



HAL
open science

Une approche granulaire 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. Une approche granulaire pour la prédiction automatique de l'intelligibilité de la parole. 9èmes Journées de Phonétique Clinique (JPC 2023), Jun 2023, Toulouse, France. pp.75–76. hal-04313911

HAL Id: hal-04313911

<https://hal.science/hal-04313911>

Submitted on 1 Dec 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.

Une approche granulaire pour la prédiction automatique de l'intelligibilité de la parole

Sebastião QUINTAS¹
Julie MAUCLAIR¹
Julien PINQUIER¹

¹IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France

La prédiction automatique de l'intelligibilité de la parole peut être considérée comme une alternative pertinente aux évaluations perceptives réalisées en clinique. Ces évaluations sont connues pour être subjectives, biaisées et variables, puisque l'évaluation peut être conditionnée 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). Dans ces conditions, le développement d'une approche automatique peut être considéré comme une alternative plus robuste, reproductible et objective.

Cependant, les évaluations automatiques peuvent être difficiles à interpréter, un aspect qui n'aide pas à l'adaptation clinique de celles-ci. Ainsi, dans cette étude, nous présentons une approche automatique de prédiction de l'intelligibilité, en utilisant les niveaux de granularité de la phrase, du mot et du phonème. Au niveau des phrases, le paradigme des *X-vector speaker embeddings* est utilisé (Quintas et al., 2020). Au niveau des mots, un mécanisme d'auto-attention est appliqué à la tâche de lecture des pseudo-mots du même corpus (Quintas et al., 2022a). Finalement, au niveau des phonèmes, un réseau siamois est appliqué pour prédire un score d'intelligibilité en utilisant la similarité phonémique (Quintas et al., 2022b). La fusion de ces trois systèmes, voir *Figure 1*, crée une couche supplémentaire d'interprétation de grande pertinence clinique. Ce travail est appliqué au corpus français du cancer des voies aérodigestives supérieures (Woisard et al., 2020). Les scores d'intelligibilité de référence, obtenus dans le cadre de ce corpus, correspondent à l'évaluation perceptive de six juges professionnels. Ils sont compris entre 0 (inintelligible) et 10 (parfaitement intelligible).

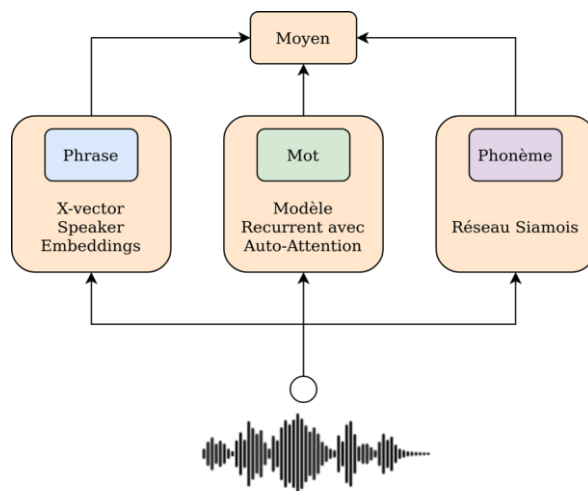


Figure 1. Schéma de l'approche granulaire proposée et de la fusion

Les résultats, illustrés dans le *Tableau 1*, montrent la corrélation et l'erreur (RMSE) obtenue avec chaque système ainsi que leur combinaison. Individuellement, il est possible d'obtenir de hautes corrélations pour chaque système. Toutefois, avec leur fusion, le système est encore plus performant. La comparaison avec un système « oracle » (qui choisit manuellement le meilleur système pour chaque patient) montre que des améliorations sont encore possibles, notamment au niveau de l'erreur. Cette différence est principalement due à la sous-performance de notre système sur les patients à faible intelligibilité, une classe sous-représentée dans le corpus.

L'approche granulaire permet l'analyse de plusieurs niveaux qui sont intéressants pour les cliniciens et permet l'interprétabilité par des humains du fonctionnement du système automatique. Si le système automatique fournit des scores interprétables, cela favorisera son acceptabilité par les thérapeutes.

Tableau 1. Corrélations (Spearman) et erreur quadratique moyenne (RMSE) obtenues avec les systèmes individuels ainsi que leur fusion

Granularité	Système		RMSE
Phrase	X-Vector Speaker Embeddings (Quintas, S. 2020)	0,85	1,623
Mot	Modèle récurrent avec auto-attention (Quintas, S. 2022a)	0,82	1,548
Phonème	Réseau Siamois (Quintas, S. 2022b)	0,89	2,080
Fusion des 3 systèmes (score moyen)		0,91	1,588
<i>Oracle (Choix manuel du meilleur score)</i>		<i>0,92</i>	<i>1,075</i>

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. *Head & neck*, 42(1), 111-130. <https://doi.org/10.1002/hed.25949>
- Woisard, V., Astésano, C., Balaguer, M., Farinas, J., Fredouille, C., Gaillard, P., Ghio, A., Giusti, L., Laaridh, I., Lalain, M., Lepage, B., Mauclair, J., Nocaudie, O., Pinquier, J., Pouchoulin, G., Puech, M., Robert, D., & Roger, V. (2021). C2SI corpus : a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers. *Language Resources and Evaluation*, 55(1), 173-190. <https://doi.org/10.1007/s10579-020-09496-3>
- Quintas, S., Abad, A., Mauclair, J., Woisard, V., Pinquier, J. (2022a). Utilisation de modèles transformers pour la prédiction de l'intelligibilité de la parole de patients atteints de cancers des voies aérodigestives supérieures. Journées d'Études sur la parole.
- Quintas, S., Mauclair, J., Woisard, V., & Pinquier, J. (2022b). Automatic Assessment of Speech Intelligibility using Consonant Similarity for Head and Neck Cancer. Proceedings of Interspeech.
- 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.