Analyse des performances de systèmes de reconnaissance automatique de la parole spontanée après cancer oral ou oropharyngé

Mathieu Balaguer (1), Julien Pinquier (1), Jérôme Farinas (1), Virginie Woisard (2,3)

(1) IRIT, Université de Toulouse, CNRS, UT3, Toulouse, France – (2) CHU Larrey, Toulouse, France – (3) Laboratoire de Neuro-Psycho-Linguistique LNPL, Université Toulouse II, France

Introduction

Analyse de la **parole spontanée** des patients traités pour un cancer de la cavité buccale ou de l'oropharynx

- **Essentielle en clinique** courante : situation de production la plus « écologique » [1,2]
- Mais nécessite de pouvoir s'appuyer sur une **transcription fiable**

Systèmes de reconnaissance automatique de parole (RAP)

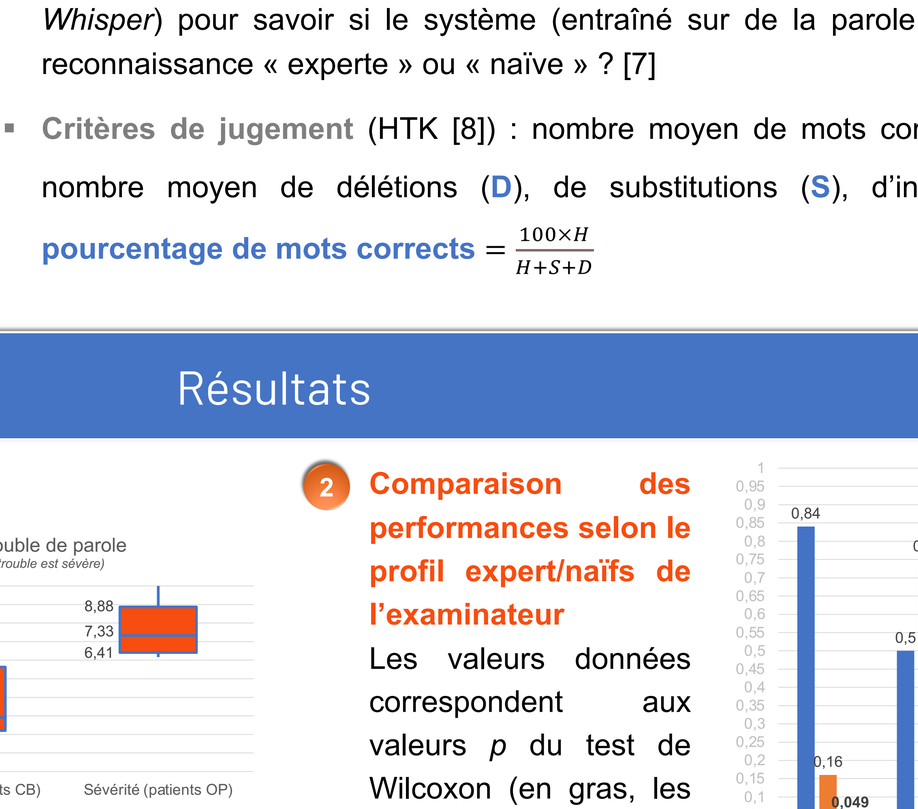
- **Utiles** pour
  - Faciliter la transcription
  - Autoriser des analyses plus précises et complètes du discours des patients : niveaux lexical, sémantique, discursif…
  - Accéder à de **nouvelles informations sur la dynamique psychosociale** des patients.
- Mais sont **entraînés sur la parole typique**
  - Mauvaise prise en compte des spécificités de la parole après cancer
  - Taux d'erreurs mots très élevé [3]

Objectif

Étudier les performances de systèmes de reconnaissance automatique de parole au niveau du mot, appliqués à la parole spontanée après cancer.

Matériel et méthode

Enregistrements (entretiens semi-dirigés) de la **parole spontanée** de 12 patients



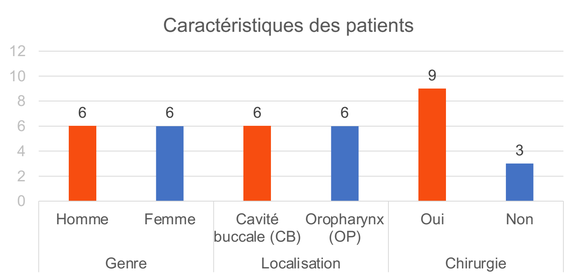
Système	Architecture	Corpus d'entraînement (heures)
CRDNN	End-to-end Convolutional Recurrent Deep Neural Network [6]	LibriSpeech (960 h)
TDNNf	Time-Delay Neural Network factorisé – Hidden Markov Model [7]	CommonVoice (147 h)
Whisper (Open AI)	Transformer Seq2seq [8]	Multilingue (> 117 kh)

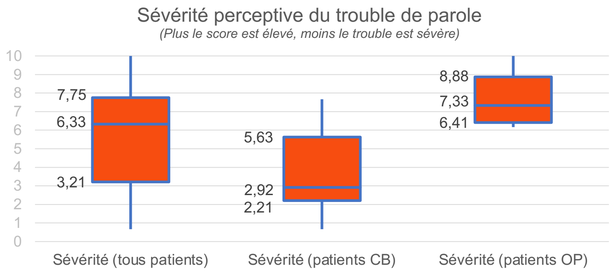
Comparaisons statistiques (tests de Wilcoxon)

- **Performances des systèmes de RAP** entre eux (**1a** – *chez un expert* – et **1b** – *chez les naïfs*)
- **Comparaison des résultats automatiques à la référence perceptive** (**2a** – *CRDNN*, **2b** – *TDNNf* et **2c** – *Whisper*) pour savoir si le système (entraîné sur de la parole typique) se rapproche davantage d'une reconnaissance « experte » ou « naïve » ? [7]
- **Critères de jugement** (HTK [8]) : nombre moyen de mots correctement reconnus par le système (**H**), nombre moyen de délétions (**D**), de substitutions (**S**), d'insertions (**I**), **exactitude** $= \frac{100 \times (H-I)}{H+S+D}$ et **pourcentage de mots corrects** $= \frac{100 \times H}{H+S+D}$

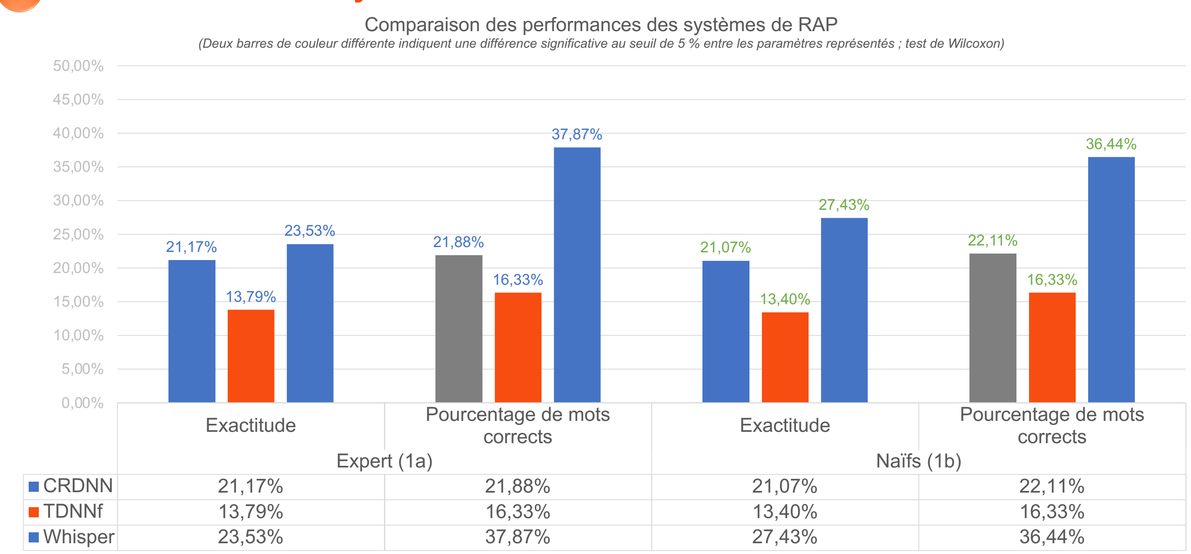
Résultats

12 patients (âge médian : 65 ans, EIQ 12,75)





1 Performances des systèmes de RAP



	Expert (1a)				Naïfs (1b)			
	H	**D**	**S**	**I**	**H**	**D**	**S**	**I**
CRDNN	20	39,83	18,33	0,5	18,83	32,58	18,42	0,75
TDNNf	14,33	26,67	37,17	1,75	13,54	22,04	37,5	2,21
Whisper	32,83	15,92	29,42	8	29,42	10,25	31	9,83

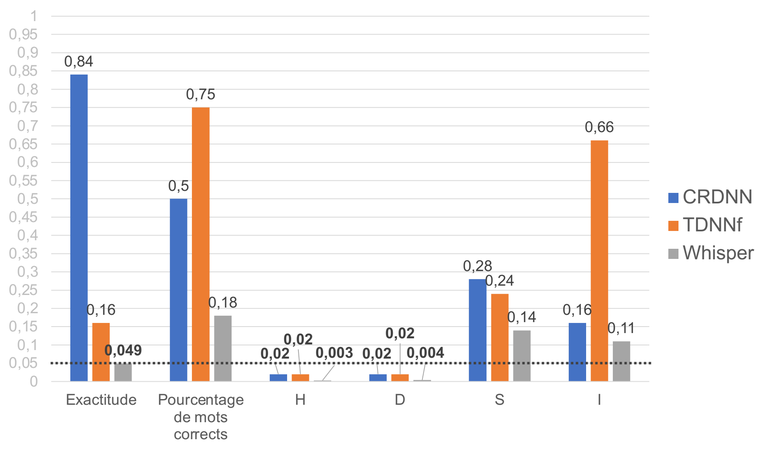
Meilleures performances de reconnaissance de parole par **Whisper**

- **Exactitude significativement meilleure** que le **TDDNf** (p=0,004 chez l'expert et p=0,007 chez les naïfs) mais différence **non significative** avec le **CRDNN** (p>0,08)
- **Pourcentage de mots corrects reconnus significativement meilleur** que le TDNNf (p<0,01) et que le CRDNN (p<0,003)
- Nombre de mots correctement reconnus significativement plus élevé (p<0,02), nombre de délétions significativement plus faible qu'avec les deux autres systèmes (p<0,003)
- Nombre d'insertions significativement plus élevé qu'avec les autres systèmes (p≤0,05)

Malgré tout, dans le meilleur système (Whisper) : **taux d'erreurs mots très élevé** = 76,47 %

2 Comparaison des performances selon le profil expert/naïfs de l'examinateur

Les valeurs données correspondent aux valeurs *p* du test de Wilcoxon (en gras, les valeurs *p* significatives au seuil de 5 %).



Discussion

Étude exploratoire

- **Architecture** des systèmes de RAP et **taille de leur corpus d'entraînement** (sur **parole typique**) **influencent** les performances de RAP après cancer
- **Exactitude faible** dans tous les cas

Perspectives

- Valider les résultats sur un **échantillon plus large**
- Étudier les **performances** des systèmes en **sous-groupes par sévérité** du trouble de parole
- **Adapter les modèles acoustiques** de parole typique à la parole pathologique pour améliorer l'exactitude
- À terme : réaliser des **analyses de contenus du discours** informant sur la **dynamique psychosociale** des patients

Références

[1] Knuijt, S., Kalf, J. G., van Engelen, B. G. M., de Swart, B. J. M., & Geurts, A. C. H. (2017). The Radboud Dysarthria Assessment: Development and Clinimetric Evaluation. *Folia Phoniatrica et Logopaedica*, 69(4), 143–153. https://doi.org/10.1159/000484556

[2] Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091. https://doi.org/10.1080/02687030444000534

[3] Balaguer, M. (2021). *Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé*. Thèse de doctorat, Université Paul Sabatier Toulouse III. https://theses.hal.science/tel-03557511

[4] Heba, A. (2021). *Reconnaissance automatique de la parole à large vocabulaire : des approches hybrides aux approches End-to-End*. Thèse de doctorat, Université Paul Sabatier Toulouse III. https://tel.archives-ouvertes.fr/tel-03616588

[5] Gelin, L., Daniel, M., Pinquier, J., & Pellegrini, T. (2021). End-to-end acoustic modelling for phone recognition of young readers. *Speech Communication*, 134, 71–84. https://doi.org/10.1016/j.specom.2021.08.003

[6] Radford, A., Wook, J., Tao, K., Greg, X., Christine, B., & Ilya, M. (2021). Robust Speech Recognition via Large-Scale Weak Supervision. Openai.Com. https://doi.org/10.48550/arXiv.2212.04356

[7] Fex, S. (1992). Perceptual evaluation. *Journal of Voice*, 6(2), 155-158

[8] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X. (Andrew), Moore, G., Odell, J., Ollason, D., Pover, D., Ragni, A., Valtchev, V., Woodland, P., & Zhang, C. (2015). The HTK Book (for HTK Version 3.5, documentation alpha version) (Issue December). Cambridge University Engineering Department.