi que les valeurs des amplitudes des formants et des harmoniques avec une implémentation python inspirée des méthodes utilisées dans "VoiceSauce" (Shue *et al.*, 2009). Ces paramètres sont ensuite moyennés sur chaque phonème issu de la transcription automatique avant analyse.

5 Résultats

Afin d'analyser les différences dans les caractéristiques de la qualité de la voix entre les styles de parole, nous avons analysé nos données à l'aide de Modèles linéaires mixtes (Linear Mixed Models - LMM) calculés grâce au package R *lme4* (R Development Core Team, 2019). Un modèle a été produit pour chaque paramètre acoustique présenté dans la section 4. Le style de parole a été inclus comme effet fixe pour tous les modèles. En ce qui concerne les effets aléatoires, des intercepts ont été inclus

pour les sujets.

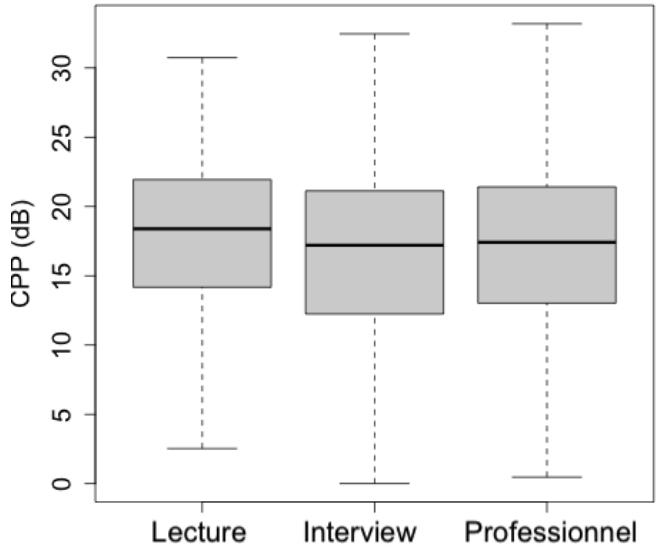

FIGURE 1 – Diagramme en boîte des valeurs normalisées de CPP pour les trois styles de parole.

Un CPP plus faible est observé (Figure 1) chez les soignants lorsqu'ils sont interviewés que lorsqu'ils s'occupent d'un patient [β = -1.26721 ; t = -2.240 ; SE = 0.56567]. Aucune différence significative n'est mesurée entre la lecture et la parole professionnelle des soignants. Ces résultats suggèrent que la parole des soignants est plus soufflée lorsqu'ils sont interviewés que lorsqu'ils accomplissent d'autres tâches.

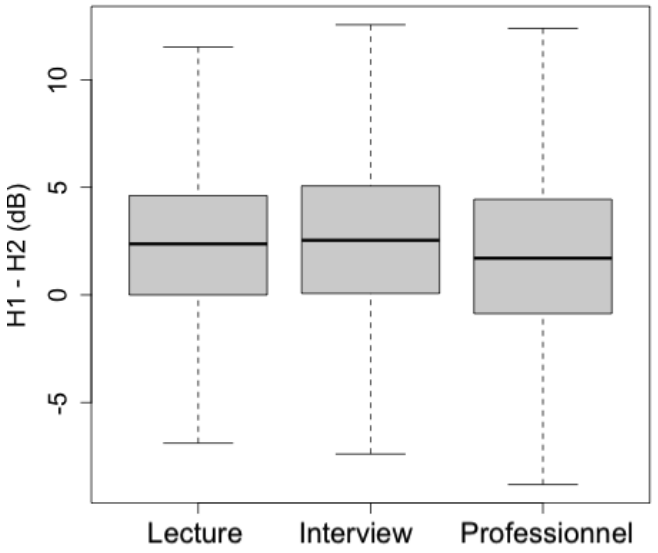

FIGURE 2 – Diagramme en boîte des valeurs normalisées de H1-H2 pour les trois styles de parole.

Le H1-H2 est plus élevé en interview [β = 0.8989 ; t = 2.841 ; SE =0.3164] que dans la parole professionnelle des soignants (Figure 2). Aucune différence significative n'est mesurée entre la lecture et la parole professionnelle des soignants. Les valeurs de H1 et H2 ont également été analysées séparément et montrent un comportement similaire (valeurs plus élevées pour l'interview et la voix professionnelle que pour la lecture).

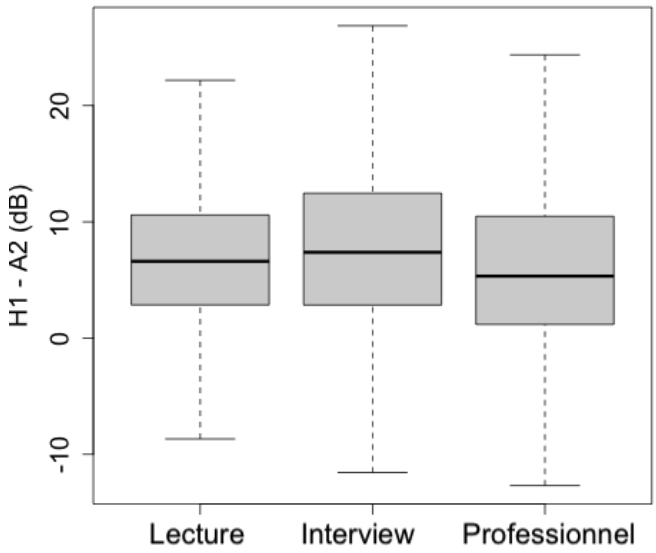

FIGURE 3 – Diagramme en boîte des valeurs normalisées de H1-A2 pour les trois styles de parole.

Les observations sur H1-A2 montrent également des valeurs plus élevées en interview [β = 2.1232 ; t = 3.066 ; SE =0.6925] que dans la parole professionnelle (Figure 3). Aucune différence significative n'est observée entre la lecture et la parole professionnelle.

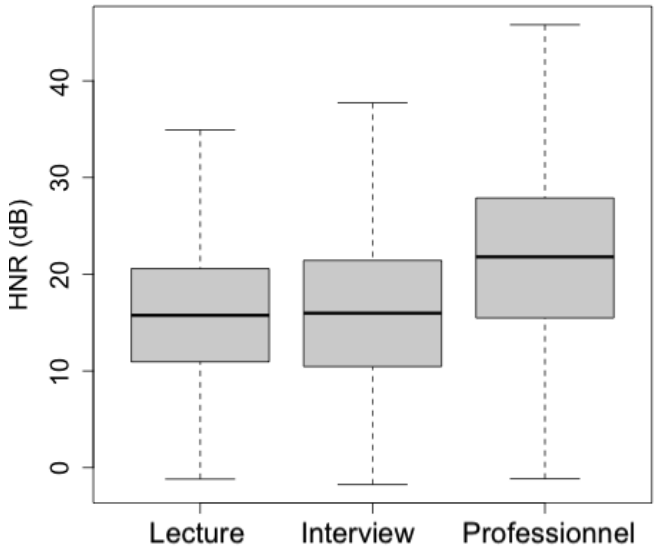

FIGURE 4 – Diagramme en boîte des valeurs normalisées de HNR pour les trois styles de parole.

Des valeurs de HNR plus faibles sont observées en interview [β = -5.5048 ; t = -5.170 ; SE = 1.0648] et en lecture [β = -5.6953 ; t = -5.337 ; SE = 1.0672], par rapport à celles observées dans la parole professionnelle (Figure 4). Ces résultats indiqueraient que les soignants utilisent moins de voix craquée dans les soins professionnels que lors de la lecture ou de l'interview.

Pour compléter ces analyses, nous fournissons les résultats LMM pour tous les paramètres dans le Tableau 2. Des différences significatives sont observées pour les paramètres H1, H2, H1-A1 et H1-A2 dans les comparaisons Interview/Professionnel et Lecture/Professionnel. Aucune différence n'est

	CPP	H1	H2	H1-H2	H1-A1	H1-A2	H1-A3	HNR
IP	*	***	*	**	NS	**	***	***
LP	NS	***	***	NS	NS	NS	**	***

TABLE 2 – Résultats LMM sur les caractéristiques de voix évaluées. « IP » fait référence à l'interview (I) comparée à la parole professionnelle (P) ; « LP » indique la lecture (L) comparée à la parole professionnelle (P). NS : Non significatif ; $* : p < .05; ** : p < .01; *** : p < .001$

trouvée pour la comparaison LP en utilisant CPP et HNR.

6 Discussion

Nous avons étudié dans cet article les styles de parole des soignants dans trois conditions différentes : la lecture de textes, les entretiens et la voix professionnelle produite lors de soins. Les résultats de l'analyse de la qualité de la voix indiquent que, lors de la prise en charge professionnelle, les soignants ont tendance à avoir des valeurs plus élevées pour le HNR et des valeurs plus faibles pour le CPP, H1, H2, H1-H2 et H1-A2.

Dans l'ensemble, ces observations ne nous fournissent malheureusement pas de conclusions claires sur la quantité de souffle utilisée dans la voix professionnelle. Les résultats sur CPP semblent illustrer une tendance vers plus de craquement dans la voix pendant les soins professionnels, tandis que les résultats du HNR sont contradictoires. L'analyse de l'amplitude des harmoniques (H1-H2, H1-A2, H1-A3) semble montrer que les soignants ont moins de souffle dans la voix lors des situations professionnelles. Toutefois, l'estimation de l'amplitude des harmoniques peut ne pas être très fiable, comme discuté dans Chai & Garellek (2022), où il est mentionné que ces mesures sont impactées par le niveau de pression sonore.

Ces résultats montrent que la proportion de souffle dans la voix, et plus largement tous les aspects de qualité de la voix, sont difficiles à mesurer et que les résultats obtenus à partir de ces mesures peuvent varier en fonction des ensembles de données et de l'objectif des études. En particulier, il nous semble évident que l'utilisation d'une seule mesure de souffle (par exemple, la CPP) n'est pas toujours suffisante.

Remerciements

Les auteurs remercient les fondateurs de l'"Humanitude" Yves Gineste et Rosette Marescotti pour leur aide dans la construction de ce projet. Un grand merci à Jean-Yves Nou, Hervé Tomassi et surtout Aurélie Rives pour nous avoir permis d'enregistrer en milieu professionnel. Nous sommes profondément redevables à tous les personnels que nous avons enregistrés pour leur confiance et leur temps.

Références

BICKLEY C. (1982). *Acoustic Analysis and Perception of Breathy Vowels*. Speech Communication Group Working Papers I, Research Laboratory of Electronics, MIT, Cambridge, MA.

BOYER F. (2021). *Reconnaissance de Parole Pour Le Français et Intégration Dans Un Système de Compréhension Du Langage Parlé*. Thèse de doctorat, Université de Bordeaux.

CHAI Y. & GARELLEK M. (2022). On H1–H2 as an acoustic measure of linguistic phonation typea). *The Journal of the Acoustical Society of America*, **152**(3), 1856–1870. DOI : 10.1121/10.0014175.

DE MAREÜIL P. B., CÉLÉRIER P. & TOEN J. (2002). Generation of Emotions by a Morphing Technique in English, French and Spanish. In *Speech Prosody*.

GALLIANO S., GRAVIER G. & CHAUBARD L. (2009). The ester 2 evaluation campaign for the rich transcription of french radio broadcasts. In *Interspeech, Brighton (United Kingdom*.

GINESTE Y., MARESCOTTI R. & PELLISSIER J. (2008). L'humanitude dans les soins. *Recherche en soins infirmiers*, **94**(3), 42–55. DOI : 10.3917/rsi.094.0042.

GINESTE Y. & PELLISSIER J. (2007). *Humanitude*. Nouvelle édition : Armand colin édition.

GOBL C. & NÍ CHASAIDE A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, **40**(1), 189–212. DOI : 10.1016/S0167-6393(02)00082-1.

GRICHKOVTSOVA I., MOREL M. & LACHERET A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, **54**(3), 414–429.

GUPTA V., KENNY P., OUELLET P. & STAFYLAKIS T. (2014). I-vector-based speaker adaptation of deep neural networks for French broadcast audio transcription. In *ICASSP*. DOI : 10.1109/ICASSP.2014.6854823.

HEBA A. (2021). *Reconnaissance Automatique de La Parole à Large Vocabulaire : Des Approches Hybrides Aux Approches End-to-End*. Theses, Université toulouse 3 Paul Sabatier.

HILLENBRAND J., CLEVELAND R. A. & ERICKSON R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, **37**(4), 769–778.

HONDA M., ITO M., ISHIKAWA S., TAKEBAYASHI Y. & TIERNEY L. (2016). Reduction of Behavioral Psychological Symptoms of Dementia by Multimodal Comprehensive Care for Vulnerable Geriatric Patients in an Acute Care Hospital : A Case Series. *Case Reports in Medicine*, **2016**, 4813196. DOI : 10.1155/2016/4813196.

HONDA M., MORI M., HAYASHI S., MORIYA K., MARESCOTTI R. & GINESTE Y. (2013). The effectiveness of French origin dementia care method ; Humanitude to acute care hospitals in Japan. *European Geriatric Medicine*, **4**, S207. DOI : 10.1016/j.eurger.2013.07.689.

ISELI M. & ALWAN A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *ICASSP 2004*, volume 1, p. I–669. DOI : 10.1109/ICASSP.2004.1326074.

ISHI C. T., ISHIGURO H. & HAGITA N. (2008). The roles of breathy/whispery voice qualities in dialogue speech. In *Speech Prosody*.

ITO M. (2004). Politeness and Voice Quality – The Alternative Method to Measure Aspiration Noise. In *Speech Prosody*, Nara, Japan.

ITO M. & HONDA M. (2015). An examination of the influence of Humanitude caregiving on the behavior of older adults with dementia in Japan. In *Proceedings of the 8th International Association of Gerontology and Geriatrics European Region Congress*, volume 2018.

KANE J. & GOBL C. (2011). Identifying regions of non-modal phonation using features of the wavelet transform. In *INTERSPEECH*, p. 177–180.

LAVER J. (1980). *The Phonetic Description of Voice Quality*. Cambridge Studies in Linguistics.

PATEL R. R., AWAN S. N., BARKMEIER K. J., COUREY M., DELIYSKI D., EADIE T., PAUL D., ŠVEC J. G. & HILLMAN R. (2018). Recommended Protocols for Instrumental Assessment of Voice : American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech-Language Pathology*, **27**(3), 887–905. DOI : 10.1044/2018_AJSLP-17-0009.

PEARSELL S. & PAPE D. (2023). The effects of different voice qualities on the perceived personality of a speaker. *Frontiers in Communication*, **7**.

PITTMAN J. (1985). *Voice Quality : Its Measurement and Functional Classification - UQ eSpace*. Thèse de doctorat, The University of Queensland, School of English, Media Studies and Art History.

POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P., SILOVSKY J., STEMMER G. & VESELY K. (2011). The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, Hilton Waikoloa Village, Big Island, Hawaii, US : IEEE Signal Processing Society.

R DEVELOPMENT CORE TEAM (2019). *R : A La nguage and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROUAS J.-L., WU Y. & SHOCHI T. (2023a). Affective attributes of French caregivers' professional speech. In *INTERSPEECH 2023*, p. 1239–1243 : ISCA. DOI : 10.21437/Interspeech.2023-848.

ROUAS J.-L., WU Y. & SHOCHI T. (2023b). A study on caregivers speech in retirement homes. In *ICPhS 2023*.

SHUE Y.-L., KEATING P. & VICENIK C. (2009). VOICESAUCE : A program for voice analysis. *The Journal of the Acoustical Society of America*, **126**(4_Supplement), 2221–2221. DOI : 10.1121/1.3248865.

SJÖLANDER K. (2004). The snack sound toolkit.

STUART-SMITH J. (1999). Glasgow : Accent and voice quality. p. 201–222. Leeds, UK : Arnold.

XU C., FOULKES P., HARRISON P., HUGHES V. & WORMALD J. H. (2023). Contributions of acoustic measures to the classification of laryngeal voice quality in continuous English speech. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* : ASSTA.

YANUSHEVSKAYA I., GOBL C. & CHASAIDE A. N. (2006). Mapping Voice to Affect : Japanese listeners. In *Speech Prosody*, Dresden, Germany.

YANUSHEVSKAYA I., GOBL C. & NÍ CHASAIDE A. (2018). Cross-language differences in how voice quality and f contours map to affecta). *The Journal of the Acoustical Society of America*, **144**(5), 2730–2750. DOI : 10.1121/1.5066448.