



HAL
open science

Représentation automatique du rythme de la parole pathologique via le spectre de modulations d'amplitude

Robin Vaysse, Corine Astésano, Jérôme Farinas

► To cite this version:

Robin Vaysse, Corine Astésano, Jérôme Farinas. Représentation automatique du rythme de la parole pathologique via le spectre de modulations d'amplitude. 9èmes Journées de Phonétique Clinique (JPC 2023), Université de Toulouse, Jun 2023, Toulouse, France. pp.59-61. hal-04189576

HAL Id: hal-04189576

<https://hal.science/hal-04189576>

Submitted on 28 Aug 2023

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.

Représentation automatique du rythme de la parole pathologique via le spectre de modulations d'amplitude

Robin VAYSSE^{1,2}
Corine ASTESANO²
Jérôme FARINAS¹

¹IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3,
Toulouse, France

²Laboratoire de Neuro-Psycho-Linguistique LNPL, Université
Toulouse II, France

Une bonne structuration prosodique en groupements cohésifs hiérarchisés optimise la segmentation de la parole et l'accès au sens pour l'auditeur. Parmi les trois principes organisateurs de la parole (intonation, accentuation et rythme ; Di Cristo, 2011), le rythme constitue le socle de la prosodie puisqu'il permet l'intégration temporelle des proéminences accentuelles et des unités intonatives en lien avec les règles métriques sous-jacentes d'une langue donnée. Les processus de programmation de la parole (pathologique ou non) induisent des disfluences caractérisées par des ruptures du rythme de la parole, pouvant nuire à la compréhension et à l'intelligibilité de la parole. Caractériser le rythme linguistique est donc un enjeu fondamental en linguistique et en traitement automatique de la parole. Au-delà des méthodes basées sur le rythme linéaire (notamment pour la classification rythmique automatique des langues), il est nécessaire de développer des méthodes permettant de rendre compte de l'intégration temporelle de la hiérarchie prosodique (Arvaniti, 2012).

Notre recherche propose une représentation automatique du rythme à partir du signal de parole. Cette méthode permet de visualiser les répartitions d'énergie des différents niveaux prosodiques : syllabes, mot prosodique (pw ; Astésano, 2019), syntagme accentuel (AP), syntagme intonatif (IP ; Di Cristo, 2011). Le spectre de modulation d'amplitude (Envelope Modulation Spectrum – EMS) est obtenu en calculant l'enveloppe du signal auquel est appliqué un filtre 300-1000 Hz afin de capturer l'énergie

des voyelles (Tilsen & Johnson, 2008 ; Vaysse et al., 2020 ; voir courbe noire, figure 1). Un spectre de puissance est ensuite appliqué pour représenter les fréquences de 0 à 10 Hz (courbe orange). Un lissage (courbe rouge pointillée) est également appliqué pour englober les constituants prosodiques de niveau similaire. La détection automatique des constituants prosodiques a été validée par une annotation prosodique manuelle sur un échantillon du corpus (10 patients et 10 contrôles).

Le corpus de texte lu par 137 témoins, 86 patients atteints de cancers VADS et 205 patients atteints de la maladie de Parkinson est issu du projet RUGBI (Woisard et al., 2021). La figure 2 illustre la répartition spectrale rythmique sur les deux populations de patients.

La répartition énergétique sur l'EMS permet de visualiser les stratégies rythmiques des patients (préservation des cibles articulatoires au détriment de la fluence rythmique par une structuration essentiellement aux niveaux prosodiques inférieurs *vs* préservation de la fluence et des groupes prosodiques hiérarchisés). Nous présenterons les différents profils rythmiques caractérisant les pathologies et nous ouvrirons sur les applications potentielles pour l'évaluation automatique longitudinale des patients.

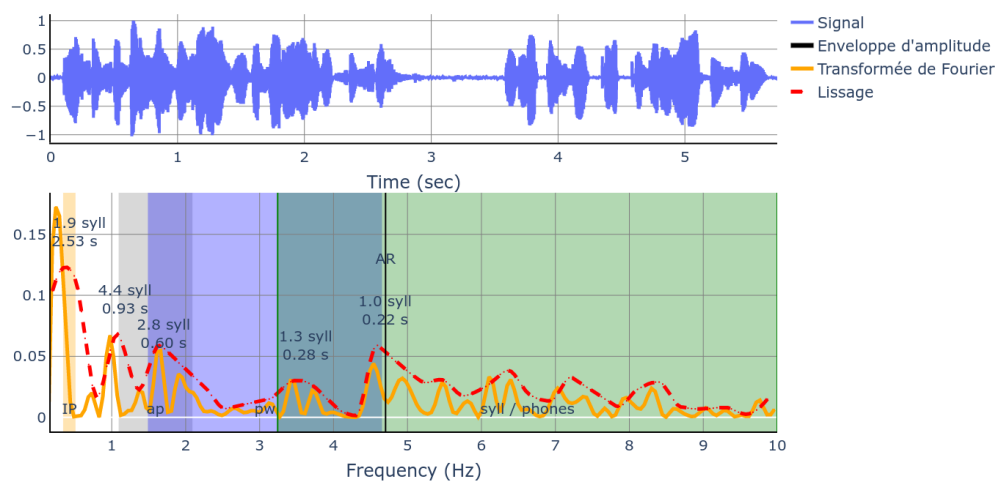


Figure 1 : EMS d'un locuteur contrôle sur l'extrait "Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon". Les intervalles correspondant aux niveaux prosodiques annotés manuellement sont indiqués en couleur : orange pour l'IP, gris

pour l'ap, bleu pour le pw et vert pour la syllabe. AR (*articulation rate*) désigne le débit articulatoire.

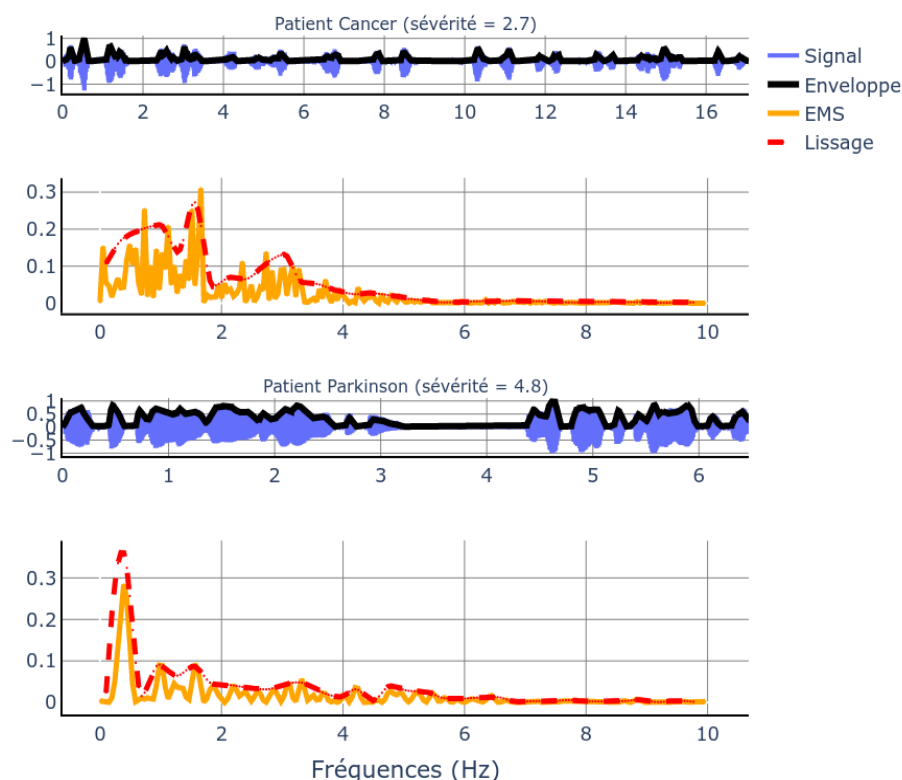


Figure 2 : Comparaison des EMS de deux patients sur la lecture des phrases "Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon.". L'EMS du sujet cancer VADS est en haut, celui du patient Parkinson est en bas. Les patients sont des personnes avec une sévérité de maladie élevée. L'énergie des spectres de modulations pour ces patients est plus concentrée vers les basses fréquences que pour le locuteur contrôlé sur la figure 1.

Références bibliographiques

- Di Cristo, A. (2011). Une approche intégrative des relations de l'accentuation au phrasé prosodique du français. *Journal of French Language Studies*, 21(1), 73-95.
- Arvaniti, A. 2012. The Usefulness of Metrics in the Quantification of Speech Rhythm. *Journal of Phonetics*, 3, 351–373.
- Astésano, C. (2019) The prosodic word as the domain of French accentuation - Empirical evidence. *Phonetics and Phonology in Europe, PaPE 2019*, Lecce : 170-171.
- Tilsen, S. & Johnson, K. (2008). Low-frequency fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*, 124(2): 34–39. <https://doi.org/10.1121/1.2947626>

- Vaysse, R., Farinas, J., Astésano, C., André-Obrecht, R. (2021) Automatic Extraction of Speech Rhythm Descriptors for Speech Intelligibility Assessment in the Context of Head and Neck Cancers. *Interspeech 2021*, 1912-1916, <https://doi.org/10.21437/Interspeech.2021-1736>
- Woisard, V. & Astésano, C. & Balaguer, M. & Farinas, J. & Fredouille, C. & Gaillard, P. & Ghio, A. & Giusti, L. & Laaridh, I & Lalain, M. & Lepage, B. & Maclair, J. & Nocaudie, O. & Pinquier, J. & Pouchoulin, G. & Puech, M. & Robert, D. & 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, 173-190. <https://doi.org/10.1007/s10579-020-09496-3>