

Proposition de travaux de thèse, IRISA/LLF, août 2016

Titre : «*Caractérisation et la génération de l'expressivité en fonction des styles de parole pour la construction de livres audio* »

Dans le cadre de projets ANR portant sur la modélisation de la prosodie et sur l'utilisation de modèles prosodiques en synthèse de parole, une thèse sur la «*Caractérisation et la génération de l'expressivité en fonction des styles de parole pour la construction de livres audio* » est proposée.

Contexte : Pour obtenir une voix de synthèse de qualité utilisable dans des contextes particuliers, il est fondamental d'améliorer l'expressivité de la parole qui constitue un vecteur d'émotion, d'intention ou d'état d'esprit. Une part importante de l'expressivité de la voix est liée au contexte d'élocution et est notamment influencée par le type de texte lu. Tous ces éléments participent à ce que l'on peut nommer des styles de parole. Les poèmes, les contes, les discours politiques, ou les journaux télévisés sont des textes dont l'oralisation se fait selon des styles différents, et de nombreux lecteurs, s'ils lisent des documents intégrant plusieurs types de textes ou style de parole, sont capable d'adapter leur élocution aux textes à oraliser.

Dans le cadre de la thèse, le candidat travaillera à la caractérisation des styles de parole à partir de l'étude de différents paramètres prosodiques (rythme, intonation, etc.) et segmentaux (réalisation des segments, liaisons, etc.). Les résultats de ces analyses serviront de base à la construction de modèles permettant aux systèmes de synthèse de générer des styles de parole divers. L'objectif global du projet est d'améliorer le contrôle et le rendu expressif des systèmes de synthèse de la parole, notamment pour les systèmes par concaténation. L'enjeu d'un tel sujet concerne tout particulièrement la création de nouvelles applications dans les domaines du jeu vidéo, de l'apprentissage, du divertissement ou encore du médical.

Proposition de travail : Le traitement de l'expressivité dans la parole et l'adaptation de la prosodie à des styles particuliers constituent des questions de recherche importante à l'heure actuelle. Des études très récentes comme [1], mettent en avant le manque de naturel et de qualité dans la parole synthétique expressive. Concernant les styles de parole, [2] propose un modèle permettant la génération de quelques genres de discours. Dans [3] sont présentés quelques résultats d'une étude récente visant à déterminer les principaux éléments caractéristiques de quelques genres en vue de les re-synthétiser.

Partant de ces travaux, le candidat se focalisera sur les changements de styles, comme, par exemple, lors de passage au style direct, et sur l'expression par la parole de certaines émotions comme la colère, la peur, etc. Pour cela, le candidat s'intéressera à l'expression dans des livres audio car ces types de textes permettent de regrouper certaines de ces caractéristiques. Le déroulement de la thèse pourra se faire sur la base suivante :

- sélection des données pertinentes (livres audio existants ou enregistrements de textes spécifiques) ;
- analyse de ces données afin d'en extraire un ensemble de descripteurs pertinents sur la base des travaux existants dans le domaine l'expressivité et la prosodie de la parole ;

- proposition de modèles prosodiques adaptés aux différents styles/ types de texte ;
- intégration de ces modèles dans le processus de synthèse de parole.

Environnement de travail : Ce travail de thèse sera dirigé conjointement par un linguiste spécialiste de prosodie (Elisabeth DELAIS-ROUSSARIE, UMR 7110-LLF, Université Paris-Diderot) et par un informaticien spécialiste de synthèse de parole (Damien LOLIVE, IRISA – ENSSAT Lannion, Université Rennes 1). Il bénéficiera d'un financement sur trois ans. Le candidat sera d'ailleurs rattaché aux deux établissements.

Pendant la thèse, le candidat devra séjourner assez longuement à Lannion pour travailler en étroite collaboration avec les spécialistes de synthèse de parole. Il sera accueilli au sein de l'équipe EXPRESSION du département Multimédia et Interactions de l'IRISA qui cible l'étude des données produites par l'homme (en particulier langagières) et véhiculées par différents médias : le geste ou le mouvement, la parole et le texte. Un des principaux axes de recherche de cette équipe concerne la synthèse de la parole expressive. Dans ce cadre, l'équipe dispose d'une forte expérience dans le domaine et de collaborations dans les domaines complémentaires (linguistique et prosodie). L'équipe dispose d'un moteur de synthèse de la parole par corpus, d'un moteur statistique (HTS) et également de moyens techniques d'enregistrement qui lui permettront de faciliter la génération de données utiles à la caractérisation de l'expressivité.

Parallèlement, le candidat devra collaborer avec des linguistes spécialisés en phonétique, en phonologie, en prosodie, mais aussi en discours. Ces collaborations seront menées au sein de Labex EFL, qui regroupe plusieurs laboratoires de recherche en linguistique tels le Laboratoire de Linguistique Formelle (UMR 7110-LLF, Université Paris-Diderot), le Laboratoire de Phonétique et Phonologie (UMR 7018-LPP, Université Paris 3), Alpage (UMR INRIA, Université Paris-Diderot).

Informations pratiques : Peuvent candidater à cette thèse des étudiants ayant un master recherche en linguistique (spécialité phonétique et phonologie, prosodie, discours), en linguistique informatique ou en informatique intéressé par un travail collaboratif, par le traitement automatique de la parole et par le machine learning.

Contacts : Elisabeth DELAIS-ROUSSARIE (elisabeth.roussarie@wanadoo.fr) et Damien LOLIVE (damien.lolive@irisa.fr)

Début souhaité : septembre ou octobre 2016

Durée : 36 mois

Références

[1] D Govind & SR Mahadeva Prasanna. Expressive speech synthesis : a review. *International Journal of Speech Technology*, pages 1–24, 2013.

[2] Nicolas Obin. *MeLos : Analysis and Modelling of Speech Prosody and Speaking Style*. Theses, Université Pierre et Marie Curie - Paris VI, June 2011.

[3] Mathieu Avanzi, George Christodoulides, Damien Lolive, Elisabeth Delais-Roussarie, and Nelly Barbot. Towards the adaptation of prosodic models for expressive text-to-speech synthesis. In *Proceedings of Interspeech*, 2014.