



# Cartographie des protocoles perceptifs d'hétéro-évaluation de la qualité vocale : voix normophoniques, pathologiques et chantées

---

Lola Terny<sup>1</sup>, Nicolas Audibert<sup>2</sup>, Lise Crevier-Buchman<sup>2</sup>, Véronique Delvaux<sup>1</sup>, Sophie Fagniart<sup>1</sup>, Bernard Harmegnies<sup>1</sup>, Kathy Huet<sup>1</sup>, Myriam Piccaluga<sup>1</sup>, Reina Remman<sup>2</sup>, Virginie Roland<sup>1</sup>, Claire Pillot-Loiseau<sup>2</sup>, Amélia Pettirossi<sup>2</sup>

<sup>1</sup>Service de Métrologie et Sciences du Langage (Université de Mons, Belgique)

<sup>2</sup>LPP, UMR 7018, CNRS (Université Paris 3-Sorbonne Nouvelle, France)

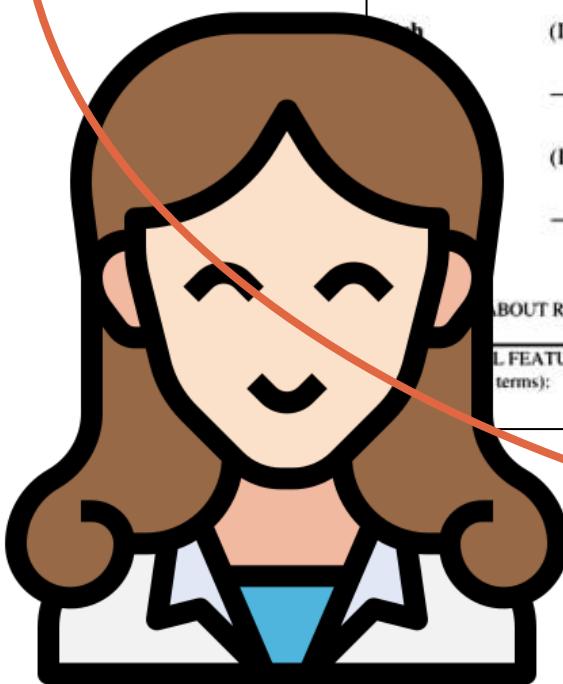
# Sommaire

---

- 01. Introduction**
- 02. Questions de recherche**
- 03. Méthodologie**
- 04. Cartographie des outils : quelques résultats**
- 05. Discussion & perceptives**

# protocole

attributs



Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Name: \_\_\_\_\_ Date: \_\_\_\_\_

The following parameters of voice quality will be rated upon completion of the following tasks:

1. Read the following sentence: "I like to eat eggs every Easter. My mother makes lemon muffins. I will keep at the peak. Please tell me how your voice is functioning."

Legend: I=Inconsistent  
D=Deviant  
MO=Moderately Deviant  
SE=Severely Deviant

Overall Severity	MI	MO	SE	SCORE C1 /100
Roughness	MI	MO	SE	SCORE C1 /100
Breathiness	MI	MO	SE	SCORE C1 /100
Strain	MI	MO	SE	SCORE C1 /100
(Indicate the nature of the abnormality): _____				SCORE C1 /100
(Indicate the nature of the abnormality): _____				SCORE C1 /100
ABOUT RESONANCE: NORMAL OTHER (Provide description): _____				
ADDITIONAL FEATURES (for example, diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, wet/gurgly, or other terms): _____				
Clinician: _____				

outils

[a]  
[badego]  
Spontaneous  
speech

...

Etc.



94 GRBAS (cont'd)

Grade 0 – non-hoarse or normal 1 – slight 2 – moderate 3 – extreme

Rough	0 – non-hoarse or normal	1 – slight	2 – moderate	3 – extreme
Breathy	0 – non-hoarse or normal	1 – slight	2 – moderate	3 – extreme
Asthenic	0 – non-hoarse or normal	1 – slight	2 – moderate	3 – extreme
Strained	0 – non-hoarse or normal	1 – slight	2 – moderate	3 – extreme
G R B A S	Total /15			

(G=grade, R=rough, B=breathy, A=asthenic, S=strain, Rating: 0=normal, 1=slight, 2=moderate, 3=severe)

Adopted from: Clinical Examination of Voice, M. Hirano, 1981

# 01. Introduction

---



- Collaboration franco-belge dans le cadre d'un Projet Hubert Curien (PHC) Tournesol
- Origine du projet et constat : diversité des protocoles d'évaluation de la qualité de voix; peu d'études dans un contexte qui ne soit pas pathologique; hétéroévaluation et subjectivité
- Revue exploratoire → comprendre les différents outils d'évaluation perceptuelle développés pour décrire la qualité de la voix d'un individu
- Originalité → **1<sup>er</sup> compte rendu exhaustif de l'ensemble des protocoles** indépendamment :
  - du **contexte** d'évaluation
  - de l'**objectif** d'évaluation
  - du **profil** des personnes évaluées
  - des **tâches** utilisées pour obtenir du matériel vocal
- Hétéroévaluation : la question de la fiabilité inter/intrajuges



# 01. Introduction

---

- Contributions antérieures → cliniques (Kreiman et al., 1993 ; De Bodt et al., 1996 ; Suhail et al., 2016 ; Schuering et al., 2021; Carding et al. 2000)
- (Hétéro)Évaluation perceptive = subjectif. D'autres études ont tenté de se baser sur des données plus objectives (IRMf, analyses acoustiques...) mais « qualité de voix est perceptive par nature » (Kreiman et al. 1993)
- Pour pallier les limites de cette « subjectivité » perceptive → créer des protocoles d'hétéroévaluation standardisés, avec des outils précisant les critères et conditions d'utilisation, pour une évaluation plus fiable et valide
- Pratique courante pour les « voix pathologiques » mais la difficulté majeure est qu'elle est dépendante de **l'expérience personnelle** et de la **formation** de l'évaluateur
  - Domaine de la logopédie : approche EBP valorisée + revues systématiques fondées sur des preuves pour élaborer des lignes directrices pratiques (ASHA, 2004 ; Choi, 2013 ; Kempster et al., 2009 ; Patel et al., 2018 ; Roy et al., 2013).
  - Consensus ELS-UEP : proposition d'un ensemble minimal multidimensionnel pour l'évaluation de la qualité de la voix qui inclut l'évaluation perceptive comme l'un de ses éléments centraux (Dejonckere et al., 2001 ; Lechien et al., 2023)
- Objectifs initiaux de notre projet :



Cartographier les outils disponibles



Les analyser en fonction de nos questions de recherche



Identifier les limites des outils existants et les besoins futurs

## 02. Questions de recherche

---



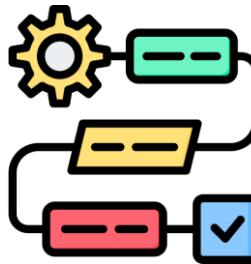
**Question 1** Quels sont les outils d'évaluation perceptuelle de la qualité vocale actuellement disponibles, et comment sont-ils structurés en termes d'échelles d'évaluation et de paramètres évalués ?

**Question 2** Comment ces outils ont-ils été appliqués dans les études sur la qualité vocale, compte tenu des voix évaluées (voix chantées, voix parlées, voix pathologiques), du sexe et de l'âge des locuteurs, de la nature des stimuli présentés aux évaluateurs et de la langue dans laquelle les évaluations ont été réalisées ?

**Question 3** Quels sont les attributs les plus fréquemment pris en compte dans les outils d'évaluation identifiés ?

**Question 4:** Dans quelle mesure les évaluations obtenues sont-elles fiables et valides, compte tenu de la concordance intra- et inter-évaluateurs ainsi que de leur lien avec les mesures acoustiques ?

# 03. Méthodologie



- **« Scoping » review?** “Scoping reviews are particularly well-suited for mapping the breadth and nature of evidence in a given field, especially when the literature is **heterogeneous** and not yet amenable to systematic synthesis. They allow for the inclusion of a **wide range of study types**, the identification of key concepts, **gaps** in knowledge, and **variations** in methodologies (...)"

Inclusion	Exclusion
Mentionner un protocole d'hétéro-évaluation perceptif de la voix contenant au moins un paramètre purement vocal	Être une étude de cas
	N'utiliser que des voix synthétiques
	Porter exclusivement sur l'effet d'un traitement, d'une thérapie ou d'un effort vocal
	N'utiliser qu'une sous-partie d'un protocole connu
Les articles présentant un protocole unique ont été inclus même s'ils violaient des critères d'exclusion	

- **Bases de données:** ProQuest, PsycINFO, PubMED, Science Direct GCPN

# Équation & diagramme PRISMA (Tricco et al. 2018)

Table 1. Equation used per database with number of results obtained before and after removing duplicates and dates of extraction

Database	Equation used	Results (n)	Without duplicates (n)	Date
Proquest	abstract((perceptual OR auditory) AND (assessment OR evaluation) AND (voice OR “voice quality”) AND (“rating scale” OR rating-scale))	86	49	17/02/2025
PsycINFO	(perceptual OR auditory) AND (assessment OR evaluation) AND (voice OR “voice quality”) AND (“rating scale**” OR rating-scale*)	228	137	17/02/2025
PubMed	(perceptual OR auditory) AND (assessment OR evaluation) AND (voice OR “voice quality”) AND (“rating scale**” OR rating-scale*)	199	102	17/02/2025
Science Direct	(perceptual OR auditory) AND (assessment OR evaluation) AND (voice OR “voice quality”) AND (“rating scale” OR rating-scale)	377	345	05/09/23
<b>TOTAL</b>		<b>890</b>	<b>633</b>	

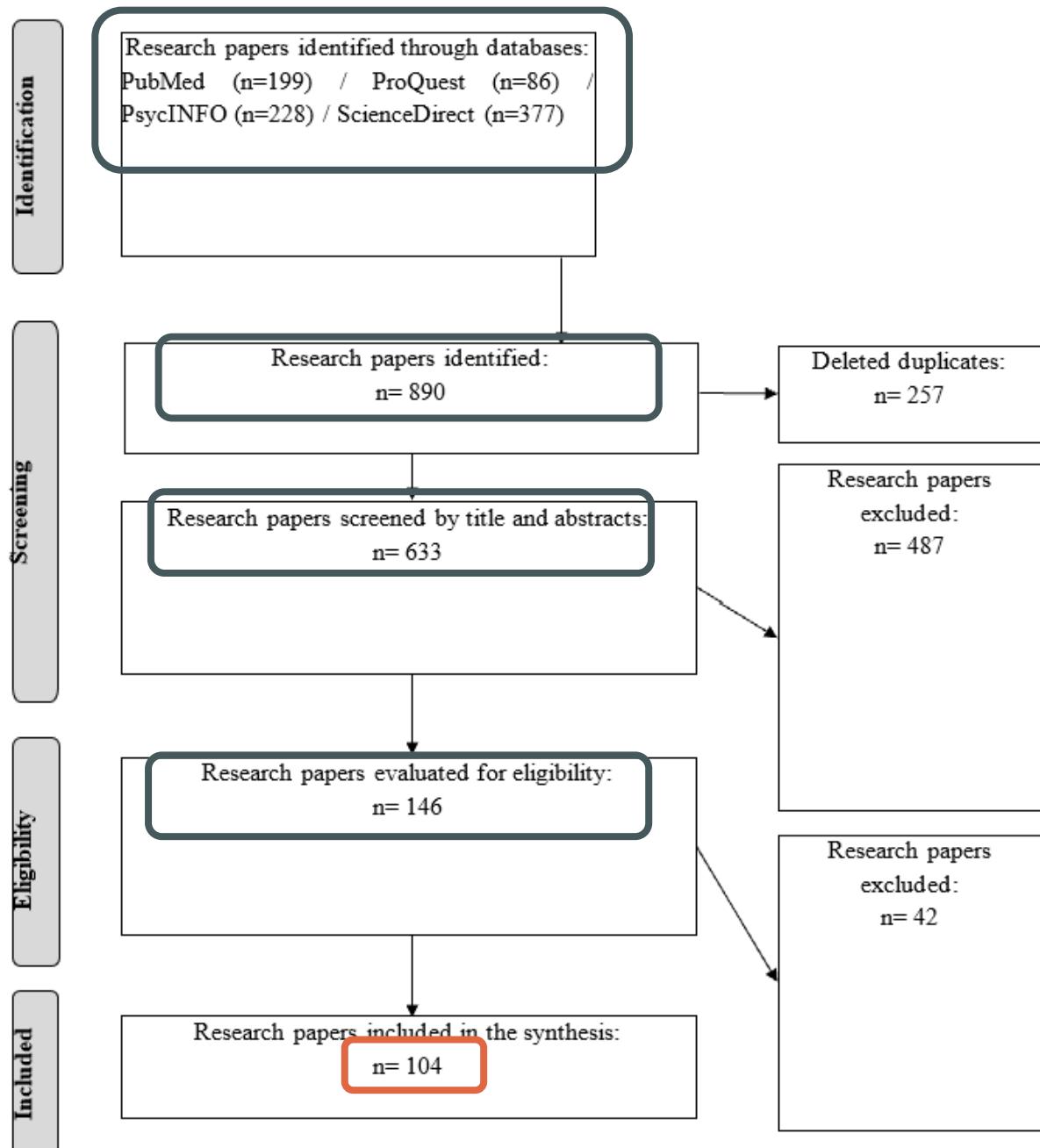


Figure 1 PRISMA scoping-review flow chart

# 04. Cartographie des outils: quelques résultats

---

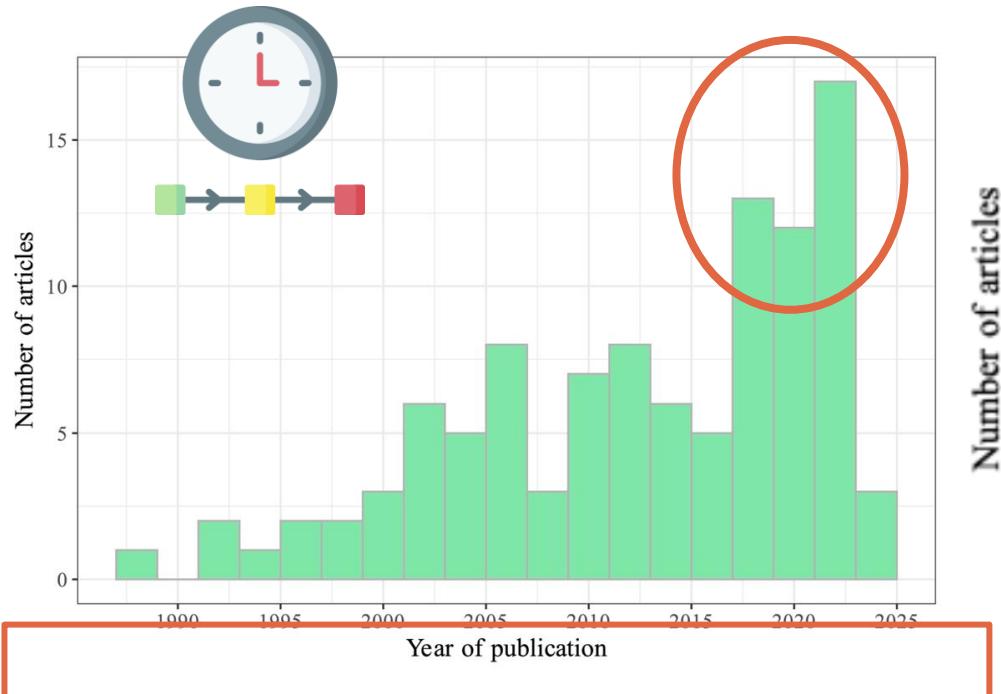
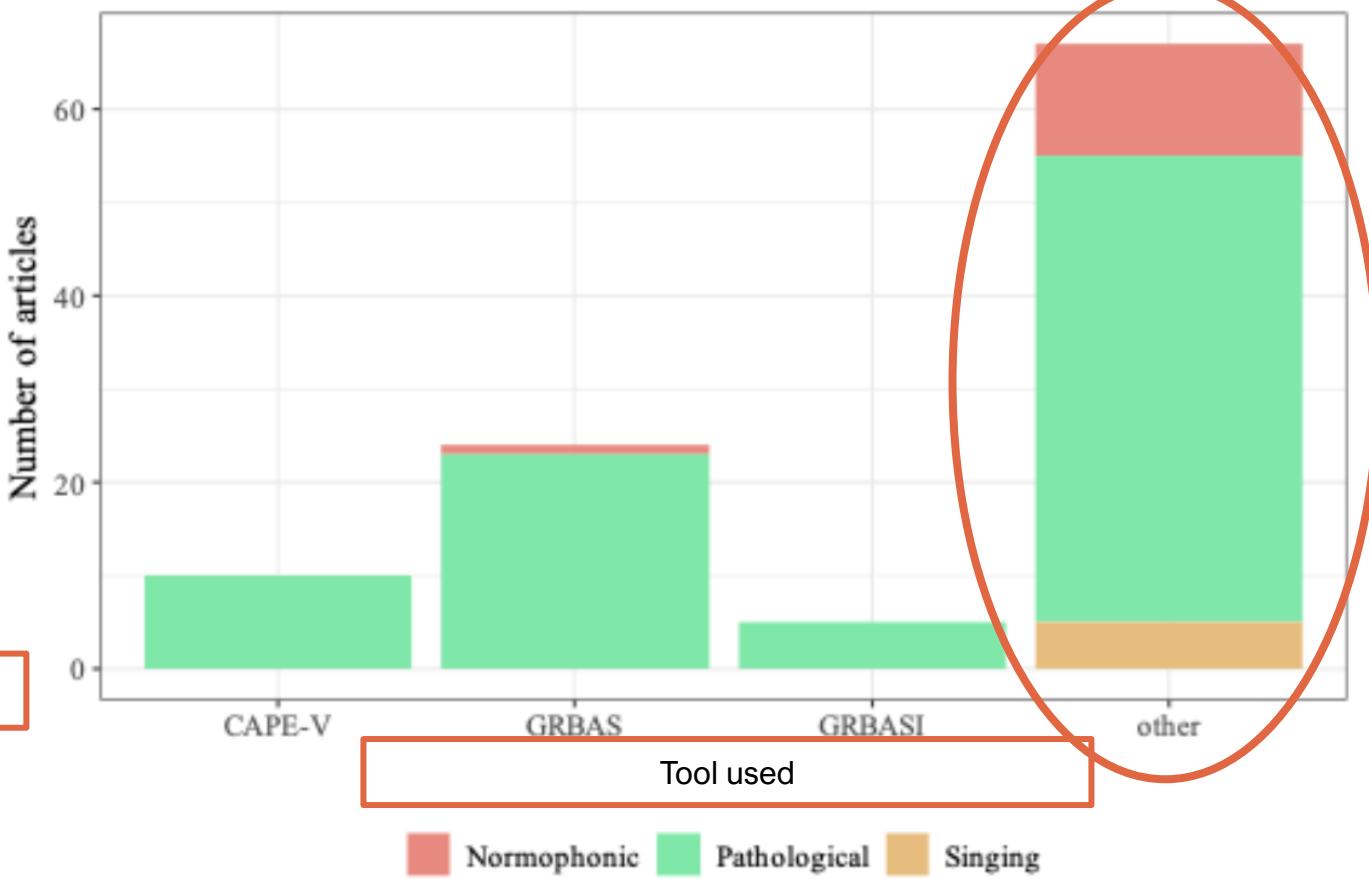


Figure 2 Distribution of the number of papers published per year as a function of year (N=104)



# 04. Quelques résultats

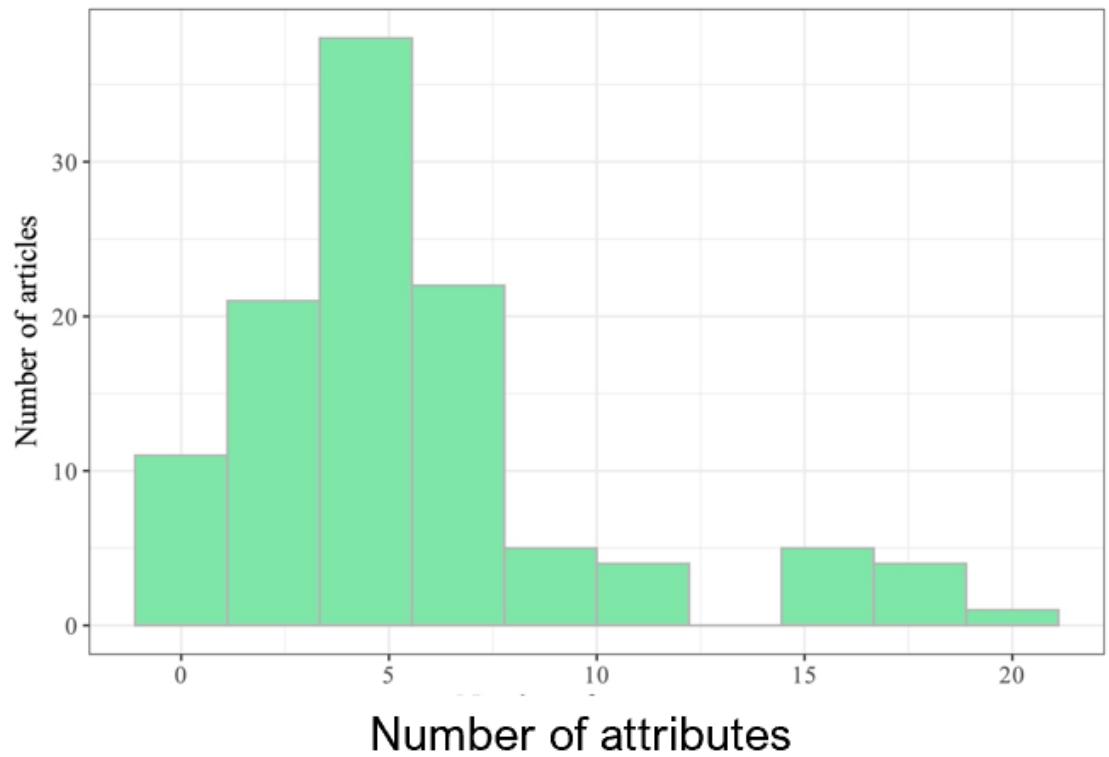
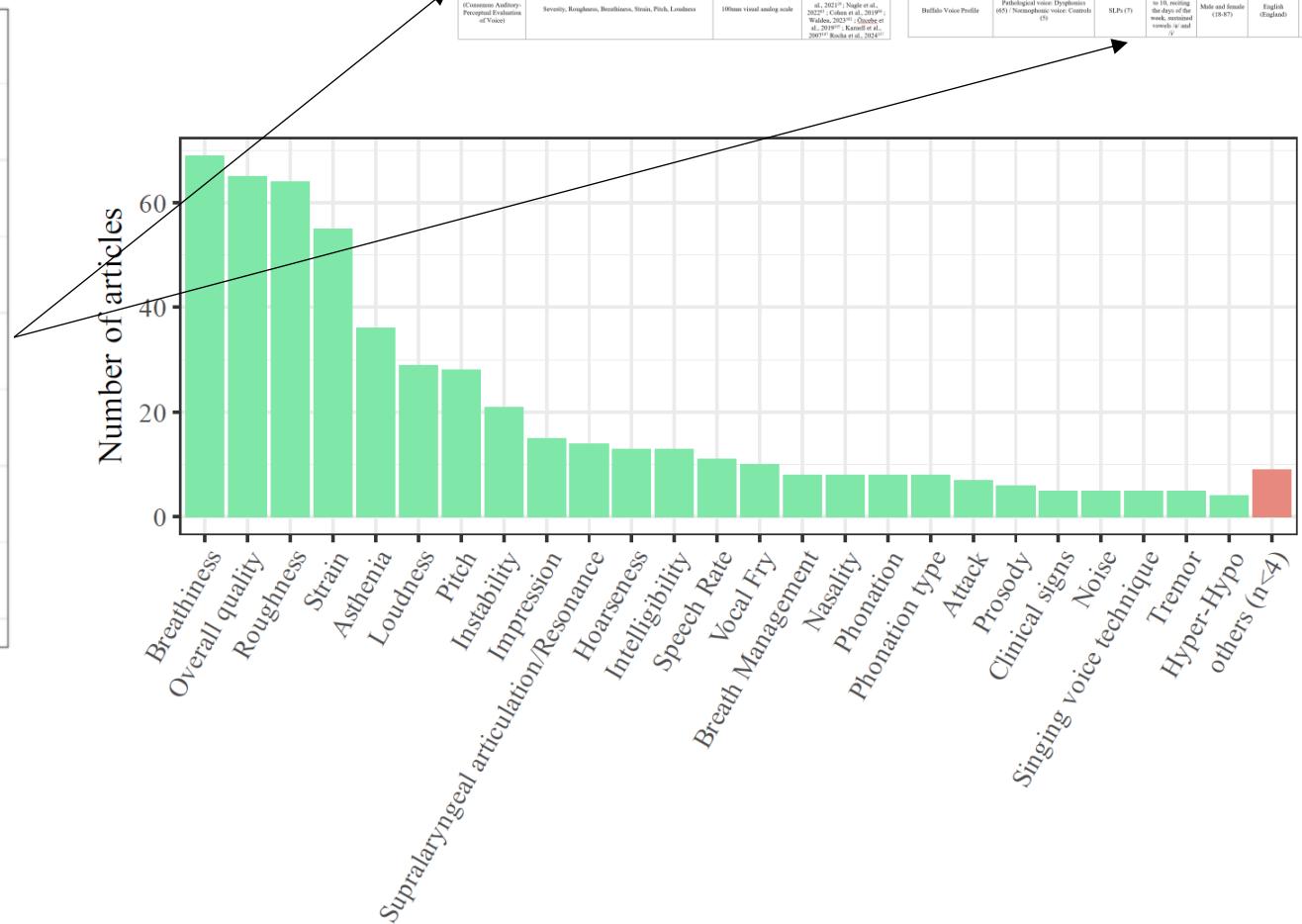
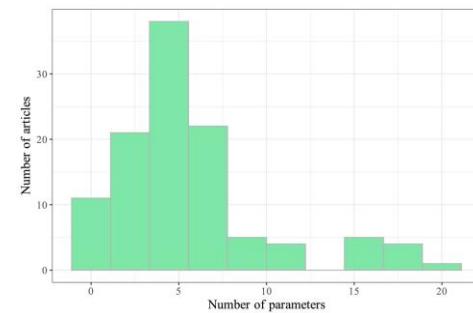


Table 2: Perceptual hetero-evaluation protocols available in the literature (listed in alphabetical order by name): name, parameters used, type of scale and corresponding reference.

Table 3. Implementation of the various protocols ordered by protocol name: specific types of voice evaluated, gender and age of the speakers,



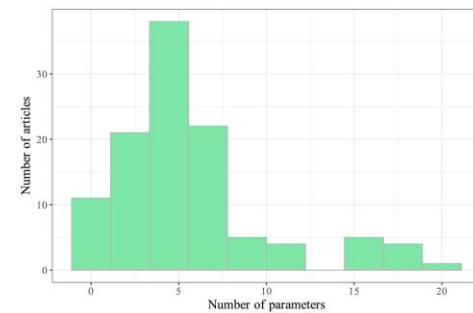
# 04. Quelques résultats



Protocol name	Attributes	Scale type	Reference(s)
Auditory-Perceptual Ratings Instrument for Operatic Singing Voice	Overall vocal performance, appropriate vibrato, resonance balance (chiaroscuro), ring, pitch accuracy, breath management, evenness throughout the range, strain	1 to 10 ordinal scale	Ahmadi et al., 2022 <sup>75</sup>
Bele's protocol	Pitch, pitch range, sonority, variation in loudness, clarity of articulation, overall voice quality, breathiness, hyperfunctional-pressed voice production, hypofunctional-lax voice production, vocal fry-creaky voice production intermittent, vocal fry-creaky voice production at phrase endings, roughness-grating, weak phonation at phrase endings, instability of pitch, soft glottal attacks, ringing voice quality	200mm Visual analog scale for pitch, pitch range, sonority, variation in loudness, clarity of articulation, overall voice quality and 100mm visual analog scale for all the other parameters	Bele (2005 and 2013) <sup>71, 106</sup>
BoDyS (Bogenhausen Dysarthria Scales Speech dimensions)	Respiration, pitch/loudness, voice quality, voice stability, articulation, nasal resonance, speaking rate, fluency, prosodic modulation	0 to 4 ordinal scale	Brendel et al., 2013 <sup>11</sup>
Buffalo Voice Profile	Laryngeal tone severity rating, High pitch rating, Low pitch rating, Voice too loud, Voice too soft, Hypernasal, Hyponasal, Oral resonance (throatiness), Breath supply, Hypertensive muscles, Hypotensive muscles, Rate too fast, Rate too slow, Speech intelligibility, Overall voice rating	1 to 5 ordinal scale	Webb et al., 2004 <sup>37</sup>
CAPE-V (Consensus Auditory-Perceptual Evaluation of Voice)	Severity, Roughness, Breathiness, Strain, Pitch, Loudness	100mm visual analog scale	Ehrlich et al., 2018 <sup>4</sup> ; Muss et al., 2010 <sup>5</sup> ; Joshi et al., 2020 <sup>6</sup> ; Helou et al., 2010 <sup>25</sup> ; Duvall et al., 2021 <sup>29</sup> ; Nagle et al., 2022 <sup>63</sup> ; Cohen et al., 2019 <sup>98</sup> ; Walden, 2023 <sup>102</sup> ; Özcebe et al., 2019 <sup>105</sup> ; Karnell et al., 2007 <sup>107</sup> Rocha et al., 2024 <sup>137</sup>

**Table 2. Perceptual hetero-evaluation protocols available in the literature (listed in alphabetical order by name): name, attributes, type of used, type of scale and corresponding reference**

# 04. Quelques résultats

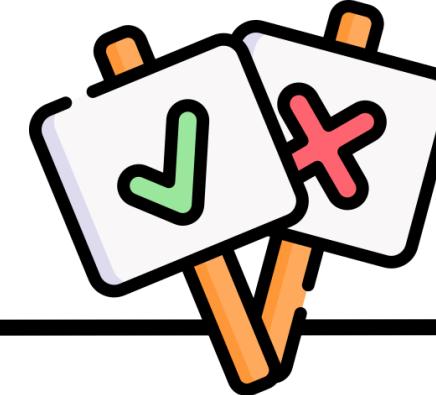


Protocol name <sup>2</sup>	Voice type	Rater(s) (N)	Stimulus(i)	Speakers' Gender (Age range)	Language(s)	Reference(s)
Auditory-Perceptual Ratings Instrument for Operatic Singing Voice	Pathological voice / singed voice: Traditional Iranian singers with primary muscle tension dysphonia (32)	Singing coach (2)	Sustained vowels, singing a traditional Iranian song "Morph-e Sahar"	Male and female (mean 36.75)	Iranian	Ahmadi et al., 2022 <sup>75</sup>
Bele's protocol	Normophonic voice: actors (36) and teachers (35)	SLPs (7) and SLP students (3)	Sustained vowel /a/ at three intensity levels, reading task at two intensity levels	Male (21-73)	Norwegian	Bele (2005 and 2013) <sup>71, 106</sup>
BoDyS	Pathological voice: Friedreich's ataxia (20) / Normophonic voice: Controls (10)	SLP used to BoDyS (2)	Answering three questions, repetition task, reading tasks, induced speech	Male and female (15-71)	German	Brendel et al., 2013 <sup>11</sup>
Buffalo Voice Profile	Pathological voice: Dysphonics (65) / Normophonic voice: Controls (5)	SLPs (7)	Reading task, counting from 1 to 10, reciting the days of the week, sustained vowels /a/ and /i/	Male and female (18-87)	English (England)	Webb et al., 2004 <sup>37</sup>

**Table 3. Implementation of the various protocols : specific types of voice evaluated, gender and age of the speakers, Raters, Nature of the stimuli, Language of administration and corresponding references**

# Concordance intra et inter-évaluateurs

---



## Niveau de fiabilité global

- 41/104 études rapportent à la fois des accords intra **et** inter-juges (d'autres études rapportent parfois intrajuge seulement ou interjuge seulement; d'autres étude 1 seul juge)
- Accord souvent « acceptable », mais très variable selon : l'**attribut** évalué (ex. CAPE-Vf : bonne fiabilité intra pour sévérité/tension, modérée pour souffle/rugosité, inter-juges souvent faible), l'**outil**, la **tâche** utilisée...

## Autres sources de variabilité identifiées:

- Expérience des évaluateurs (formation, SLPs ou étudiants; ancrages perceptifs oui/non, ...)
- Conditions d'enregistrement des productions (qualité « audio » du matériel)
- Définition parfois floue des attributs à évaluer (par exemple: « speech rate » = paramètre de qualité de voix ?)

Relations avec les mesures acoustiques → souvent rapportées mais très hétérogènes, en raison de la diversité des tâches, des attributs et des outils

Plus grande cohérence lorsque perceptif + acoustique sont combinés

# 05. Discussion

---

- 18 outils qui ont un nom et 45 « unnamed »
- GRBAS(I) et CAPE-V dominant : tous deux centrés sur la voix pathologique (populations cliniques)
- 4 à 6 attributs évalués à travers des échelles ordinales 4 points ou VAS 100 mm
- Profil dominant : voix pathologiques, adultes non âgés, hommes et femmes
- Stimuli les plus fréquents : voyelles tenues /a, i, u, lecture courte en anglais MAIS grande diversité (conversation, répétition, comptage, diadochocinésie...)
- 20 langues couvertes mais surreprésentation de l'anglais
- Plus de 250 labels de paramètres regroupés en 25 catégories
- Dimensions les plus fréquentes : sévérité globale, souffle, raucité, tension (« strain »)
- Certains domaines ajoutent des attributs esthétiques (projection, couleur, liberté dans la tessiture...) ou techniques (vibrato, attaques, précision mélodique)
- Fiabilité acceptable mais variable selon les attributs et les conditions d'évaluation (intra > inter)
- Variabilité liée à de multiples facteurs (formation, tâches, matériel) : intéressante car elle est désormais connue, peut être identifiée et dans une certaine mesure, contrôlée

Q1: les outils existants

Q2: types de voix et évaluateurs

Q3: caractéristiques des attributs

Q4: inter-intra-juges

# 05. Conclusion & perspectives

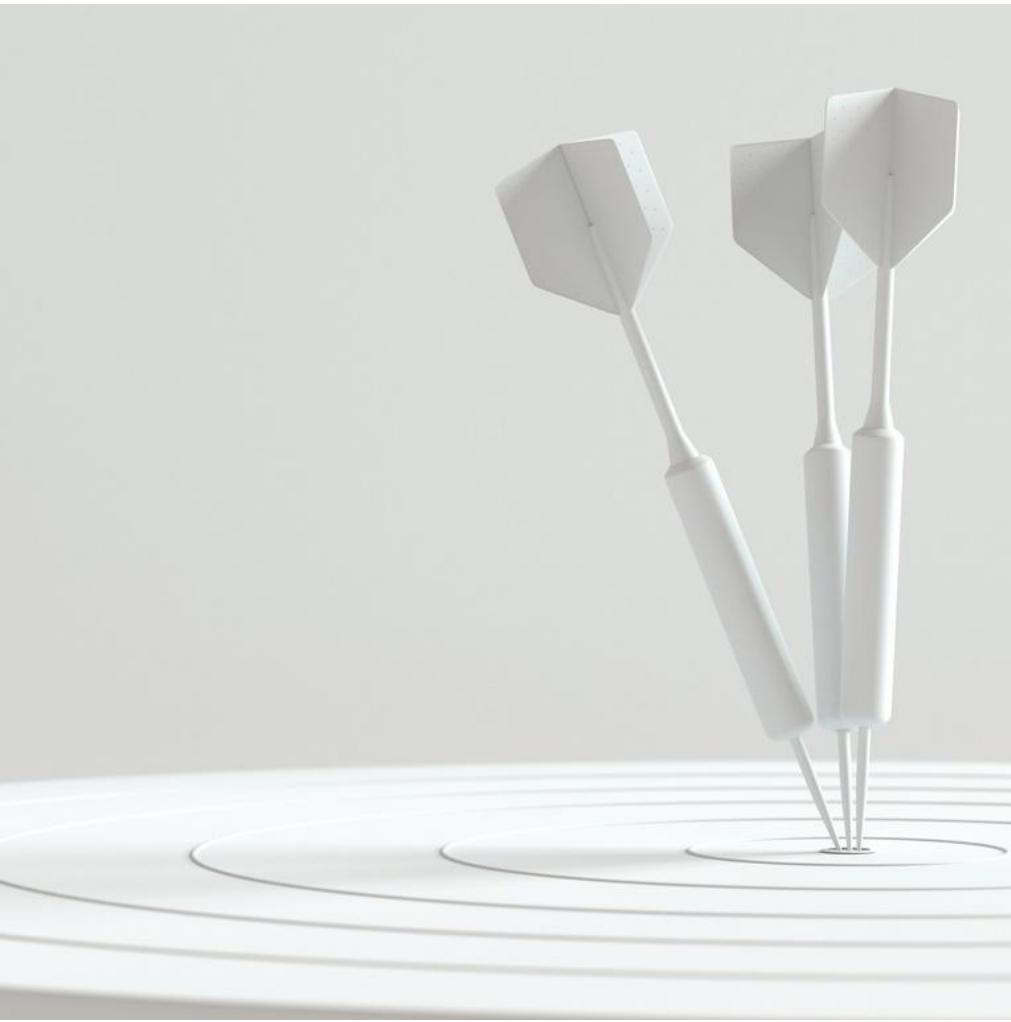
---

## Constat à la suite de la scoping review

- (hétéro)évaluation perceptive: pratique courante mais qui présente des limites (ex : accès au matériel)
- Fiabilité variable → liée aux “prototypes” (representations mentales des évaluateurs de la “qualité”) et à un manque de formation
- “*(...) all these observations highlight nevertheless the relevance of offering guidance for clinical and pedagogical practice, particularly by outlining key criteria to consider when selecting an appropriate evaluation tool.*” (extrait du manuscript)

## Perspectives

- Importance de la formation des évaluateurs et de l'utilisation “d'ancrage” perceptifs (modèles)
- Variabilité intéressante mais besoin d'une standardisation de nouveaux outils pour certains types de voix (identifier l'adéquation de certains protocoles et leur modalité d'utilisation avec certaines pathologies)
- Populations et types de voix sous-étudiés (enfants, personnes âgées normophoniques, voix chantée, ...)
- Objectif final : disposer de protocoles plus robustes, comparables et adaptés aux besoins cliniques, pédagogiques et de recherche.



# Merci

---

## Cartographie des protocoles perceptifs d'hétéro-évaluation de la qualité vocale : voix normophoniques, pathologiques et chantées

Lola Terny<sup>1</sup>, Nicolas Audibert<sup>2</sup>, Lise Crevier-Buchman<sup>2</sup>, Véronique Delvaux<sup>1</sup>, Sophie Fagniart<sup>1</sup>, Bernard Harmegnies<sup>1</sup>, Kathy Huet<sup>1</sup>, Myriam Piccaluga<sup>1</sup>, Reina Remman<sup>2</sup>, Virginie Roland<sup>1</sup>, Claire Pillot-Loiseau<sup>2</sup>, Amélia Pettirossi<sup>2</sup>

<sup>1</sup>Service de Métrologie et Sciences du Langage (Université de Mons, Belgique)

<sup>2</sup>LPP, UMR 7018, CNRS (Université Paris 3-Sorbonne Nouvelle, France)

**UMONS**  
University of Mons

Métrologie  
Sciences  
Langage



UNIVERSITÉ  
**SORBONNE**  
NOUVELLE  
PARIS 3