Sujet de doctorat Equipe SAMOVA/IRIT Université Paul Sabatier 2016-2017

Enrichissement et amélioration de transcriptions automatiques de la parole par des informations non verbales

Contexte

Les système de transcription automatique de parole grand vocabulaire ont été démocratisés assez récemment auprès du grand public, suite à l'amélioration des performances apportées par les modélisation acoustiques utilisant des réseaux de neurones profonds [1]. Le signal acoustique est découpés en tronçons et est analysé pour découvrir la séquence de sons élémentaires qui le constitue. Les informations prosodiques et non verbales ne sont en général pas utilisées à ce niveau. Or la prosodie joue un rôle très important dans la structuration des énoncés et le codage d'informations sur l'état émotionnel du locuteur. La fréquence fondamentale, l'intensité, la vitesse d'élocution, les pauses entre mots participent grandement à l'information véhiculée par le signal acoustique.

Il existe de nombreuses théories [2,3,4] qui proposent des modélisations linguistiques de la prosodie. Mais celle-ci n'est pas utilisée en totalité par les systèmes de reconnaissance de la parole.

Objectif

L'objectif général de ce travail de doctorat est d'approfondir la relation entre les modélisations classiques de systèmes de reconnaissance automatique de la parole et les modélisations automatiques des informations non verbales/prosodiques, et de proposer des améliorations en vue d'enrichir les transcriptions résultantes mais également augmenter la qualité de la transcription.

Le travail de doctorat consistera dans un premier temps à un approfondissement des théories prosodiques existantes et de l'analyse de leur transposition en vue de la réalisation de modélisations automatiques.

Une modélisation de la prosodie sera appliqué à la problématique de la ponctuation et à la segmentation des énoncés. L'analyse des informations non verbales permettra de segmenter une séquence acoustique avec les marques de séparation classiques virgule, point, point virgule. Ce découpage permettra d'améliorer le décodage du système de reconnaissance automatique de la parole en relançant une nouvelle passe de décodage acoustique.

Dans un second temps, les séquences acoustiques seront analysées au niveau prosodiques pour extraire des informations sur la modalité des phrases et autres informations non verbales. Cela permettra d'aboutir à un enrichissement des transcriptions mais également à l'adaptation des modélisations utilisées par les systèmes de reconnaissance de la parole.

Profil du candidat

Candidat titulaire d'un master en informatique ou en traitement du signal ayant des compétences en programmation. Des notions sur l'apprentissage automatique, le traitement du signal et la reconnaissance des formes seraient un plus. Un esprit scientifique, créatif, et aimant le travail en équipe est souhaitable.

Contact

Jérôme Farinas, jerome.farinas@irit.fr, 0561558343, https://goo.gl/d3iqSY

Références

1. G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," in IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, Nov. 2012. (doi: 10.1109/MSP.2012.2205597)
2. Crombie, W. (1987) "Intonation in English: A systematic perspective".
3. Albert Di Cristo. La prosodie au carrefour de la phon´etique, de la phonologie et de l'articulation formes-fonctions. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), Laboratoire Parole et Langage, 2004, 23, pp.67-211. <hal-00285554>
4. Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). "The original ToBI system and the evolution of the ToBI framework". In S.-A. Jun, ed., Prosodic Typology: The Phonology of Intonation and Phrasing, pp. 9-54. Oxford University Press.