



HAL
open science

La modélisation automatique de l'intelligibilité pour améliorer la qualité de vie des patients atteints de troubles de la parole : un exemple d'IA dans le domaine médical

Jérôme Farinas

► To cite this version:

Jérôme Farinas. La modélisation automatique de l'intelligibilité pour améliorer la qualité de vie des patients atteints de troubles de la parole : un exemple d'IA dans le domaine médical. A.N.A.E. Approche neuropsychologique des apprentissages chez l'enfant, 2024, 36 (189), pp.137-141. hal-04602548v1

HAL Id: hal-04602548

<https://hal.science/hal-04602548v1>

Submitted on 5 Jun 2024 (v1), last revised 5 Jun 2024 (v2)

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

Titre : La modélisation automatique de l'intelligibilité pour améliorer la qualité de vie des patients atteints de troubles de la parole : un exemple d'IA dans le domaine médical

Auteur : J. Farinas

Adresse : Université Paul Sabatier, IRIT Équipe SAMOVA, 31062 Toulouse cedex 9

Adresse électronique : Jerome.farinas@irit.fr

Affiliation : Maître de conférence à l'Institut de Recherche en Informatique de Toulouse à l'Université Toulouse 3

Mots clés (français) : parole pathologique, intelligibilité, mesure automatique, démonstrateur

Résumé (français) : La mesure automatique de l'intelligibilité est une fonctionnalité qui permet d'améliorer la pratique de soin de personnes présentant des altérations de la parole. La modélisation par modèle de neurones profonds permet une mesure objective et instantanée. Un démonstrateur installé à l'hôpital est en train d'être évalué afin d'envisager une utilisation dans les soins courants. Un tel système permettrait également une évolution des techniques de remédiation et une amélioration de la qualité de vie des patients présentant des handicaps au niveau de la communication.

Mots clés (anglais) : pathological speech, intelligibility, automatic measurement, demonstrator

Résumé (anglais) : Automatic intelligibility measurement is a feature that improves the care of people with speech impairments. Deep neural model modeling enables objective and instantaneous measurement. A demonstrator installed in a hospital is currently being evaluated for use in routine care. But such a system would also enable an evolution in remediation techniques and improve the quality of life of patients with communication disabilities.

Mots clés (espagnol) : palabra patológica, inteligibilidad, medidas automáticas, demostrador

Résumé (espagnol) : La medida automática de la inteligibilidad es una función que puede mejorar la atención a las personas con deficiencias del habla. La modelización mediante modelos neuronales profundos ofrece una medición objetiva e instantánea. Actualmente se está evaluando un demostrador instalado en un hospital con vistas a su utilización en la asistencia cotidiana. Pero un sistema de este tipo también permitiría evolucionar las técnicas de remediación y mejorar la calidad de vida de los pacientes con discapacidades de comunicación.

La modélisation automatique de l'intelligibilité pour améliorer la qualité de vie des patients atteints de troubles de la parole : un exemple d'IA dans le domaine médical

1. Introduction

Les troubles de la parole concernent différents types de pathologies qui portent sur la voix et le langage. La parole est un mouvement volontaire qui trouve sa source au niveau cérébral. Les troubles de la parole peuvent être d'ordre neurologique. Par exemple cela peut correspondre aux accidents vasculaires cérébraux ou à des maladies neurodégénératives, comme la maladie de Parkinson ou la sclérose latérale amyotrophique. Parfois ce sont les muscles qui sont atteints, comme dans le cas des myopathies ou de myasthénie grave. Enfin les différents éléments de la sphère ORL (lèvres, bouche, langue, pharynx, palais, larynx) peuvent présenter des altérations et influencer grandement sur la compréhensibilité de la parole produite.

Les troubles de la parole entraînent généralement une déficience fonctionnelle au niveau de la communication. Cet aspect est d'autant plus attendu après le traitement de pathologies qui affectent spécifiquement la voix, comme dans le cas des cancers de la tête et du cou. Les cancers de la cavité buccale et de l'oropharynx ont une incidence élevée en France (voir (Defossez, 2019) pour plus de détails) : plus de 13 500 nouveaux cas ont été déclarés en 2018 (dont environ trois quarts chez les hommes). Il s'agit du troisième cancer le plus incident, notamment chez les hommes avec 18,3 cas pour 100 000 personnes-années (taux standardisé mondial, contre 5,8 chez les femmes). En fonction de la zone touchée, des répercussions fonctionnelles majeures sur les voies aérodigestives supérieures sont susceptibles d'apparaître, affectant la respiration, la déglutition et la capacité de phonation. Une perte d'intelligibilité de la parole est donc fréquemment constatée, ce qui a un impact sur la qualité de vie du patient (De Graeff, 2000).

Étant donné la nature dégénérative de la majorité de ces troubles de la parole, un diagnostic et un suivi précoces sont généralement corrélés à un meilleur pronostic (Lim, 2015) (Leifer, 2003), car ils permettent une mise en œuvre progressive et programmée des mesures post-traitement. Par conséquent, l'évaluation perceptive de l'intelligibilité de la parole et de la gravité des troubles de la parole constitue une méthode clé pour l'évaluation de la parole pathologique (Miller, 2013). La définition de l'intelligibilité n'est pas une notion simple et l'on peut rencontrer différentes conceptions : elle peut être vue comme la proportion de discours compris (Keintz, 2007), mais également comme le taux de mots correctement transcrits (Hustad, 2008), (Pommée, 2021). Il s'agit donc d'une mesure fortement liée à la perception par des tiers, ce qui la rend difficile à quantifier, en raison des stratégies de compensation humaines qui renforcent le décodage du signal par des indices reposant sur la structure et le sens de la langue. La gravité des troubles de la parole sert d'indice alternatif dans les contextes cliniques. Elle est basée sur le degré d'altération de l'intelligibilité associé à d'autres variables du signal de parole telles que la qualité d'émission du code acoustico-phonétique, la vitesse d'élocution et d'autres paramètres temporels et/ou prosodiques pertinents pour la difficulté perçue (Kent, 1992). La gravité des troubles peut être considérée comme un concept plus large, incluant l'intelligibilité et d'autres stratégies de compensation. Dans ces conditions, l'intelligibilité et la sévérité peuvent être considérées comme des mesures similaires servant des objectifs distincts, qui sont également intrinsèquement subjectifs (Fex, 1992). Cette subjectivité va normalement au-delà du biais et de la variance que l'on trouve habituellement dans les évaluations cliniques, puisqu'elle peut également se

retrouver dans les éléments qui composent ces deux mesures (Klopfenstein, 2009) (Kent, 2003).

Il est donc difficile pour un médecin d'arriver à porter un jugement objectif sur l'état de la qualité de la voix d'un patient, car à force de proximité, l'être humain s'habitue aux altérations de la voix et finit par deviner ce qui est prononcé. Afin d'augmenter la stabilité de la mesure, celle-ci peut être traitée par un groupe d'experts (Woisard et Lepage, 2010). Néanmoins, le recours à un groupe d'experts pour obtenir un score objectif complique le suivi des patients. L'utilisation d'un outil automatique, calculé sur un échantillon de parole enregistré, pourrait fournir une solution objective et libérer du temps pour les experts. Dans cet article, la section 2 va permettre de présenter des solutions possibles afin d'arriver à une modélisation automatique de la mesure de l'intelligibilité de la parole. Dans la section 3, il sera présenté une première expérimentation clinique d'un dispositif permettant d'évaluer automatiquement l'intelligibilité.

2. Modélisation automatique de l'intelligibilité

L'automatisation du calcul de mesure de l'intelligibilité, passe par la disponibilité de données pour alimenter les apprentissages. Le Groupement d'Intérêt Scientifique Parolothèque vise à faciliter l'accès d'enregistrements audio de patients atteint de cancers. Sur le modèle des thumorotheques, cela permet aux chercheurs un accès facilité aux données issues des hôpitaux. La voix est une donnée identifiante, et donc tombe sur le coup du « Règlement Général sur la Protection des Données ». Pour accéder à ces données, il est nécessaire de disposer d'un consentement de la personne pour l'exploitation scientifique de ses enregistrements, de l'informer sur ses droits, notamment au niveau de la rétractation possible de l'usage de ses données dans les études, et d'assurer un traitement sécurisé de ces données. Différents intervenants assurent ces droits : le Comité de la Protection des Personnes, les Comités d'Éthique, les délégués à la Protection des Données, la Commission Nationale Informatique et Libertés. La Parolothèque permet d'accélérer et de faciliter l'accès aux données dans le cadre de projets scientifiques. Plusieurs projets de recherche ont permis d'alimenter en données la base de données disponible : le projet INCA C2SI (Carcinologic Speech Severity Index : mesure de l'impact des traitements des cancers de la cavité buccale et du pharynx sur la production de la parole par l'Indice de sévérité des troubles de la production de la parole) (Woisard, 2021), le projet ANR RUGBI (Recherche d'unités linguistiques pertinentes pour améliorer la mesure de l'intelligibilité de la parole altérée par des troubles de production pathologique), le projet Programme Hospitalier de Recherche Infirmière et Paramédicale DAPADAF-E (Validité d'une tâche de Décodage Acoustico-Phonétique Automatisée sur les Déficits Anatomo-Fonctionnels dans l'Évaluation paramédicale des troubles de la parole des patients traités pour un cancer de la cavité buccale ou de l'oropharynx). Cela représente en 2024 les données de 288 personnes, enregistré suivant des protocoles permettant d'évaluer différents aspects de la communication (parole lue, prononciation de pseudo mots, tâche prosodique, tâche de compréhension, parole spontanée). De nombreuses métadonnées ont également été collectées afin de compléter ces enregistrements : questionnaires sur la communication, la qualité de vie, mesures perceptives de la sévérité et de l'intelligibilité de la parole (réalisé par un groupe de 6 experts).

Les premières études ont permis de valider la construction d'un score permettant de modéliser les mesures de sévérité et d'intelligibilité à partir des questionnaires des patients

(Balaguer, 2019a), puis l'impact sur la communication (Balaguer, 2021). Puis des études ont évalué la faisabilité d'une automatisation de ces mesures à partir d'indices calculés sur les enregistrements audio (Balaguer, 2019b) (Woisard, 2022).

Des études utilisant des modélisations par réseaux de neurones profonds ont également été réalisées. Cela a permis de valider l'usage de ces méthodes, qui nécessitent en général de nombreuses données pour être apprises, sur les corpus de parole de patients atteints de cancers (Roger, 2022a) (Roger, 2022b) (Quintas, 2022). Il a également été étudié l'explicabilité des scores, en analysant les organisations internes dans les réseaux de neurones profonds (Abderrazek, 2022).

3. Mise en œuvre d'un démonstrateur à l'hôpital

Les modélisations automatiques mises en place par Roger et Quintas, nécessite en général l'exécution sur des stations de travail disposant de matériel spécifique tel que des cartes graphiques dédiées, pour pouvoir être exécutées. Un projet de pré-maturation soutenu par la Société d'accélération du transfert de technologies Toulouse Tech Transfert (prototpitch 2020) a permis de réaliser un prototype pouvant être déployé et testé à l'hôpital : le Système Automatisé de Mesure de l'Intelligibilité (Roger, 2022c).



Figure 1 : exemple d'utilisation de la tablette du démonstrateur SAMI

SAMI est constitué d'un serveur installé à l'hôpital, qui permet de calculer des scores d'intelligibilité avec les modèles de réseaux de neurones profonds développés par Roger (2022b) et Quintas (2022). Les données sont acquises grâce à une tablette et sont transmises au serveur pour évaluation. La figure 1 illustre un cas d'usage. Un tel système permet un accès quasi instantané à la mesure de l'intelligibilité. L'interface, illustrée sur les figures 2 et 3, a été conçue pour être facile à utiliser et intuitive.

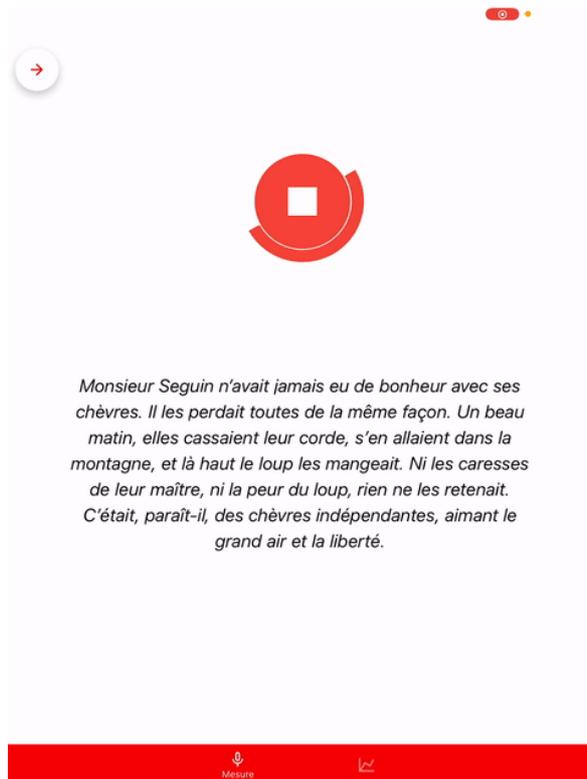


Figure 2 : Capture écran de l'interface d'enregistrement de la tablette SAMI

La figure 2 présente une capture de l'écran de la tablette, vue par le patient, pendant l'enregistrement. Les utilisateurs sont invités à enregistrer une tâche de lecture de passage, pour laquelle un score d'intelligibilité et de sévérité, compris entre 0 et 10, est renvoyé par la suite. Les utilisateurs peuvent facilement sélectionner, effacer et ajouter de nouveaux patients.



Figure 3 : Capture écran du suivi de la mesure de l'intelligibilité d'un patient sur plusieurs séances

L'entrée d'un nouveau patient est composée de son nom complet, d'un IP (index du patient) et de sa date de naissance. Tous les enregistrements vocaux sauvegardés sont stockés dans un espace de stockage asynchrone crypté en AES 256. Un deuxième onglet est à la disposition du praticien (illustré sur la figure 3). Il lui permet de visualiser l'historique des mesures du patient sélectionné. Il peut trier les mesures par ordre chronologique ou par score. Le praticien peut également choisir d'appliquer des filtres aux données du patient, afin de n'afficher que les données qui l'intéressent. Enfin, le praticien peut facilement supprimer une mesure. Le dispositif SAMI est actuellement en cours d'évaluation au CHU Larrey de Toulouse.

4. Conclusion

Le démonstrateur SAMI présenté dans cet article constitue un exemple d'utilisation d'Intelligence Artificielle dans un cadre clinique. Cela illustre un aspect de ce que les nouvelles technologies peuvent apporter aux soins. Le sujet a été débattu dans la table ronde « Les nouvelles technologies et l'intelligence artificielle favorisent-elles réellement l'autonomie, l'accessibilité des espaces sociaux et, par là même, une plus grande effectivité des droits ? » qui a eu lieu le 6 décembre 2023 dans le cadre du 41ème congrès scientifique de la FISAF à la cité Internationale Universitaire à Paris. L'usage de nouvelles technologies peut permettre d'apporter des avantages. Dans le cas de SAMI, la mise à disposition de scores automatiques permet d'avoir un retour instantané sur la mesure automatique de l'intelligibilité de la parole, et donc de ne pas nécessiter la mobilisation de plusieurs experts pour la réalisation d'une mesure perceptive. Ce dispositif permet également d'envisager de nouveaux usages : par exemple, un suivi qui pourrait être réalisé en dehors de l'hôpital et qui permettrait une meilleure prise en compte de l'évolution d'un patient. Mais on pourrait

également l'intégrer à des pratiques de remédiation ou réhabilitation afin de permettre une meilleure remise sur pied du patient. Les patients qui sont affecté par des cancers de la tête et du cou présentent en général de très grave altérations de la parole. Ceci a une très forte incidence sur la qualité de vie de la personne. La perte soudaine de la voix comme moyen de communication est très handicapante dans la vie de tous les jours. Le gain de temps dans la remédiation permettrait d'offrir une meilleure vie à ces personnes. Bien évidemment, l'usage de nouvelles technologies ne peut être utilisé sans une analyse de fiabilité complète, afin de se prémunir de biais de modélisation, mais également d'évaluer les erreurs de modélisations. Mais l'usage des technologies de la voix constitue des avancées pour de nombreuses personnes présentant des handicaps. Ces technologies sont maintenant disponibles avec des degrés de maturités assez élevés, et permettent d'envisager de nouveaux usages. Les technologies grand public sont matures car elles utilisent de très nombreuses ressources (par exemple Whisper proposé par Open AI nécessite plus de 880 000 heures d'enregistrements pour réaliser son apprentissage). Malheureusement elles présentent des performances très mauvaises quand il s'agit de traiter des enregistrements de personnes présentant des altérations de la voix. La recherche dans le domaine de la parole pathologique permet d'être plus inclusif et de proposer une prise en charge des handicaps.

5. Références

Abderrazek, S., Fredouille, C., Ghio, A., Lalain, M., Meunier, C., & Woisard, V. (2022).

Interpreting Deep Representations of Phonetic Features via Neuro-based Concept Detector: Application to speech disorders due to Head and Neck cancer. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 1–14.

doi:10.1109/TASLP.2022.3221039

Balaguer, M., Boisguerin, A., Galtier, A., Puech, M., Farinas, J., & Woisard, V. (2019a, May).

Facteurs influençant l'intelligibilité et la sévérité du trouble chronique de la parole des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. Retrieved from

<https://hal.science/hal-02919628>

Balaguer, M., Farinas, J., Pinquier, J., & Woisard, V. (2019b, August). Construction of an

automatic Carcinologic Speech Severity Index (C2SI) score. *31st World Congress of the International Association of Logopedics and Phoniatrics (IALP 2019)*. Retrieved from

<https://hal.science/hal-02977799>

Balaguer, M., Pinquier, J., Farinas, J., Lepage, B., & Woisard, V. (2021, May). *Construction*

d'un index holistique d'impact sur la communication des troubles de la parole chez les

patients traités pour un cancer oral ou oropharyngé. Retrieved from

<https://hal.science/hal-03258582>

Defossez, G., Le Guyader-Peyrou, S., Uhry, Z., Grosclaude, P., Colonna, M., Dantony, E., ...

Monnereau, A. (2019). *Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Volume 1 : Tumeurs solides :*

Étude à partir des registres des cancers du réseau Francim (p. 371 p.). Saint Maurice,

France : Santé Publique France.

De Graeff, A., De Leeuw, R. J., Ros, W. J. G., Hordijk, G.-J., Blijham, G. H., & Winnubst, J.

A. M. (2000). Long-Term Quality of Life of Patients With Head and Neck Cancer. *The Laryngoscope*, *110*(1), 98–106. doi:10.1097/00005537-200001000-00018

Fex, S. (1992). Perceptual evaluation. *IEEE Transactions on Acoustics, Speech and Signal*

Processing, *6*, 155–158.

Keintz, C. K., Bunton, K., & Hoit, J. D. (2007). Influence of visual information on the

intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*,

16(3), 222–234. doi:10.1044/1058-0360(2007/027)

Kent, Raymond D. (1992). *Intelligibility in Speech Disorders*. Amsterdam/Philadelphia: John

Benjamins Publishing Company.

Kent, Ray D., & Lim, Y.-J. (2003). Toward an acoustic typology of motor speech disorders.

Clinical Linguistics & Phonetics, *17*(6), 427–445. doi:10.1080/0269920031000086248

Klopfenstein, M. (2009). Interaction between prosody and intelligibility. *International*

Journal of Speech-Language Pathology, *11*(4), 326–331.

doi:10.1080/17549500903003094

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility

scores for speakers with dysarthria. *Journal of Speech, Language and Hearing*

Research, *51*(3), 562–573. doi:10.1044/1092-4388(2008/040)

- Leifer, B. P. (2003). *Early Diagnosis of Alzheimer's Disease: Clinical and Economic Benefits*. *51(5)*, 281–288.
- Lim, H., Kim, D. H., Jung, H.-Y., Gong, E. J., Na, H. K., Ahn, J. Y., ... Kim, J.-H. (2015). Clinical Significance of Early Detection of Esophageal Cancer in Patients with Head and Neck Cancer. *Gut Liver*, *9(2)*, 159–165. doi:10.5009/gnl13401
- Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, *48(6)*, 601–612. doi:10.1111/1460-6984.12061
- Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021). Intelligibility and comprehensibility: A delphi consensus study. *International Journal of Language & Communication Disorders*, *57(1)*, 21–41. doi:10.1111/1460-6984.12672
- Quintas, S. (2022). *Deep learning approaches to assess speech intelligibility of head and neck cancer* (Université Paul Sabatier - Toulouse III). Retrieved from <https://theses.hal.science/tel-04094765>
- Roger, V., Farinas, J., & Pinquier, J. (2022a). Deep neural networks for automatic speech processing: a survey from large corpora to limited data. *EURASIP Journal on Audio, Speech, and Music Processing*, (article 19), 1–15. doi:10.1186/s13636-022-00251-w
- Roger, V., Farinas, J., Woisard, V., & Pinquier, J. (2022b, June). Création d'une mesure entropique de la parole pour évaluer l'intelligibilité de patients atteints de cancers des voies aérodigestives supérieures. *34èmes Journées d'Études Sur La Parole (JEP 2022)*, 117–125. doi:10.21437/JEP.2022-13
- Roger, V. (2022c). *Modélisation de l'indice de sévérité du trouble de la parole à l'aide de méthodes d'apprentissage profond : d'une modélisation à partir de quelques exemples à un apprentissage auto-supervisé via une mesure entropique* (Université Paul Sabatier - Toulouse III). Retrieved from <https://theses.hal.science/tel-03935738>

Woisard, V., & Lepage, B. (2010). Perception of speech disorders: Difference between the degree of intelligibility and the degree of severity. *Audiological Medicine*, *8*, 171–178. doi:10.3109/1651386X.2010.525375

Woisard, V., Astésano, C., Balaguer, M., Farinas, J., Fredouille, C., Gaillard, P., ... 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. doi:10.1007/s10579-020-09496-3

Woisard, V., Balaguer, M., Fredouille, C., Farinas, J., Ghio, A., Lalain, M., ... Lepage, B. (2022). Construction of an automatic score for the evaluation of speech disorders among patients treated for a cancer of the oral cavity or the oropharynx: The Carcinologic Speech Severity Index. *Head & Neck*, *44*(1), 71–88. doi:10.1002/hed.26903