Modélisation automatique du rythme de la parole

Stage Master 2 ou 3ème année école d'ingénieur en informatique

Contexte : La prosodie de la parole regroupe de nombreuses fonctions dans la parole. Elles sont en général laissées de côté lors de la reconnaissance automatique de la parole, car les systèmes cherchent à optimiser la reconnaissance des sons et des mots. Mais si vous souhaitez connaître les émotions qui sont encodées dans le signal de parole, mais également les informations sur le style de parole (modalité de la phrase, registre de parole…) ou l'attitude des locuteurs (adhésion, conviction, doute, invitation…) il est nécessaire de s'intéresser à la mélodie, au rythme et à l'accentuation qui constituent la prosodie. Les grandeurs physiques présentent dans le signal sont bien connues, mais la matérialisation de la prosodie dans le signal n'est pas simple et résultat d'interactions complexes qui font l'objet de nombreuses recherches dans le domaine [1]. Nous proposons dans ce stage, de proposer une automatisation de la modélisation du rythme, afin de pouvoir proposer des représentations et des caractérisations qui pourront être utiles aux linguistes et aux personnes travaillant sur le traitement automatique de la parole.

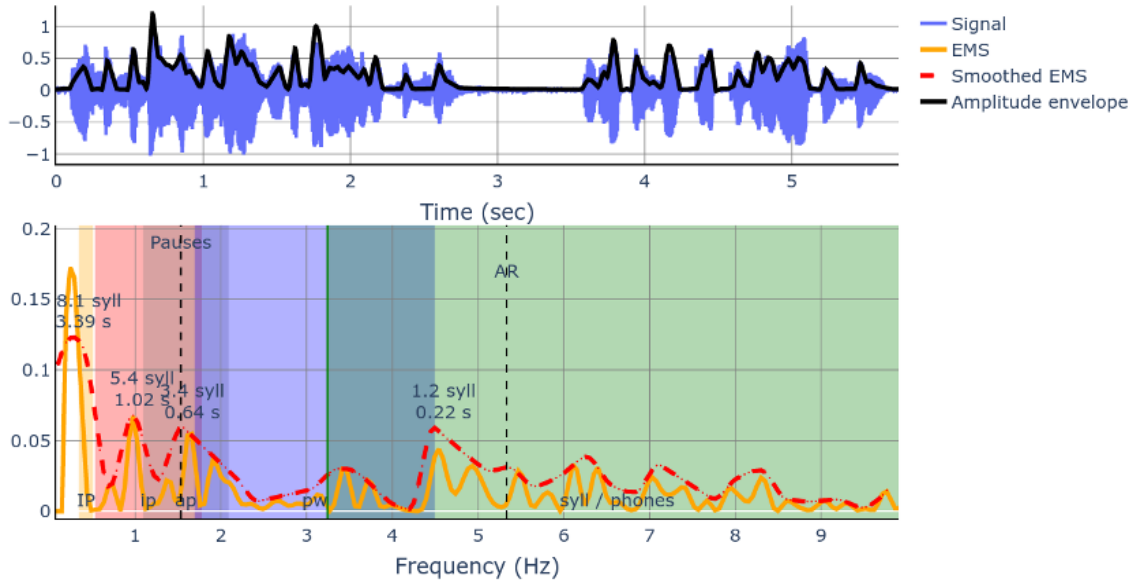

Figure 1 : EMS sur l'extrait *"Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon"*. Les intervalles correspondants aux niveaux prosodiques annotés manuellement sont indiqués en couleur : orange pour l'IP, rouge pour l'ip, gris pour l'ap, bleu pour le pw et vert pour la syllabe.

Objectif : Proposer une représentation obtenue automatiquement à partir du signal du rythme de la parole. Ce stage constitue une continuation des travaux du doctorat de Robin Vaysse [3]. Il a proposé une représentation par spectre de modulation d'amplitude

(Envelope Modulation Spectrum – EMS). Cette méthode permet de visualiser les répartitions d'énergie du rythme de la parole. L'EMS est obtenu en calculant l'enveloppe du signal auquel est appliqué un filtre 300-1000 Hz afin de capturer l'énergie des voyelles ([5, 6]; 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. Sur la figure 1, les niveaux prosodiques sont représentés en couleur et sont issus d'une annotation manuelle : syllabes, mot prosodique (pw ; [4]), syntagme accentuel (AP), syntagme intermédiaire (ip), syntagme intonatif (IP) ; [1].

Le stage se basera sur cette représentation, et proposera une détection automatique des zones des niveaux prosodiques. Il sera également combiné à cette représentation, pour les fréquences de 0 à 4 Hz, une représentation issue de la courbe de l'intonation, ce qui permettrait d'améliorer la précision dans cette zone. La représentation sera testée sur de nombreux styles de parole pour en évaluer la robustesse et la pertinence de cette représentation.

Localisation : le stage aura lieu au Laboratoire de Recherche en Informatique de Toulouse sur le campus de l'université Toulouse III Paul Sabatier.

Encadrants :

- Jérôme Farinas, UT3, laboratoire IRIT, jerome.farinas@irit.fr
- Corine Astésano, UT2, laboratoire LNPL, corine.astesano@univ-tlse2.fr

Compétences : Les outils seront développés en python. Une expérience sur le traitement du signal et en linguistique serait un plus.

Références :

[1] 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.

[2] Arvaniti, A. 2012. The Usefulness of Metrics in the Quantification of Speech Rhythm. *Journal of Phonetics, 3*, 351–373.

[3] Vaysse, R. (2023) *Caractérisation automatique du rythme de la parole : application aux cancers des voies aéro-digestives supérieures et à la maladie de Parkinson.* Sciences de l'information et de la communication. Université Paul Sabatier - Toulouse III, 21 mars 2023. https://theses.hal.science/tel-04198849

[4] Astésano, C. (2019) The prosodic word as the domain of French accentuation - Empirical evidence. *Phonetics and Phonology in Europe, PaPE 2019*, Lecce : 170-171.

[5] 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

[6] 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