

## Extraction de voix en temps-réel par approche Deep Learning

Sujet proposé par Maxime BAELEDE (A-Volute / MODAL, INRIA Lille - Nord Europe)

### Encadrants

- Maxime BAELEDE, ingénieur R&D, doctorant (A-Volute / INRIA) maxime.baelde@a-volute.com
- Raphaël GREFF, directeur R&D (A-Volute) raphael.greff@a-volute.com

### Contexte

A-Volute est une société basée à Roubaix qui édite des logiciels d'amélioration audio pour des PC sur le marché gaming, notamment le logiciel Nahimic présent sur les PC du fabricant MSI. A-Volute a développé une solution de son 3D entièrement numérique et en temps-réel. La suite d'effets audio proposée par A-Volute comporte notamment des effets d'amélioration de la voix dans des contenus multimédia (musique ou film), ainsi que des effets micro pour la communication (réduction de bruit). C'est dans ce contexte que se tient ce projet IMPACT.

### Problématique

L'extraction de voix fait partie d'un domaine plus générale qui est la séparation de sources sonores. Celle-ci consiste à extraire les différentes sources sonores présentes dans un signal audio (mono ou multicanal). Beaucoup de techniques de séparation existent mais fort peu parviennent à fonctionner en temps-réel (car non conçu dans ce but ou à cause de la charge computationnelle).

### Sujet

Il s'agit de développer un algorithme d'extraction de signaux vocaux en temps-réel par un approche de type Deep Learning [1]. Le projet commencera par un état de l'art des publications scientifiques (voire brevets) relatives à ce sujet [2], [3]. L'aspect temps-réel est un point essentiel du projet : il s'agit de modifier un algorithme existant ou en créer un nouveau qui puisse fonctionner avec très peu d'informations audio et dont la charge computationnelle soit la plus faible possible. Le prototypage de l'algorithme sur Matlab ou Python en fin de projet serait un plus.

### Références

- [1] I. GOODFELLOW, Y. BENGIO et A. COURVILLE. *Deep Learning*. 2016.
- [2] A. A. NUGRAHA, A. LIUTKUS et E. VINCENT. « Multichannel audio source separation with deep neural networks ». In : *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2016), p. 1–1.
- [3] D. WANG et J. CHEN. « Supervised Speech Separation Based on Deep Learning : An Overview ». In : *arXiv preprint arXiv :1708.07524* (2017).