

## Phd proposal, IRISA/LLF, August 2016

**Title:** “Characterisation and generation of expressivity in function of speaking styles for audiobook synthesis”

In the context of the ANR project SynPaFlex dealing with prosody modelling and the use of prosodic models in speech synthesis, a Phd thesis on the characterisation and generation of expressivity in function of speaking styles for audiobook synthesis is proposed.

**Context:** To obtain a quality synthetic voice usable in specific contexts, it is fundamental to improve speech expressivity, which constitutes a vector of emotion, intention or mood. A large part of voice expressivity is related to the elocution context and is influenced, particularly, by the type of text. Poems, tales, political speeches or newscasts are texts for which the speaking style is different. Generally, readers adapt their speaking style to the text they’re reading, even when reading documents mixing several types of text.

The Phd candidate will contribute to the characterisation of speaking styles from the study of different prosodic parameters (rhythm, intonation, etc.) and segmental parameters (segments realisation, liaisons, etc.). These results will be used as a basis to build models enabling synthesis systems to generate various speaking styles. The main goal of the project is to improve the control over expressivity in speech synthesis, in particular for unit-selection and hybrid systems. Among the potential outcomes, new applications in the fields of video games, learning, advertisement or healthcare are expected.

**Work proposal:** Expressivity processing in speech and prosody adaptation to specific styles are important research questions. Recent studies, like [1], show the lack of natural and quality in expressive synthetic speech. Concerning speaking styles, [2] proposes a model to generate a few discourse styles. In [3], some results of a study whose objective is to determine the main characteristics of some speaking styles are presented.

From these studies, the Phd candidate will focus on style changes, as, for instance when changing to narrative style, and also on the expression of some emotions in speech (like anger, fear, etc.). To do this, the Phd student will study expression in audio books as this type of text gathers some characteristics of interest. A probable planning for this thesis might be as follows:

- Selection of relevant data (existing audio books or recording of specific texts);
- Data analysis and extraction of relevant features based on existing studies in the fields of expressivity and speech prosody;
- Proposition of prosodic models adapted to the different speaking styles/types of text;
- Integration of these models in the speech synthesis process.

**Environment:** This Phd will be directed jointly by a linguist, specialist in Prosody (Elisabeth DELAIS-ROUSSARIE, UMR 7110-LLF, Université Paris-Diderot) and by a computer scientist, specialist in Speech Synthesis (Damien LOLIVE, IRISA – ENSSAT Lannion, Université Rennes 1). The financial support is running for 3 years.

The student will be located in Lannion and will be member of the Expression team, Media and Interaction Department at IRISA, which studies the expressivity in gesture, speech and text. One of the main research axes of the team is Expressive Speech Synthesis. In this context, the team has a strong experience in the field and also collaborations in complementary fields. The team has a state of the art text-to-speech system and a recording booth to facilitate the generation of useful data to characterise expressivity.

At the same time, the student will have to collaborate with linguists, specialists in phonetics, phonology, prosody, and also in discourse analysis. These collaborations are available through the Labex EFL, which gathers several research labs in linguistics such as the Laboratoire de Linguistique Formelle (UMR 7110-LLF, Université Paris-Diderot), the Laboratoire de Phonétique et Phonologie (UMR 7018-LPP, Université Paris 3), Alpage (UMR INRIA, Université Paris-Diderot).

**Practical information:** Anyone with a Master's degree in Linguistics or Computer Science can apply. Good skills in Computer Science and in Speech and Language Processing are required in any case. Personal Interest in Artificial Intelligence, Machine Learning and Deep Learning is a clear plus.

**Contacts:** Elisabeth DELAIS-ROUSSARIE ([elisabeth.roussarie@wanadoo.fr](mailto:elisabeth.roussarie@wanadoo.fr)) and Damien LOLIVE ([damien.lolive@irisa.fr](mailto:damien.lolive@irisa.fr))

**Start in:** October 2016

**Duration:** 36 months

**References:**

- [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.