

Classification de signaux audio 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. Un de ces produits phares est le Sonic Radar : celui-ci analyse de manière aveugle la scène sonore et indique les directions de sources sonores prédominantes. Dans le cadre d'une thèse CIFRE, on souhaite qu'il indique également le type de son et qu'il puisse séparer plusieurs sources sonores. C'est dans ce contexte que se tient ce projet.

Problématique

La classification de signaux audio repose sur des techniques de Machine Learning. Celles-ci consistent à chercher des attributs pertinents pour décrire le signal : dans notre contexte il s'agit des descripteurs audio [3]. Ensuite une technique de Machine Learning est utilisée, basée sur ces descripteurs (Deep Learning ou autre) [2]. Néanmoins ces techniques ne sont toujours pas adaptées pour le temps-réel ou les données de type musique ou jeux vidéos.

Sujet

Un algorithme de classification temps-réel a été développé chez A-Volute [1]. Il s'agit ici de réfléchir une implémentation à base de Deep Learning de manière à pouvoir être utilisée en amont d'une technique de séparation de sources sonores, et en temps-réel. L'objectif minimal du projet consiste à proposer une étude des différentes structures de réseaux de neurones pertinentes pour l'objectif, et si possible une implémentation Matlab ou Python pour en tester la validité.

Références

- [1] M. BAELEDE, C. BIERNACKI et R. GREFF. « A mixture model-based real-time audio sources classification method ». In : 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Mar. 2017, p. 2427–2431.
- [2] E. CAKIR et al. « Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection ». In : *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25.6 (juin 2017), p. 1291–1303.
- [3] G. PEETERS et al. « The Timbre Toolbox : extracting audio descriptors from musical signals ». In : *The Journal of the Acoustical Society of America* 130.5 (nov. 2011), p. 2902–2916.