Parametric Speech Synthesis with Deep Neural Networks

Mots clés : Array

- Directeur de thèse : Axel Roebel
- Co-encadrant(s) :
- Unité de recherche : Sciences et Technologies de la Musique et du Son
- Ecole doctorale : Ecole Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Sciences et technologies de l'information et de la communication

Résumé du projet de recherche (Langue 1)

The project proposes to investigate into new models for parametric speech synthesis that allow unlocking the full potential of deep neural networks to create parametric representation of speech signals. The parametric speech synthesizer to be developed should integrate specific knowledge about speech and singing signals without introducing the approximations and simplifications that are used in traditional parametric speech models. The formulation of the speech model should allow training the models with training databases that are significantly smaller than the 20h that are currently required for Tacotron 2. Transfer learning strategies, that benefit from the huge amount of speech databases that are publicly available today might be employed. The parametric speech synthesizer to be developed should be integrated and evaluated in the existing speech and singing synthesis software that is maintained by the Analysis/Synthesis team of IRCAM and support.