

Une méthode automatique non supervisée pour évaluer le score de sévérité de la parole chez les patients traités pour un cancer ORL

Vincent ROGER¹
Jérôme FARINAS¹
Virginie WOISARD²
Julien PINQUIER¹

¹*Institut de Recherche en Informatique de Toulouse, CNRS,
Université Toulouse III, France*

²*Service ORL Hôpital Larrey, Hôpitaux de Toulouse, France*

Entre 2007 et 2016, le nombre de cancers de la cavité buccale et du pharynx combinés a augmenté, avec une incidence de 3 % aux États-Unis chaque année [1]. De nombreuses autres pathologies peuvent entraîner des troubles de la parole en influant sur les mécanismes de production de la parole [2]. Cela peut réduire la qualité de vie des patients affectés [3]. Pour quantifier l'impact de ces pathologies sur la production de la parole, il est courant d'utiliser des scores qui reflètent l'intelligibilité de la production et l'indice de sévérité de la parole. Ils mesurent l'impact des pathologies. Ainsi, ces scores ne mesurent pas la qualité du signal audio, mais la capacité des patients à produire des sons intelligibles. Les experts obtiennent ces scores en écoutant leurs patients. La familiarité avec les patients et l'habitude d'entendre la voix pathologique peuvent influencer ce score [4]. Afin d'augmenter la stabilité de la mesure, celle-ci peut être traitée par un groupe d'experts [5]. Néanmoins, le fait de demander à un groupe d'obtenir un score objectif complique le suivi des patients. L'utilisation d'un outil automatique, calculé par un ordinateur sur un échantillon de discours enregistré, pourrait fournir un moyen objectif et libérer du temps pour les experts. Dans notre étude, nous proposons une nouvelle méthode non supervisée utilisant un modèle appris sur la parole standard en Anglais. Nous proposons un formalisme basé sur des caractéristiques de la parole qui produisent un score de sévérité. Cette méthode, utilisant l'encodeur PASE+[6], s'inspire du modèle *Inception* [7] (modèle de réseau de neurones

profonds, généralement utilisé en vision par ordinateur pour évaluer la qualité des images générées où les locuteurs sont ici considérés comme des générateurs). Cette approche, bien que fondée sur un modèle en Anglais, nous donne d'excellents résultats sur les données en Français de lecture du corpus de voix pathologiques C2SI [8]. En effet, une corrélation de Spearman de 0,87 est obtenue par rapport à l'indice de sévérité, réalisé manuellement par un groupe d'experts cliniques (voir figure 1). Notons que la mesure de sévérité est très corrélée ($> 0,91$) avec la mesure d'intelligibilité [9]. Nous arrivons à des résultats comparables avec ceux obtenus par des méthodes qui ont réalisé un apprentissage sur les données du corpus [10,11]. La corrélation élevée et le fait que cette méthode ne nécessite aucun apprentissage sur des enregistrements de voix pathologiques, nous permettent d'envisager des applications cliniques intéressantes.

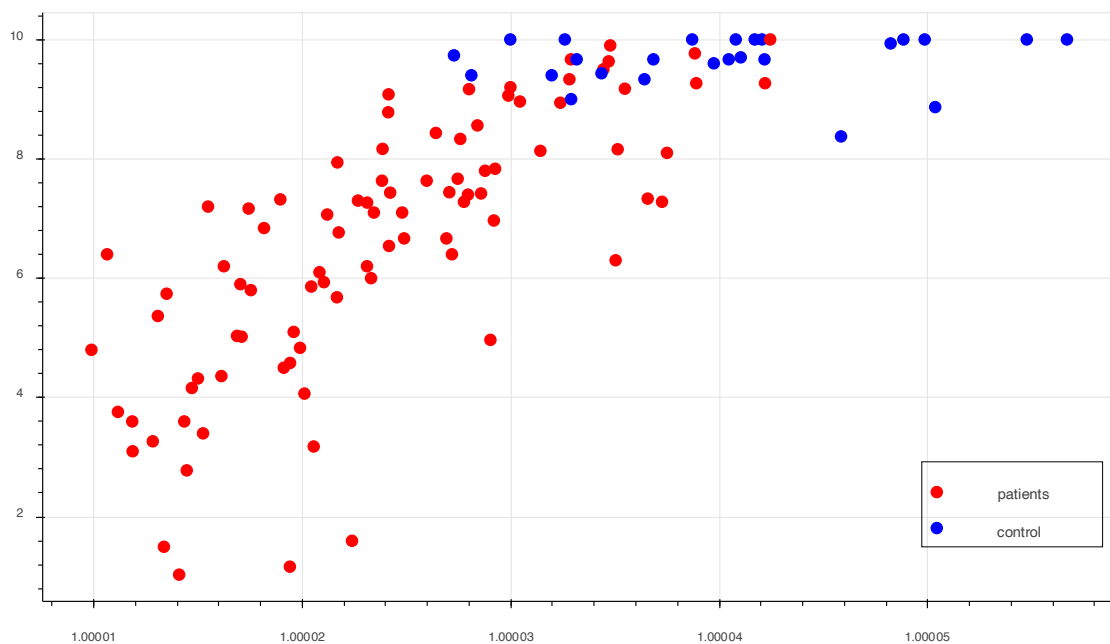


Figure 1. Nuage de points mettant en lien notre mesure (en abscisse) et l'indice de sévérité des experts (en ordonnée).

Références bibliographiques

- [1] T. D. Ellington, "Trends in incidence of cancers of the oral cavity and pharynx in United States 2007–2016," *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, 2020.

- [2] P. Enderby, “Disorders of communication: Dysarthria,” In: *Handbook of Clinical Neurology*, vol. 110, pp. 273–281. Elsevier, 2013.
- [3] M. Walshe and N. Miller, “Living with acquired dysarthria: The speaker’s perspective,” In: *Disability and Rehabilitation*, vol. 33, no. 3, pp. 195–203, 2011.
- [4] S. Landa, L. Pennington, N. Miller, S. Robson, V. Thompson, and N. Steen, “Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding,” *International journal of speech-language pathology*, vol. 16, no. 4, pp. 408–416, 2014.
- [5] V. Woisard and B. Lepage, “Perception of speech disorders: Difference between the degree of intelligibility and the degree of severity,” *Audiological Medicine*, vol. 8, pp. 171–178, Nov. 2010.
- [6] M. Ravanelli, J. Zhong, S. Pascual, P. Swietojanski, J. Monteiro, J. Trmal, and Y. Bengio, “Multi-task self-supervised learning for Robust Speech Recognition,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 6989–6993.
- [7] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved Techniques for Training GANs,” in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., pp. 2234–2242. Curran Associates, Inc., 2016.
- [8] V. Woisard, C. Astesano, M. Balaguer, J. Farinas, C. Fredouille, P. Gaillard, A. Ghio, L. Giusti, I. Laaridh, M. Lalain, B. Lepage, J. Mauclair, O. Nocaudie, J. Piquier, G. Pouchoulin, M. Puech, D. Robert, and V. Roger, “C2SI corpus: A Database of Speech Disorder Productions to Assess Intelligibility and Quality of Life in Head and Neck Cancers,” *Language Resources and Evaluation*, June 2020.
- [9] M. Balaguer, “Construction d’un score Carcinologic Speech Severity Index (C2SI) automatique”, Rapport de master 2 recherche mention santé publique, spécialité épidémiologique clinique, Université Paul Sabatier Toulouse 3 & Faculté de Médecine de Purpan, juin 2018
- [10] I. Laaridh, C. Fredouille, A. Ghio, M. Lalain, V. Woisard. Automatic Evaluation of Speech Intelligibility Based on I-vectors in the Context of Head and Neck Cancers. *ISCA Interspeech 2018*, pp. 2943-2947, Hyderabad, India
- [11] Sebastião Quintas, Julie Mauclair, Virginie Woisard, Julien Piquier. Automatic Prediction of Speech Intelligibility Based on X-Vectors in the

Context of Head and Neck Cancer. *ISCA Interspeech 2020*, pp. 4976-4980, Oct 2020, Shanghai, China