



Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Mathieu Balaguer

► To cite this version:

Mathieu Balaguer. Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé. Les Cahiers de l'ASELF, 2022, 19 (2), pp.19. hal-03884599

HAL Id: hal-03884599

<https://hal.science/hal-03884599>

Submitted on 5 Dec 2022

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.



Distributed under a Creative Commons Attribution 4.0 International License

Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Mathieu Balaguer

Orthophoniste, Docteur en Informatique et Télécommunications
IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
Hôpital Larrey, Hôpitaux de Toulouse, France

Résumé

Contexte : Les méthodes actuelles d'évaluation de la parole après traitement d'un cancer oral ou oropharyngé prédisent mal les conséquences sur les capacités de communication. L'utilisation d'outils d'analyse automatique de la parole peut permettre de donner des scores fiables prédisant mieux les capacités de communication.

Objectif : Prédire l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé.

Méthodes : Un état des lieux des outils acoustiques / automatiques utilisables en cancérologie ORL a été réalisé, par une revue systématique de littérature. Un corpus a été constitué. Il incluait onze autoquestionnaires (ECVB, DIP, CHI, PHI, QFS, ECCS, HAD, EDP, MoCA, EORTC QLQ-C30, EORTC QLQ-H&N35), permettant notamment la construction d'un score holistique de communication (SHC) par réduction de dimensionnalité. Il comprenait également des enregistrements de parole spontanée, dont différents paramètres automatiques ont été extraits (par analyse du signal et reconnaissance automatique de la parole). Le score de communication a été prédit (régression LASSO) par (1) les paramètres extraits de la parole, et (2) en y associant les données fonctionnelles des questionnaires.

Résultats : En cancérologie ORL, aucune étude ne concerne l'analyse acoustique / automatique de la parole spontanée. Ainsi, pour 25 sujets, 75 paramètres automatiques ont été extraits des enregistrements. Le SHC a été construit avec 24 items des autoquestionnaires. Une prédiction de ce score a été réalisée, par les paramètres automatiques ($rs=0,83$ entre scores prédit et réel) puis avec les facteurs associés (cercles sociaux, état anxio-dépressif, déficits associés, auto-perception du handicap : $rs=0,89$).

Conclusion : Les analyses automatiques de la parole spontanée permettent une prédiction de l'altération de communication ressentie par les patients. Cette étude ouvre de nouvelles perspectives quant à l'utilisation de la reconnaissance automatique de parole dans l'évaluation clinique, et la prise en compte des besoins exprimés par les patients.

1. Introduction

Les troubles de parole sont une problématique fréquemment rencontrée après traitement d'un cancer de la cavité buccale ou de l'oropharynx¹ (Balaguer et al., 2019; Borggreven et al., 2007). Différents modèles conceptuels permettent d'analyser les conséquences fonctionnelles du trouble de parole sur les capacités de communication ou de l'altération de la qualité de vie qui en découle (Murphy et al., 2007; Wilson & Cleary, 1995).

En clinique, l'optimisation des capacités de communication est un objectif thérapeutique majeur dans le suivi de ces patients. Or, l'évaluation habituellement réalisée présente deux limites. D'une part, l'impact fonctionnel est peu évalué par les cliniciens (Pommée et al., 2021) : un quart n'évalue pas les répercussions psychosociales du trouble de parole, ou alors avec des outils inadaptés. D'autre part, peu

¹ Rapport Santé publique France, juillet 2019 :

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-dusein/documents/rapport-synthese/estimations-nationales-de-l-incidence-et-de-lamortalite-par-cancer-en-france-metropolitaine-entre-1990-et-2018-volume-1-tumeurssolides-etud>, consulté le 07/05/2021

d'outils sont disponibles en cancérologie de la cavité buccale ou de l'oropharynx (Plisson et al., 2017). Ces outils sont majoritaires perceptifs, réalisés « à l'oreille » du clinicien, mais aboutissent à des scores prédisant mal l'impact de ces troubles sur la communication (Gray et al., 2012). De plus, l'évaluation perceptive présente une variabilité inter-juges et intra-juge importante (Kuo & Tjaden, 2016), limitant la fiabilité et la reproductibilité de la mesure. Ainsi, le développement de nouveaux outils d'analyse automatique de la parole permettrait de fournir aux cliniciens de nouveaux outils fiables et reproductibles, et d'aboutir à une meilleure prédiction des capacités de communication.

L'objectif est de prédire l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé.

2. Méthodes

2.1. Revue systématique de la littérature (RS)

Une RS a été conduite pour obtenir un état des lieux actuel et exhaustif des outils acoustiques et automatiques utilisés sur le signal de parole pour mesurer les effets des cancers oraux et oropharyngés (Balaguer et al., 2020), selon les recommandations PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2009).

Les critères d'inclusion des articles étaient : évaluer la parole, inclure des termes relatifs à l'acoustique ou à l'analyse automatique de la parole, porter sur des patients traités d'un cancer oral ou oropharyngé. Ont notamment été exclus les articles non originaux et n'étant pas en anglais.

Deux bases de données, PubMed et Embase, ont été sondées le 4 décembre 2018. Puis une démarche de screening, d'analyse de l'éligibilité des résumés et du contenu des articles finalement retenus a été menée.

2.2. Constitution du corpus

Ce travail a été réalisé dans le cadre du projet ANR RUGBI² et a été approuvé par le Comité de Protection des Personnes (CPP : Ouest IV, 19/02/2020, référence 11/20_3).

Un nouveau corpus a été constitué pour (1) obtenir des échantillons de parole spontanée en situation conversationnelle et (2) prendre en compte les dimensions fonctionnelles (niveau des activités) et psychosociales (niveau des participations), notamment les aspects relatifs à la communication.

Ont été inclus des patients majeurs, francophones natifs, traités pour un cancer oral ou oropharyngé depuis au moins six mois (stabilité des troubles).

Un premier volet de ce corpus est constitué d'autoquestionnaires, questionnant les données individuelles, cliniques et de traitement, les déficits consécutifs au traitement de la pathologie (Carcinologic Handicap Index), le statut anxiodépressif (Hospital Anxiety and Depression scale et Échelle de Détresse Psychologique) et cognitif (Montreal Cognitive Assessment), les besoins de communication liés aux cercles sociaux (Évaluation de la Constitution des Cercles Sociaux et Questionnaire de Fonctionnement Social), l'auto-perception du trouble de parole (Phonation Handicap Index), la communication (Échelle de Communication Verbale de Bordeaux, Dysarthria Impact Profile) et la qualité de vie globale et relative à la santé (EORTC QLQ-C30, H&N35).

Le second volet comprend des enregistrements de parole spontanée au cours d'un entretien semi-dirigé. La sévérité du trouble de parole et l'intelligibilité ont été évaluées perceptivement par un jury de six experts pour avoir une mesure de référence.

2.3. Analyses automatiques de la parole spontanée

Dans un premier temps, différents paramètres issus d'analyses automatiques du signal de parole spontanée ont été recherchés :

- En fonction de leur disponibilité et de leur applicabilité ;
- Sur l'ensemble des niveaux linguistiques définis par le modèle psycholinguistique de Caron (François & Nespoulous, 2011) (voir Figure 1).

² « Recherche d'unités linguistiques pertinentes pour améliorer la mesure de l'intelligibilité de la parole altérée par des troubles de production pathologique », <https://www.irit.fr/rugbi/>

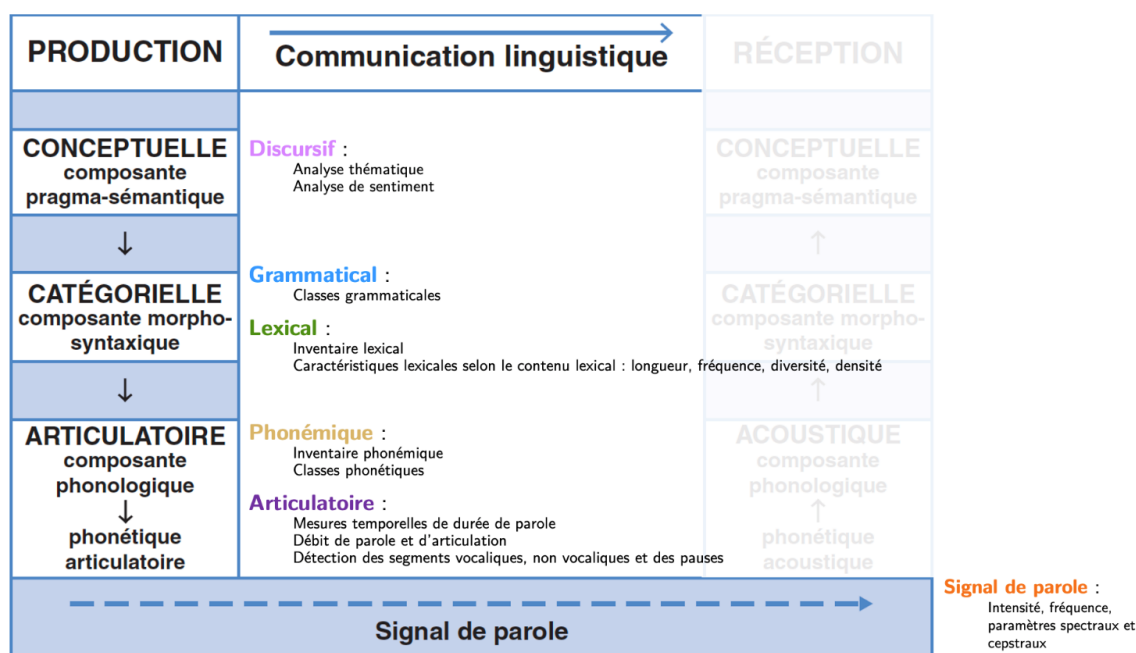


Figure 1 Schéma général de l'analyse des paramètres automatiques extraits de la parole spontanée selon le modèle psycholinguistique de Caron

Les paramètres généraux, spectraux et cepstraux du signal de parole ont été extraits au moyen du script AVQI (Pommée et al., 2020). Des mesures temporelles ont été collectées concernant la durée du signal de parole, le rythme (outil SpeechTools) (Farinas, 2002) ou le débit de parole (de Jong & Wempe, 2009).

Un système de reconnaissance automatique de la parole (RAP) de type Time-Delay Neural Network factorisé (Povey et al., 2018) a été employé pour la reconnaissance de phonèmes (Gelin et al., 2021) et de mots. Il a permis de calculer également des scores de confiances associés à l'item reconnu, ou issus de statistiques phonétiques, lexicales ou grammaticales.

Enfin, des analyses thématiques ont été conduites au moyen d'une classification hiérarchique descendante (Reinert, 1990). Le sentiment global, positif ou négatif, issu de la reconnaissance lexicale a été étudié (Microsoft Azure).

Puis, un processus de sélection a été conduit pour ne retenir que les items pertinents (mesure d'un effet présent dans l'échantillon, risque de biais limité et étant accessibles d'interprétation selon un comité d'experts) et non-redondants (matrice de corrélation).

2.4. Prédiction de l'altération de la communication

Une mesure unique d'impact fonctionnel sur la communication a été construite pour pallier son absence en cancérologie ORL. Par une méthodologie de réduction de dimensionnalité, un score holistique de communication (SHC) a été construit à partir d'items issus des autoquestionnaires. Pour cela, un comité d'experts de différentes disciplines concernées par l'étude de la communication (sciences du langage, médecine, orthophonie, informatique et traitement du signal) a sélectionné les items spécifiquement relatifs à la communication en contexte cancérologique. Puis les items redondants ont été éliminés (analyse de corrélation inter-items), et le SHC a ensuite été construit par analyse factorielle en principal-component factor (PCF) (Acock, 2018). Enfin, la fiabilité de ce score a été étudiée par validation croisée.

Ensuite, l'altération de la communication mesurée par le SHC a été prédite au moyen d'une régression LASSO (Tibshirani, 1996), d'abord avec les paramètres issus de l'analyse automatique de la parole spontanée, puis en y associant les données issues des autoquestionnaires (cercles sociaux, déficits associés, statuts cognitif et anxio-dépressif). La fiabilité de cette prédiction a été étudiée en validation croisée.

3. Résultats

3.1. Revue systématique de la littérature

488 références relatives aux effets mesurés acoustiquement ou automatiquement des cancers oraux ou oropharyngés sur la parole ont été récupérées. Après vérification de la conformité des résumés aux critères d'inclusion, un total de 196 articles complets a été évalué. 173 ont été exclus car ils ne respectaient pas les critères d'inclusion, n'étaient pas récupérables, étaient non originaux ou des duplicatas d'articles. 22 articles ont ainsi été retenus pour la suite des analyses (100 % : QualSyst $\geq$ bonne qualité méthodologique) (voir Figure 2).

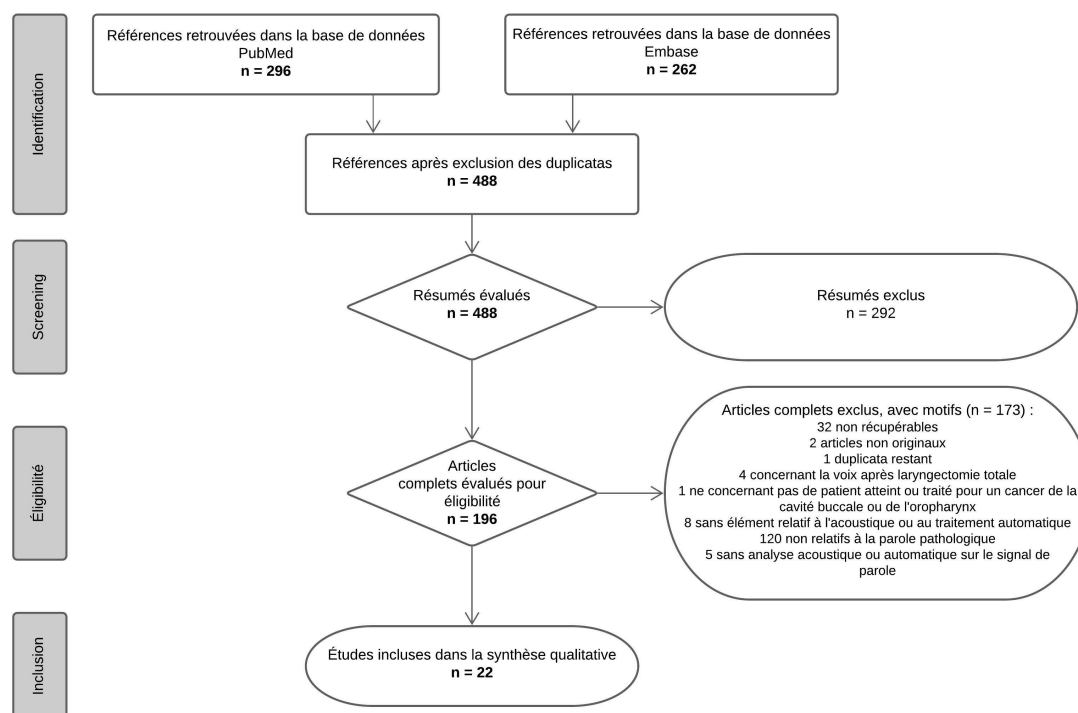


Figure 2 Diagramme du processus de revue, selon la méthodologie PRISMA

La plupart des études se sont concentrées sur la qualité du décodage acoustico-phonétique pour mesurer la parole : neuf articles prenant les mesures sur des voyelles et cinq sur les consonnes. L'analyse acoustique et automatique de la parole est ainsi principalement réalisée au niveau segmental, dans un contexte contrôlé de parole. Aucune étude n'a étudié la parole spontanée, basée sur des productions non contraintes. Pourtant, ce contexte de production est le plus proche des situations de communication vécues par les patients au quotidien.

3.2. Constitution du corpus

25 sujets ont été inclus dans cette étude : majorité d'hommes (60 %), âge médian 67 ans (écart interquartile = 12). 56 % des sujets ont été traités pour une tumeur de la cavité buccale (oropharynx : 40 %, cavité buccale + oropharynx : 4 %). Une majorité (57,9 %) de sujets a été traitée d'une tumeur de gros volume (T3 ou T4). Le traitement est majoritairement chirurgical (88 % ; radiothérapie : 96 % ; chimiothérapie : 60 % ; association chirurgie-radiothérapie : 84 %). Le délai médian depuis la fin du traitement est de 3 ans et 4 mois.

Les sujets inclus présentent des troubles de parole de tous niveaux de sévérité perceptive (voir Figure 3).

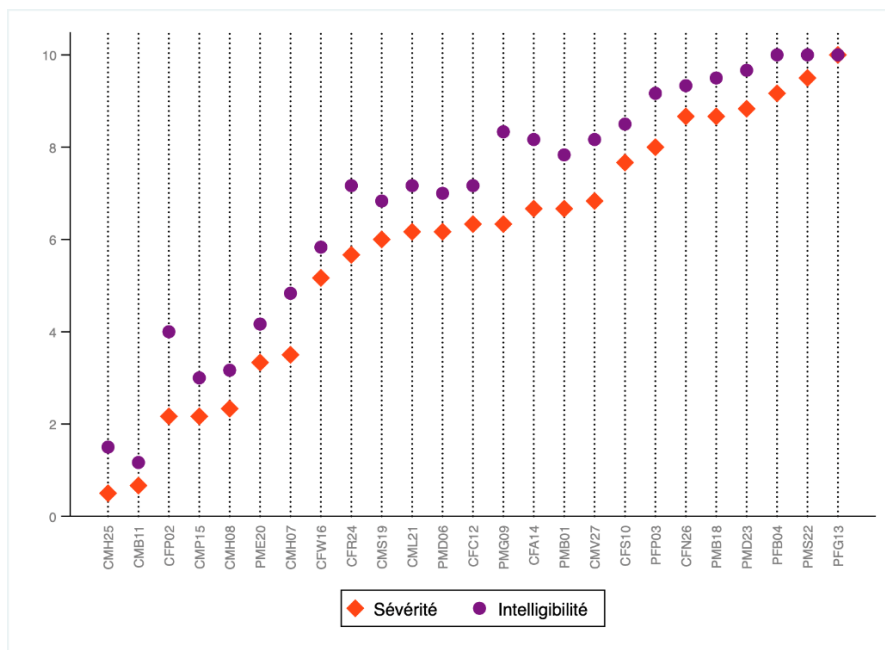


Figure 3 Scores perceptifs de sévérité et d'intelligibilité des sujets inclus

3.3. Analyses automatiques de la parole spontanée

149 paramètres automatiques ont pu être extraits des enregistrements de parole spontanée (voir Figure 4).

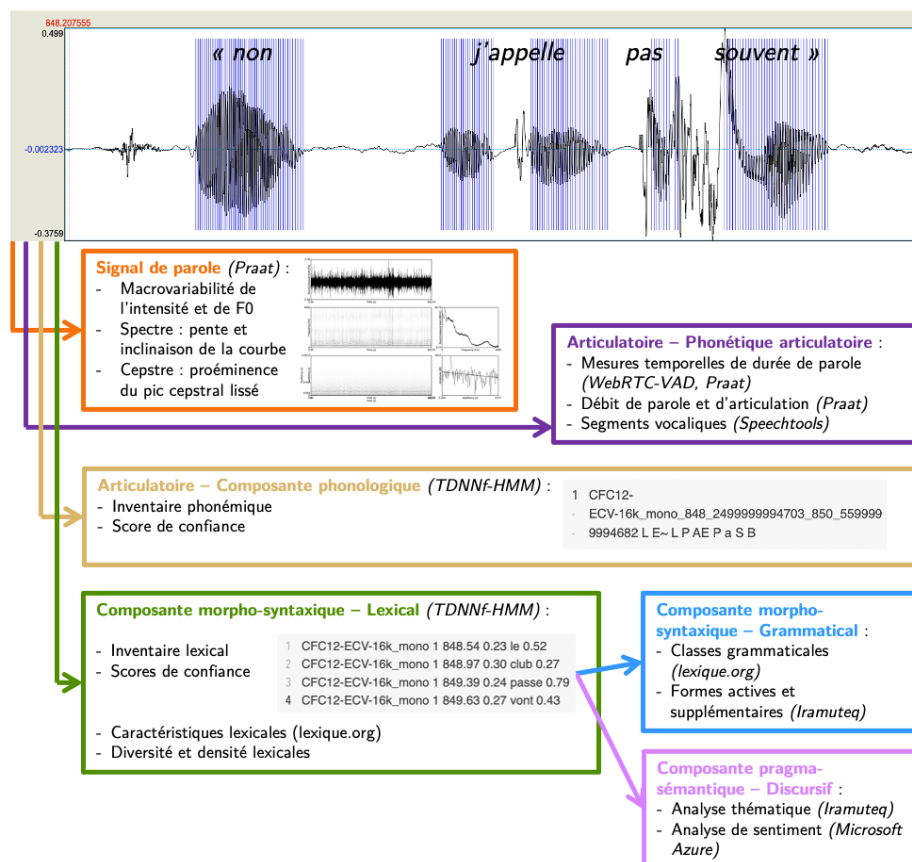


Figure 4 Schéma synthétique des paramètres extraits

31 paramètres ont été exclus en raison de leur non-pertinence :

- Trois liés au genre du locuteur donc non généralisables ;
- Deux mesurant des effets absents du corpus ;
- Quinze d'interprétation complexe, car issus d'une construction statistique par les outils d'extraction eux-mêmes (consensus d'experts).

23 paramètres supplémentaires ont été exclus pour redondance, car trop fortement corrélés ($r \geq 0,90$) à d'autres paramètres. 75 paramètres finaux sont alors conservés pour la suite des analyses.

3.4. Prédiction de l'altération de la communication

Parmi les 174 items issus des autoquestionnaires complétés par les sujets :

- 130 ont été exclus par le comité d'experts, car non pertinents : items non relatifs à l'expression orale, et/ou sans référence à une interaction à un tiers ;
- 4 items ont ensuite été exclus pour redondance ($r \geq 0,90$ avec un autre item).

40 items ont été ainsi utilisés pour la construction du SHC par analyse factorielle en PCF. Seuls 26 présentent un coefficient de régression nul et permettent de calculer le SHC par moyenne pondérée (ECVB : 13, DIP : 6, PHI : 4, EORTC QLQ-H&N35 : 1) (voir Figure 5).

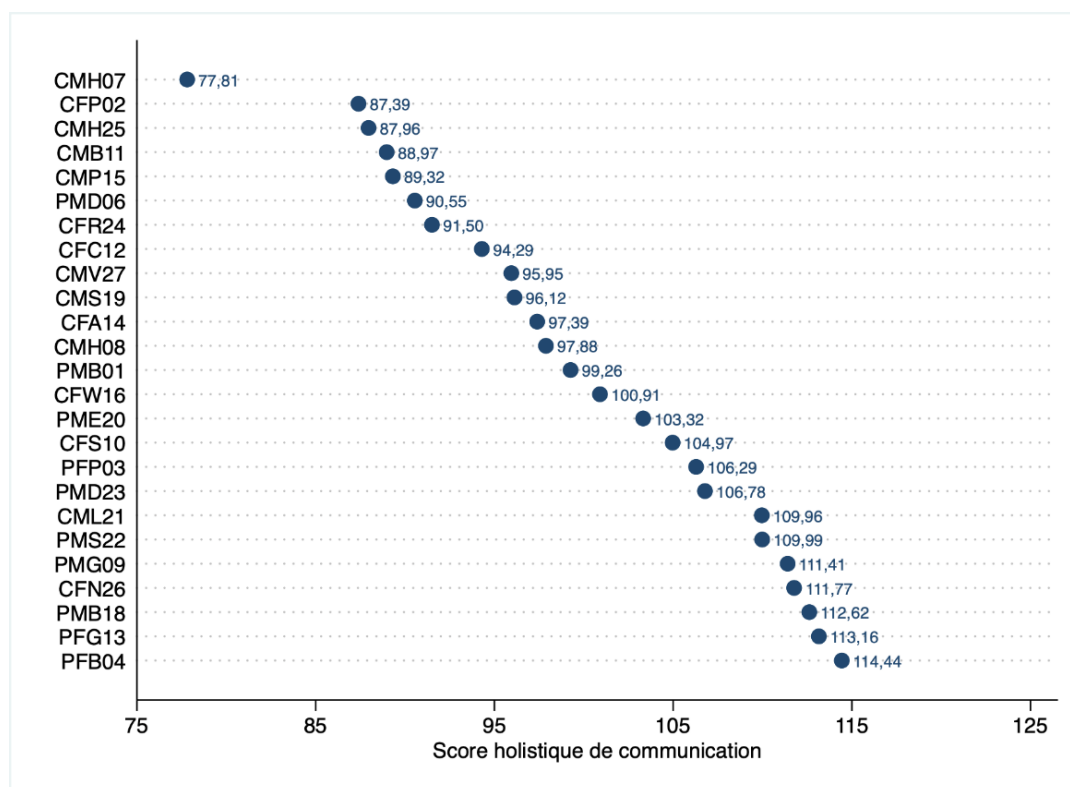


Figure 5 Scores holistiques de communication obtenus par sujet

Le SHC présente de bonnes performances en validation croisée (corrélation $r_s = 0,91$ entre SHC et score prédit par validation croisée à cinq blocs).

En prédisant le SHC au moyen des seuls paramètres automatiques, la régression LASSO aboutit à la sélection d'un modèle à deux variables explicatives : le nombre d'occlusives par seconde (*occs*) et le nombre de sonantes par seconde (*sonants*).

$$\text{SHC} = 83,672 + (10,139 \times \text{occs}) + (10,891 \times \text{sonants})$$

La corrélation entre SHC et valeurs prédites par ce modèle est de $r_s = 0,83$, et la validation croisée à cinq blocs montre une bonne fiabilité de cette prédiction ($r_s = 0,82$).

La prédiction est améliorée ($r_s = 0,89$) en incluant les facteurs associés (constitution des cercles sociaux, état anxiodépressif, déficits associés, auto-perception du handicap lié au trouble de parole) (Tableau 1).

Code	Paramètre	Coefficient
<i>Constante</i>		73,74
<i>occs</i>	Nombre d'occlusives reconnues par seconde	9,10
<i>sonants</i>	Nombre de sonantes reconnues par seconde	4,07
<i>eccs1c</i>	Satisfaction moyenne de la fréquence de contact avec les membres du cycle privé	1,32
<i>eccs1d</i>	Satisfaction moyenne de la qualité des échanges avec les membres du cercle privé	0,05
<i>qfs52</i>	Satisfaction relative aux relations avec les membres de l'entourage extrafamilial	3,58
<i>qfs62</i>	Satisfaction relative à la manière de gérer les aspects financiers et administratifs	1,56
<i>had2</i>	« Je prends plaisir aux mêmes choses qu'autrefois »	-5,22
<i>hada</i>	Score d'anxiété global	3,39
<i>chiog2</i>	« Avez-vous peur d'un accident à cause de vos problèmes d'odorat ? »	-3,99
<i>phihcp</i>	Handicap quotidien ressenti quant au problème de parole	-0,12

Tableau 1 : Résultats de la modélisation du SHC par les paramètres automatiques et les facteurs associés

Cette prédiction montre une corrélation à $r_s = 0,70$ en validation croisée (voir Figure 6).

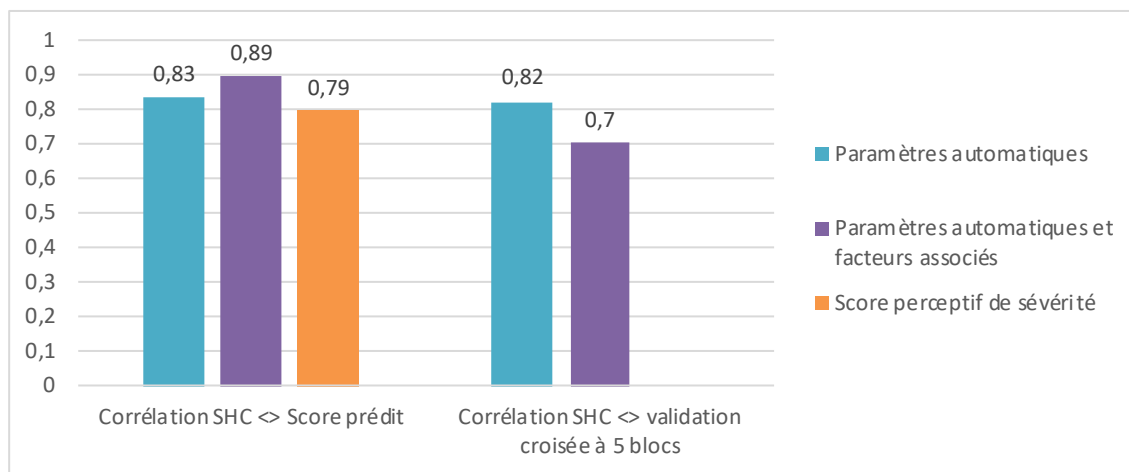


Figure 6 Comparaison des performances entre modélisations du SHC

Dans tous les cas, l'utilisation des paramètres issus des analyses automatiques de la parole spontanée présente des résultats légèrement meilleurs que le score perceptif de sévérité du trouble de parole.

4. Discussion

L'utilisation de l'analyse automatique de la parole permet donc une prédiction fiable de l'altération de communication ressentie par les patients. Cette étude ouvre de nouvelles perspectives quant à l'utilisation et l'optimisation des systèmes de reconnaissance automatique de parole dans l'évaluation

clinique d'une part, et la prise en compte des besoins fonctionnels et psychosociaux exprimés par les patients d'autre part (Balaguer, 2021).

L'élargissement de la taille de l'échantillon d'étude permettrait à la fois une meilleure généralisabilité des résultats, mais aussi de réaliser des analyses en sous-groupes afin d'étudier les paramètres prédisant le mieux l'altération de la communication selon différents facteurs comme le traitement subi. L'intégration d'indices non-verbaux à la prédiction de la communication autoriserait de s'intéresser aux moyens de compensation mis en œuvre par les sujets pour pallier les déficits fonctionnels. Du côté de l'analyse automatique, l'entraînement des systèmes de reconnaissance de la parole avec de la parole pathologique permettrait peut-être d'accéder à des informations plus précises et spécifiques de haut niveau linguistique. Enfin, l'applicabilité clinique passera par le développement d'un dispositif mobile, peu onéreux et facilement utilisable, permettant une mesure en temps réel de l'altération de communication, auprès des patients en cancérologie ORL. Des dispositifs de ce type sont en cours d'expérimentation dans des services cliniques pour offrir aux cliniciens de nouveaux outils d'évaluation fiables, rapides (par le caractère automatique) et écologiques (par l'enregistrement non contraint d'un entretien spontané).

Bibliographie

- Acock, A. C. (2018). *A Gentle Introduction to Stata- 6th Edition*. Stata Press.
- 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* [Université Toulouse III Paul Sabatier]. <http://www.theses.fr/2021TOU30109>
- Balaguer, M., Boisguerin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019). Factors influencing intelligibility and severity of chronic speech disorders of patients treated for oral or oropharyngeal cancer. *European Archives of Oto-Rhino-Laryngology*, 276(6), 1767–1774. <https://doi.org/10.1007/s00405-019-05397-6>
- Balaguer, M., Pommée, T., Farinas, J., Pinquier, J., Woisard, V., & Speyer, R. (2020). 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>
- Borggrevén, P. A., Verdonck-De Leeuw, I. M., Muller, M. J., Heiligers, M. L. C. H., De Bree, R., Aaronson, N. K., & Leemans, C. R. (2007). Quality of life and functional status in patients with cancer of the oral cavity and oropharynx: Pretreatment values of a prospective study. *European Archives of Oto-Rhino-Laryngology*, 264(6), 651–657. <https://doi.org/10.1007/s00405-007-0249-5>
- de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2), 385–390. <https://doi.org/10.3758/BRM.41.2.385>
- Farinas, J. (2002). *Une modélisation automatique du rythme pour l'identification des langues*. Université Paul Sabatier.
- François, J., & Nespoulous, J. L. (2011). *L'architecture des processus de production et de réception : aspects (neuro)psycholinguistiques* (Peeters).
- Gelin, L., Daniel, M., Pinquier, J., & Pellegrini, T. (2021). *End-to-end acoustic modelling for phone recognition of young readers*. <https://www.lalilo.com/>
- Gray, C., Baylor, C., Eadie, T., Kendall, D., & Yorkston, K. (2012). The Levels of Speech Usage rating scale: comparison of client self-ratings with speech pathologist ratings. *International Journal of Language & Communication Disorders*, 47(3), 333–344. <https://doi.org/10.1111/j.1460-6984.2011.00112.x>
- Kuo, C., & Tjaden, K. (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *Journal of Communication Disorders*, 62, 30–44. <https://doi.org/10.1016/j.jcomdis.2016.05.003>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339(jul21 1), b2535–b2535. <https://doi.org/10.1136/bmj.b2535>
- Murphy, B. A., Ridner, S., Wells, N., & Dietrich, M. (2007). Quality of life research in head and neck cancer: A review of the current state of the science. *Critical Reviews in Oncology/Hematology*, 62(3), 251–267. <https://doi.org/10.1016/j.critrevonc.2006.07.005>
- Plisson, L., Pillot-Loiseau, C., & Crevier-Buchman, L. (2017). Intelligibilité de la parole après le traitement d'un cancer de l'oropharynx : étude descriptive chez sept patients en pré-traitement et en post-traitement précoce. *7èmes Journées de Phonétique Clinique (JPC7)*.

- Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021). Assessment of adult speech disorders: current situation and needs in French-speaking clinical practice. *Logopedics Phoniatrics Vocology*, 0(0), 1–15. <https://doi.org/10.1080/14015439.2020.1870245>
- Pommée, T., Maryn, Y., Finck, C., & Morsomme, D. (2020). Validation of the Acoustic Voice Quality Index, Version 03.01, in French. *Journal of Voice*, 34(4), 646.e11-646.e26. <https://doi.org/10.1016/j.jvoice.2018.12.008>
- Povey, D., Cheng, G., Wang, Y., Li, K., Xu, H., Yarmohammadi, M., & Khudanpur, S. (2018). Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks. *Interspeech 2018, 2018-Sept*(2), 3743–3747. <https://doi.org/10.21437/Interspeech.2018-1417>
- Reinert, M. (1990). Une méthode de classification des énoncés d'un corpus présentée à l'aide d'une application. *Les Cahiers de l'analyse Des Données*, 15(1), 21–36.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA: The Journal of the American Medical Association*, 273(1), 59–65. <https://doi.org/10.1001/jama.273.1.59>