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

Sebastião Quintas, Julie Mauclair et Julien Pinquier

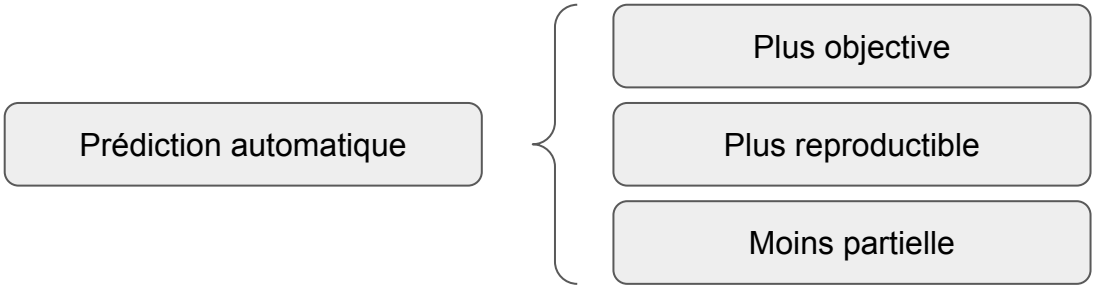
IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3

Sommaire

- Introduction
- Méthodologie
  - Approche granulaire
  - Fusion
- Expériences
  - Corpus C2SI
  - Résultats
- Discussion
- Conclusions

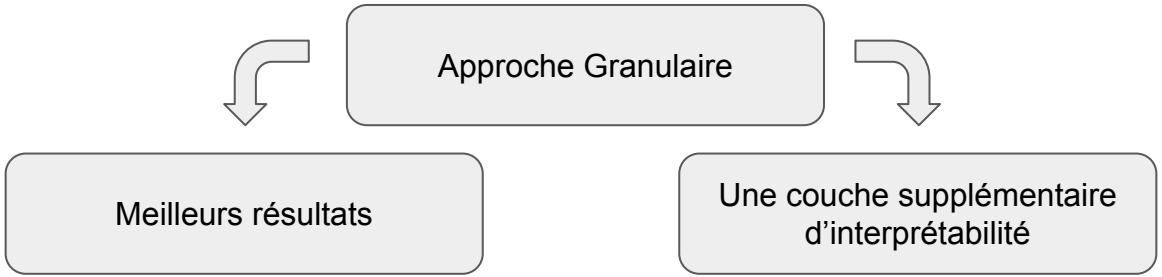
Introduction

- L'évaluation perceptive de l'intelligibilité de la parole est la méthode la plus courante d'évaluation des troubles affectant la parole tels que les cancers ORL
- Néanmoins, ces évaluations sont connues pour être **subjectives**, **biaisées** et **variables**
    - 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
- De cette façon, la **prédiction automatique de l'intelligibilité** de la parole peut-être vue comme une alternative pertinente aux évaluations perceptives cliniques



Introduction

- D’autre part, les évaluations automatiques peuvent être difficiles à interpréter, un aspect qui n’aide pas à l'adaptation clinique de celles-ci
- De cette façon, dans ce travail nous présentons une approche automatique de prédiction de l’intelligibilité, en utilisant les niveaux de granularité de la **phrase**, du **mot** et du **phonème**



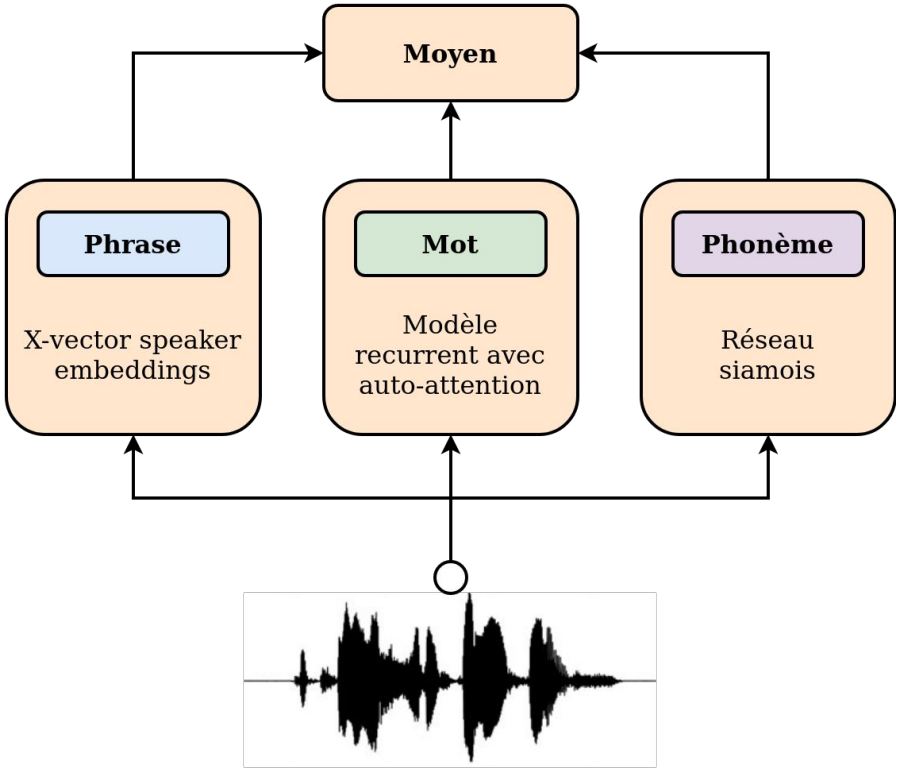
Sommaire

- Introduction
- **Méthodologie**
  - **Approche granulaire**
  - **Fusion**
- Expériences
  - Corpus C2SI
  - Résultats
- Discussion
- Conclusions

Méthodologie – Approche granulaire

- Pour chaque niveau, nous utilisons des systèmes déjà adaptés pour la tâche de prédiction de l'intelligibilité – basés sur nos travaux précédents
- Pour le niveau **phrase,** nous utilisons le paradigme des x-vector-speaker embeddings
  - ↳ Ces représentations sont utilisées comme attributs en entrée d'un réseau de neurones adapté pour une prédiction de l'intelligibilité
- Pour le niveau **mot**, un système récurrent avec auto-attention est utilisé
  - ↳ Ce système fait la prédiction de l'intelligibilité en fonction d'un ensemble de pseudo-mots en entrée
- Pour le niveau **phonème**, nous utilisons le paradigme de similarité des consonnes
  - ↳ Un réseau siamois est adapté pour prédire une mesure d'intelligibilité en fonction de la quantité de consonnes similaires. "Plus la quantité de consonnes similaires aux sujets sains est importante, plus l'intelligibilité est élevée"

Méthodologie – Fusion



Sommaire

- Introduction
- Méthodologie
  - Approche granulaire
  - Fusion
- **Expériences**
  - **Corpus C2SI**
  - **Résultats**
- Discussion
- Conclusions

Expériences – Corpus C2SI

- Toutes nos expériences sont basées sur le corpus C2SI [1]. Chaque sous-système granulaire est entraîné en utilisant des tâches différentes du même corpus
- Les systèmes utilisés étaient tous adaptés à la prédiction de la même mesure d'intelligibilité, basée sur le jugement perceptif de 6 juges différents, compris entre **0 (inintelligible) et 10 (parole saine)**
- Un total de **108 locuteurs** est utilisé dans cette analyse
    - Des méthodes d'imputation ont été utilisées pour les locuteurs qui n'ont pas effectués toutes les tâches

[1] C. Astésano, M. Balaguer, J. Farinas, C. Fredouille, A. Ghio, P. Gaillard, L. G. I. Laaridh, M. Lalain, B. Lepage, and et al., "Carcinologic speech severity index project: A database of speech disorder productions to assess quality of life related to speech after cancer," Language Resources and Evaluation Conference, 2018.

Expériences – Corpus C2SI

Granularité	Phrase	Mot	Phonème
Tâche utilisée	Lecture (Mr. Seguin)	Pseudo-Mots [2]	

Lecture de Passage

(S1) Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres.
(S2) Il les perdait toutes de la même façon.
(S3) Un beau matin, elles cassaient leur corde,
(S4) s'en allaient dans la montagne,et là-haut le loup les mangeait.
(S5) Ni les caresses de leur maître
(S6) ni la peur du loup, rien ne les retenait.
(S7) C'était paraît-il des chèvres indépendantes
(S8) voulant à tout prix le grand air et la liberté.

Pseudo-Mots (example d'un ensemble)

banfou bleja boucti brimpli chessant choniou clifant cogu crimpin daillu dinrant dredi fanrsi flinrpu fouma fravi gabi glunou grorvo guchin joutu juro lanvin lerda messo mouco nianlo niejo noksa nouillou pastu pidant ploniou pripin psila quiga rinta rurnu sanvrin scuna souquin spaclant sticho tangri tougzu tradrou virjant vumou yainzi yaltin zebou zouzant

[2] M. Lalain, A. Ghio, L. Giusti, D. Robert, C. Fredouille, and V. Woisard, "Design and development of a speech intelligibility test based on pseudowords in french: Why and how?" Journal of Speech, Language and Hearing Research, 2020.

Expériences – Résultats

Granularité	Système	ρ	RMSE
Phrase	*X-vector* speaker embeddings [3]	0,85	1,623
Mot	Modèle récurrent avec auto-attention [4]	0,82	1,548
Phonème	Réseau siamois [5]	0,89	2,080
Fusion des 3 systèmes (score moyen)		**0,91**	**1,588**
Oracle (choix manuel du meilleur score)		0,92	1,075

[3] S. Quintas, J. Mauclair, V. Woisard, J. Pinquier, “Automatic prediction of speech intelligibility using x-vectors in the context of head and neck cancer”, Proceedings of Interspeech 2020

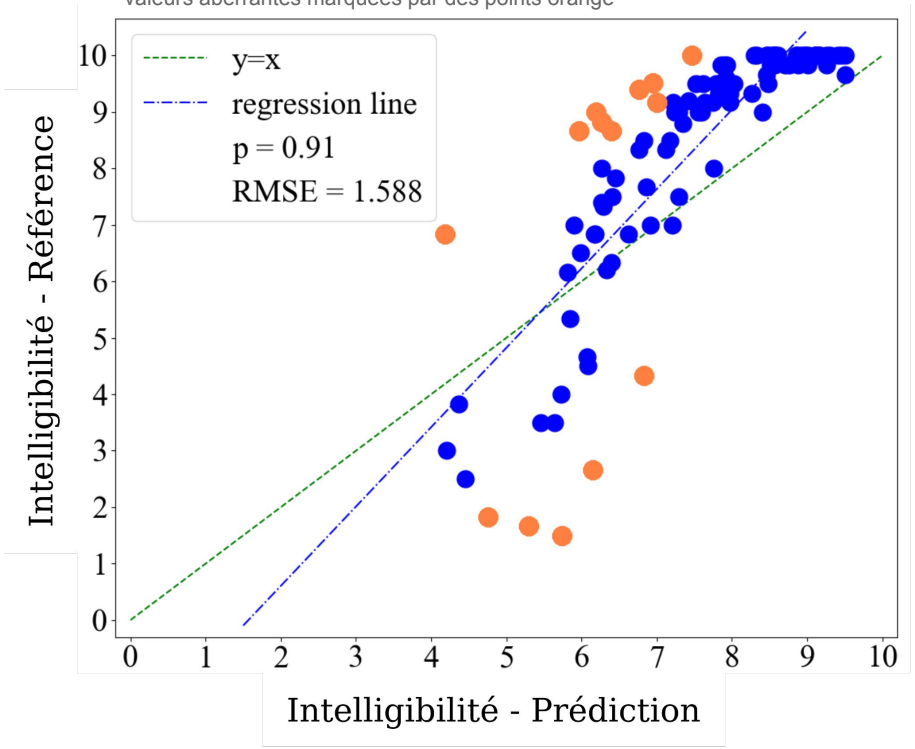
[4] S. Quintas, A. Abad, J. Mauclair, V. Woisard, J. Pinquier, “Towards reducing patient effort for the automatic prediction of speech intelligibility in head and neck cancers”, Proceedings of ICASSP 2023

[5] S. Quintas, J. Mauclair, V. Woisard, J. Pinquier, “Automatic assessment of speech intelligibility using consonant similarity for head and neck cancer”, Proceedings of Interspeech 2022

Expériences – Résultats

- Les valeurs aberrantes ont été prises en compte pour les valeurs en dehors d'une limite [−2, T, 2] (T représente la valeur cible)
- Un total de 15 valeurs aberrantes ont été trouvées (points oranges sur le graphique)
- Ces valeurs aberrantes expliquent la valeur d'erreur plus élevée par rapport à l'approche "oracle"

*valeurs aberrantes marquées par des points orange



Sommaire

- Introduction
- Méthodologie
  - Approche granulaire
  - Fusion
- Expériences
  - Corpus C2SI
  - Résultats
- **Discussion**
- **Conclusions**

Discussion

- La comparaison avec un système « oracle » (qui choisit manuellement le meilleur système pour chaque patient) montre que des améliorations au niveau de l'erreur sont encore possibles
- 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
- Curieusement, malgré des performances inférieures du système au niveau du phonème, celui-ci est le plus compétitif pour les patients à faible intelligibilité
    - Cet aspect est directement lié au type d'entraînement (réseau siamois, similarité des consonnes)
    - C'est aussi le seul système que n'a pas utilisé les scores perceptifs comme référence

Conclusions

- Outre la bonne performance de chaque système individuel, la fusion permet d'avoir une corrélation plus haute (ρ=0,91) ainsi qu'une erreur plus basse (RMSE=1,588)
- L'approche granulaire permet l'analyse de plusieurs niveaux qui sont intéressants pour les cliniciens et thérapeutes
- La même approche permet aussi une couche supplémentaire d'interprétabilité par des humains du fonctionnement du système automatique
    - Le score automatique est directement lié à la valeur de chaque niveau (phrase, mot et phonème)
    - Plus le score automatique est interprétable, plus l'acceptabilité clinique sera possible

Merci !