



HAL
open science

Efficacité de la reconnaissance automatique de phonèmes sur une tâche de production de pseudo-mots

Maëlys Leïçarrague, Tuan Nguyen, Alain Ghio, Julien Pinquier, Muriel Lalain, Virginie Woisard, Clémence Devoucoux, Corinne Fredouille, Mathieu Balaguer

► **To cite this version:**

Maëlys Leïçarrague, Tuan Nguyen, Alain Ghio, Julien Pinquier, Muriel Lalain, et al.. Efficacité de la reconnaissance automatique de phonèmes sur une tâche de production de pseudo-mots. 10èmes Journées de Phonétique Clinique, Jun 2025, Sète, France. <hal-05142338>

HAL Id: hal-05142338

<https://hal.science/hal-05142338v1>

Submitted on 3 Jul 2025

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 CC BY 4.0 - Attribution - International License

Efficacité de la reconnaissance automatique de phonèmes sur une tâche de production de pseudo-mots

Maëlys Leïçarrague^(1,2), Tuan Nguyen⁽³⁾, Alain Ghio⁽⁴⁾, Julien Pinquier⁽¹⁾, Muriel Lalain⁽⁴⁾, Virginie Woisard^(5,6), Clémence Devoucoux⁽⁶⁾, Corinne Fredouille⁽³⁾, Mathieu Balaguer⁽¹⁾

(1) IRIT, Université de Toulouse, CNRS, Toulouse, France – (2) Centre de Formation Universitaire en Orthophonie, Faculté de Santé, Université de Toulouse, France – (3) Avignon Université, Laboratoire Informatique d'Avignon, France – (4) Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France – (5) UT2J, Laboratoire de Neuro-Psycho-Linguistique, Toulouse, France – (6) Hôpital Larrey, CHU de Toulouse, France

INTRODUCTION

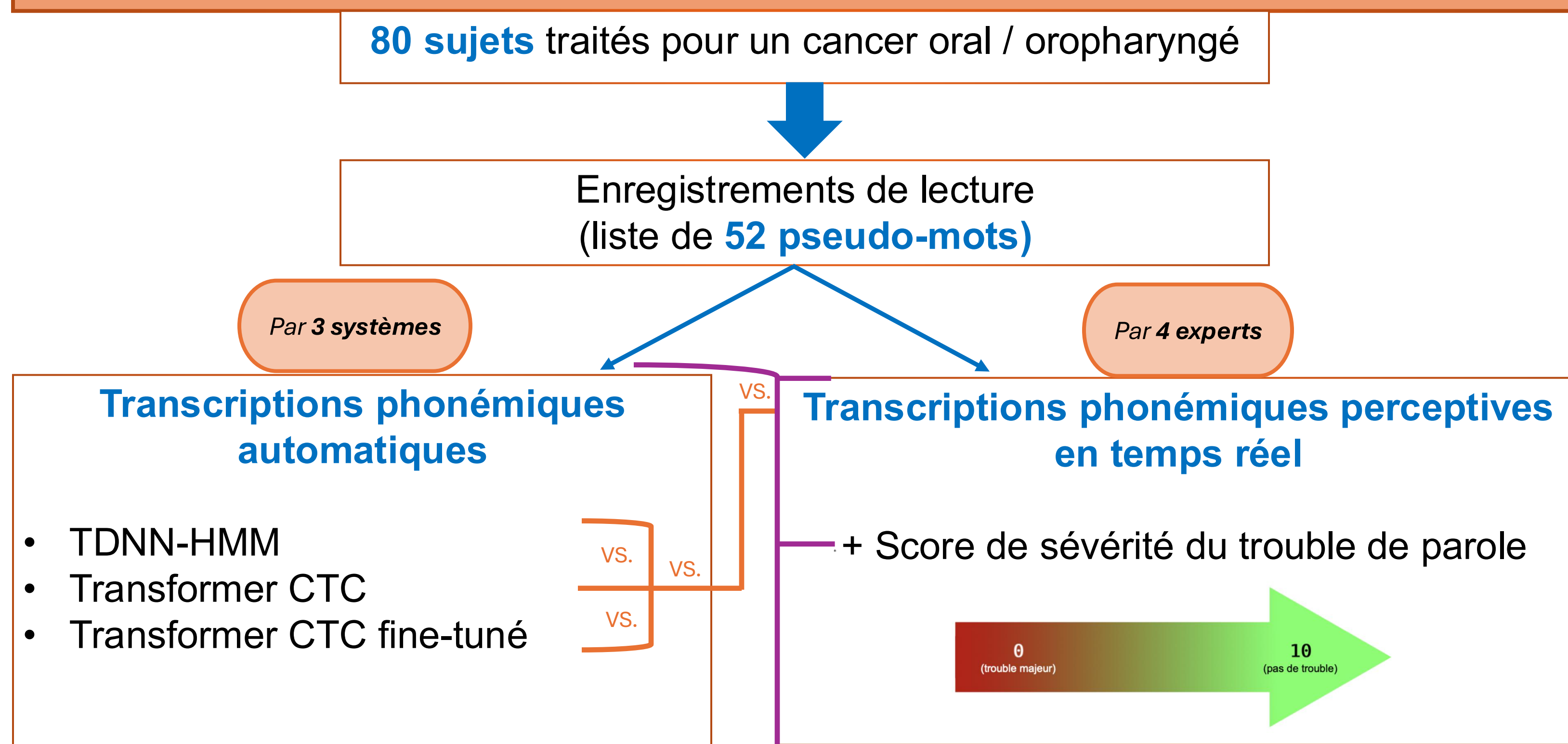
- Les traitements du cancer oral / oropharyngé induisent des **modifications anatomiques**
- Ces altérations affectent la production de la parole au niveau **segmental** [1]
- Elles motivent l'analyse des liens entre **anatomie et reconnaissance des phonèmes** [2]
- Les **systèmes automatiques de décodage acoustico-phonétique** permettent cette analyse
- Leur **efficacité** doit être évaluée dans un contexte de parole pathologique [3]

OBJECTIF

Étudier l'**efficacité*** de la reconnaissance automatique phonémique sur une tâche de production de pseudo-mots après cancer oral / oropharyngé

***Efficacité** = reconnaissance par rapport à une cible, malgré les erreurs de production / perception, et non par rapport à une référence humaine

MÉTHODE



Statistiques : test de Friedman puis comparaison paire à paire (Durbin-Conover)

- 1 Comparaison de l'efficacité des 4 systèmes entre eux (3 automatiques et 1 perceptif)
- 2 Comparaison de l'efficacité des 4 systèmes selon les groupes de sévérité

Métriques d'efficacité de reconnaissance (Conditions : pseudo-mots isolés et sans doublon)

• **PER** (Taux d'erreur de phonèmes)
=> Proportion d'erreurs

$$PER = \frac{\text{Délétions} + \text{Substitutions} + \text{Insertions}}{\text{nombre de phonèmes total du pseudomot cible}}$$

• **PPD** (Score de déviation phonologique [4])
=> Sévérité des erreurs

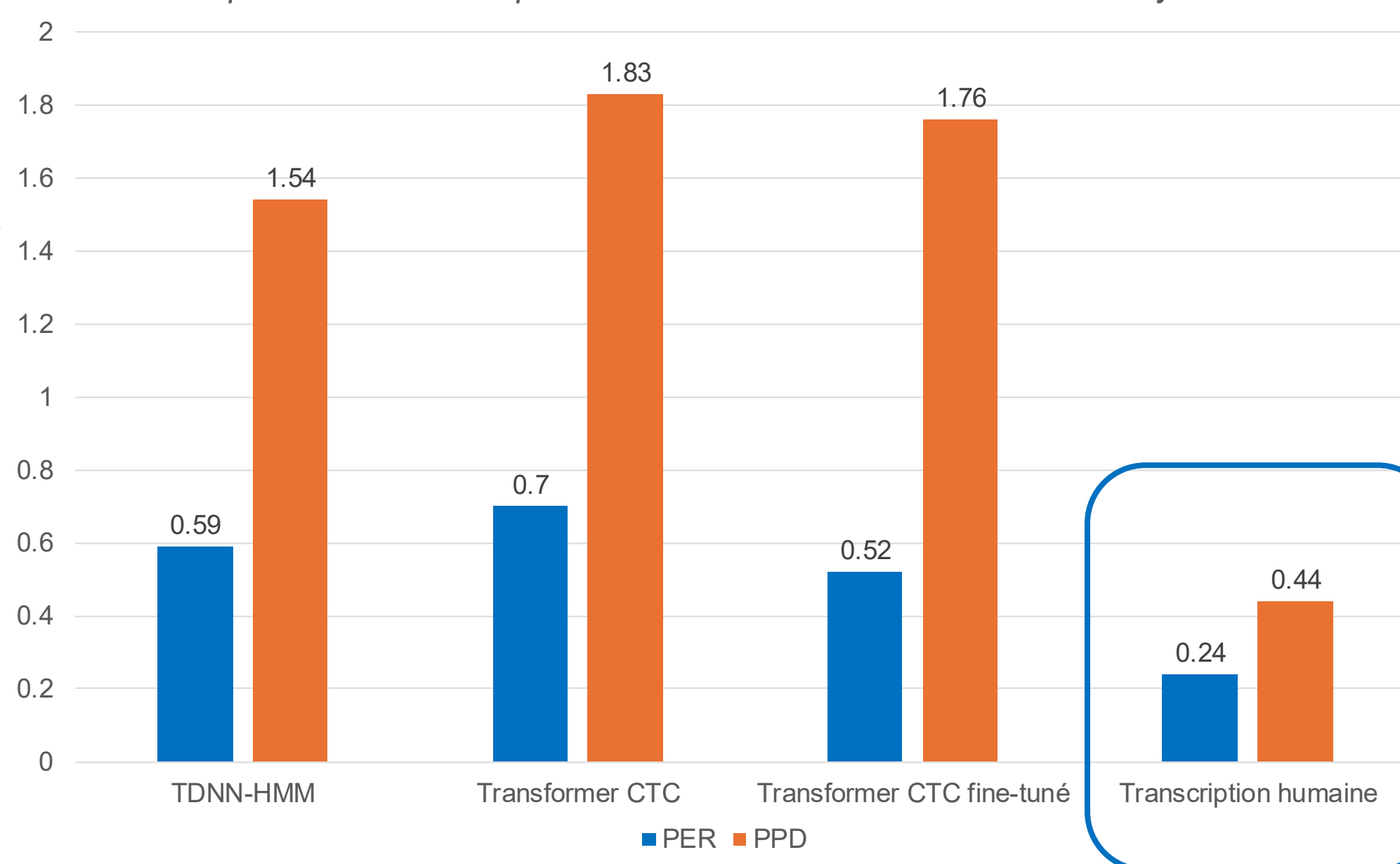
$$PPD = \text{Nombre de traits phonologiques d'écart moyen par phonème entre la reconnaissance et la cible}$$

Métriques faibles (moyennes proches de 0) = Meilleure efficacité de reconnaissance

Corpus final :
79 sujets
4107 pseudo-mots
(exclusions)

RÉSULTATS

Comparaison des métriques d'efficacité de reconnaissance des 4 systèmes



Meilleure efficacité de reconnaissance

- PER et PPD les plus faibles : **Transcription humaine**
- Parmi les systèmes automatiques
- PER le plus faible : **Transformer CTC fine-tuné**
- PPD le plus faible : **TDNN-HMM**

1 Comparaison de l'efficacité des 4 systèmes entre eux

Systèmes	PER	PPD
TDNN-HMM	0,59	1,54
Transformer CTC	0,70	1,83
Transformer CTC fine-tuné	0,52	1,76
Transcription humaine	0,24	0,44

Différence significative entre les PER et PPD de tous les systèmes ($p < 0,001$)

2 Comparaison de l'efficacité des 4 systèmes selon les groupes de sévérité

Sévérité	Systèmes	PER	PPD
0-6 (n = 30)	TDNN-HMM	0,77	2,15
	Transformer CTC	0,86	2,28
	Transformer CTC fine-tuné	0,68	2,36
	Transcription humaine	0,36	0,63
7-10 (n = 36)	TDNN-HMM	0,42	1,01
	Transformer CTC	0,57	1,43
	Transformer CTC fine-tuné	0,37	1,21
	Transcription humaine	0,15	0,25
0-9 (n = 18)	TDNN-HMM	0,69	1,87
	Transformer CTC	0,79	2,07
	Transformer CTC fine-tuné	0,61	2,09
	Transcription humaine	0,29	0,51
10 (n = 48)	TDNN-HMM	0,28	0,63
	Transformer CTC	0,47	1,13
	Transformer CTC fine-tuné	0,24	0,77
	Transcription humaine	0,12	0,20

Meilleure efficacité de reconnaissance selon la sévérité

- PER et PPD les plus faibles : **troubles peu sévères (7-10)** pour automatique et humain
- Automatique **converge vers l'humain** quand on isole les sujets peu sévères (10)

DISCUSSION

- Les **systèmes de décodage acoustico-phonétique automatiques** appliqués à la parole après cancer sont **moins efficaces que les experts humains**
- Causes possibles : **architecture des systèmes** → **insertions fréquentes ; entraînement sur de la parole « saine » ; événements sonores dans le signal ; pas d'effet de restauration phonémique (contrairement aux humains) ; ...**
- Perspectives : poursuivre la réflexion sur le **lien entre les erreurs retrouvées et les déficits analytiques des patients**

RÉFÉRENCES

- [1] Mlynarek, A., Rieger, J., Harris, J., O'Connell, D., Al-Qahtani, K., Ansari, K., Chau, J., & Seikaly, H. (2008). Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. *Journal of Otolaryngology - Head & Neck Surgery*, 37(1), 2-10. <https://doi.org/10.2310/7070.2008.1001>
- [2] Lazarus, C. L., Husaini, H., Anand, S. M., Jacobson, A. S., Mojica, J. K., Buchbinder, D., & Urken, M. L. (2013). Tongue Strength as a Predictor of Functional Outcomes and Quality of Life after Tongue Cancer Surgery. *Annals of Otolaryngology, Rhinology & Laryngology*, 122(6), 386-397. <https://doi.org/10.1177/000348941312200608>
- [3] Halpern, B. M., Feng, S., van Son, R., van den Brekel, M., & Scharenborg, O. (2022). Low-resource automatic speech recognition and error analyses of oral cancer speech. *Speech Communication*, 141, 14-27. <https://doi.org/10.1016/j.specom.2022.04.006>
- [4] Ghio, A., Lalain, M., Rebourg, M., Marczyk, A., Fredouille, C., & Woisard, V. (2022). Validation of an Intelligibility Test Based on Acoustic-Phonetic Decoding of Pseudo- Words: Overall Results from Patients with Cancer of the Oral Cavity and the Oropharynx. *Folia Phoniatrica et Logopaedica*, 74(3), 209-222. <https://doi.org/10.1159/000519427>