



HAL
open science

Modélisation individuelle de juges pour la prédiction automatique de l'intelligibilité de la parole

Sebastião Quintas, Julie Mauclair, Julien Pinquier

► **To cite this version:**

Sebastião Quintas, Julie Mauclair, Julien Pinquier. Modélisation individuelle de juges pour la prédiction automatique de l'intelligibilité de la parole. 9èmes Journées de Phonétique Clinique (JPC 2023), Université Toulouse III Paul Sabatier; Université UT2J Toulouse Jean-Jaurès; Hôpitaux de Toulouse; Laboratoire Informatique d'Avignon, Jun 2023, Toulouse, France. hal-04093783

HAL Id: hal-04093783

<https://hal.science/hal-04093783v1>

Submitted on 10 May 2023

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.

Modélisation individuelle de juges pour la prédiction automatique de l'intelligibilité de la parole

Sebastião QUINTAS
Julie MAUCLAIR
Julien PINQUIER

IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3,
Toulouse, France

Le développement d'outils automatiques pour la prédiction de l'intelligibilité de la parole peut être vu comme une alternative plus robuste, reproductible et objective face aux évaluations perceptives réalisées en clinique. En effet, ces évaluations peuvent être conditionnées par une variété d'aspects, tels que la connaissance préalable du patient ou la tâche d'évaluation de la parole elle-même (Balaguer et al., 2019). Les travaux de (De Bodt et al., 2002) ont montré que l'intelligibilité de la parole est liée à quatre paramètres perceptifs : Voix (V), Résonance (R), Prosodie (P) et Distorsions Phonémiques (PD), mais avec des niveaux de pertinence variables en fonction de chaque juge. Vu que l'intelligibilité est une mesure intrinsèquement subjective, la modélisation des juges individuels devient pertinente dans le contexte de l'intelligibilité de la parole (Woisard et al., 2021).

Dans cette étude, nous faisons la modélisation d'un ensemble de juges perceptifs, utilisés pour la prédiction de l'intelligibilité dans le cadre du corpus français du cancer des voies aérodigestives supérieures (Woisard et al., 2020). Nous nous basons sur le paradigme des *X-vector speaker embeddings* (Quintas et al., 2020) afin de proposer la modélisation illustrée sur la figure 1. Chaque modèle individuel reçoit en entrée une représentation, qui est ensuite transmise à un réseau neuronal peu profond. Ce réseau prédit les quatre paramètres perceptifs : Voix (V), Résonance (R), Prosodie (P) et Distorsions Phonémiques (PD). Ensuite, l'intelligibilité est prédite en fonction de ces quatre paramètres, en utilisant une moyenne simple (poids uniformes) ou pondérée, qui utilise des poids différents pour chaque paramètre. Pour chercher les poids optimaux

pour la moyenne pondérée, une méthode par grille de recherche (*grid search*) a été conçue.

Dans le tableau 1, nous pouvons observer deux analyses distinctes. La première montre la prédiction d'intelligibilité en utilisant les quatre paramètres perceptifs de chaque juge. La seconde montre exactement la même prédiction avec la modélisation automatique présentée précédemment. Malgré les corrélations inférieures dans la modélisation automatique de chaque juge, la différence entre les résultats obtenus avec les poids uniformes et les poids *grid search* est plus faible que dans le cas de l'analyse perceptive. Puisqu'il existe différents profils parmi les différents juges perceptifs (De Bodt et al., 2002), nous pouvons dire que la modélisation automatique favorise une prédiction, non seulement plus homogène entre les différents paramètres utilisés, mais également plus explicable. L'explicabilité des systèmes automatiques et des mesures produites devrait permettre une meilleure acceptabilité par les cliniciens.

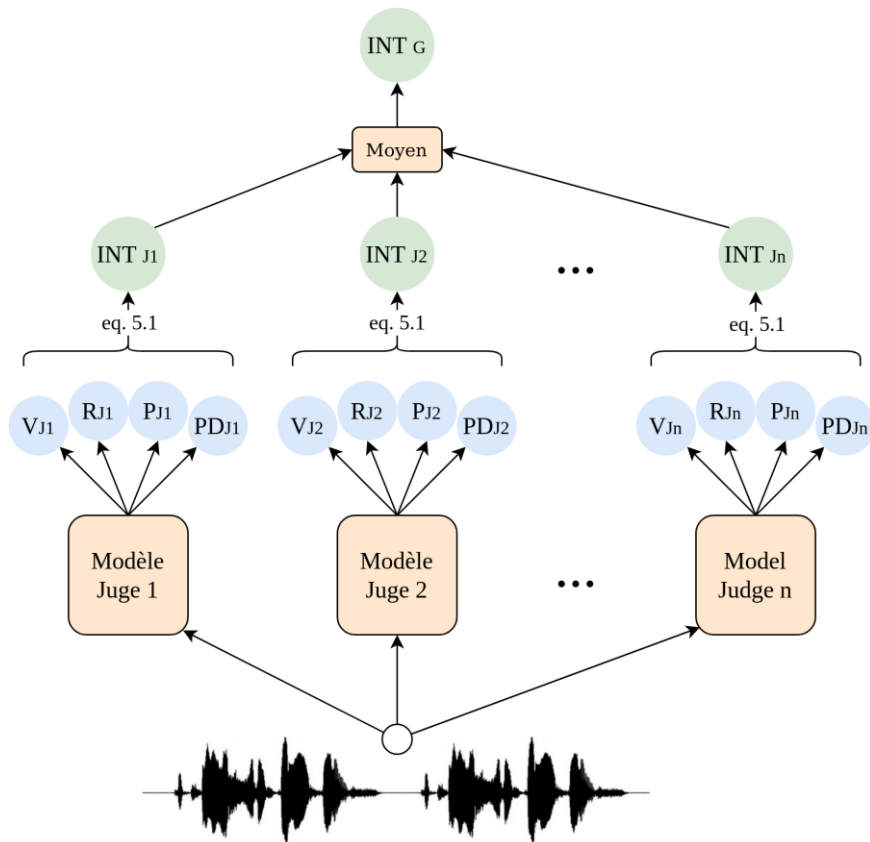


Figure 1. Schéma de la modélisation individuelle de chaque juge perceptif.

Tableau 1. Corrélations obtenues en utilisant des moyennes des 4 paramètres perceptifs.
 Comparaison entre l'analyse perceptive et la modélisation automatique des juges.

Juge	Analyse perceptive des juges		Modélisation automatique des juges	
	Poids Uniformes	Poids <i>Grid Search</i>	Poids Uniformes	Poids <i>Grid Search</i>
1	0,828	0,851	0,613	0,630
2	0,774	0,852	0,699	0,728
3	0,836	0,879	0,613	0,633
4	0,767	0,863	0,726	0,742
5	0,648	0,784	0,490	0,539
6	0,750	0,810	0,637	0,649

Références bibliographiques

- Balaguer, M., Pommée, T., Farinas J., Pinquier, J., Woisard V., & Speyer, R. (2019). Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review. *Journal of the Sciences and Specialities of Head and Neck*.
- Quintas, S., Mauclair, J., Woisard, V., & Pinquier, J. (2020). Automatic Prediction of Speech Intelligibility Based on X-Vectors in the Context of Head and Neck Cancer, *Proceedings of Interspeech*.
- De Bodt, M. S., Hernández-Diaz Huici, M. E., & Van De Heyning, P. H (2002). Intelligibility as a Linear Combination of Dimensions in Dysarthric Speech. *Journal of Communication Disorders*.
- Woisard, V., Balaguer, M., Fredouille, C., & al. (2021). Construction of an automatic score for the evaluation of speech disorders among patients treated for a cancer of the oral cavity or the oropharynx: The Carcinologic Speech Severity Index, *Head & Neck*.
- Woisard, V., & al. (2020). C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers, *Language Resources Evaluation*.