

Extraction de sources sonores dans un flux multi-canal en temps-réel.

Sujet proposé par Maxime BAELE (A-Volute / MODAL, INRIA Lille - Nord Europe)
Stage de 6 mois niveau master 2, métropole lilloise, année 2019

Encadrants

- Maxime BAELE, ingénieur R&D, doctorant (A-Volute / INRIA) maxime.baelde@a-volute.com
- Nathan SOUVIRAÀ-LABASTIE, ingénieur R&D (A-Volute) nathan.souviraà-labastie@a-volute.com
- Raphaël GREFF, directeur R&D (A-Volute) raphael.greff@a-volute.com

Contexte

A-Volute est une société basée à Villeneuve d'Ascq qui édite des logiciels d'amélioration audio pour des PC sur le marché gaming, en particulier 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 des contenus multimédia (musique ou film), ainsi que des effets micro pour la communication (réduction de bruit). C'est dans ce contexte que se tiendra ce stage.

Problématique

L'extraction de sources dans un flux multi-canal peut être vue comme à la croisée entre la séparation et la localisation de sources sonores. La séparation de sources consiste à extraire les différentes sources sonores présentes dans un signal audio (distribution fréquentielle) alors que la localisation de sources consiste à trouver la distribution spatiale des sources sonores. Beaucoup de techniques de séparation et de localisation 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 sources sonores dans un flux multi-canal en temps-réel par un approche à choisir (par exemple de type Deep Learning [1]). Le projet commencera par un état de l'art des publications scientifiques (voire brevets) relatives aux deux sujets [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 prototype de l'algorithme sur Matlab ou Python en fin de projet serait un plus.

Profil recherché

Qui recherchons nous ? En préparation d'un diplôme d'ingénieur ou d'un master (bac+5), voire d'un doctorat (bac+8) (stage de césure), tu disposes de préférence de connaissance dans le développement et l'implémentation d'algorithmes avancés de traitement numérique du signal audio.

En outre, des notions avancées dans les domaines variés suivants seraient fortement appréciés : - Audio, acoustique et psychoacoustique - Effets audio de manière générale : compression, égalisation, etc. - Machine learning et réseaux de neurones artificiels. - Statistiques, mathématiques probabilistes, optimisation. - Programmation et développement informatique : Matlab, Python.

Et des expériences dans les domaines suivants seraient des plus : - Effets de spatialisation sonore : synthèse binaurale, Ambisonics, réverbération artificielle. - Reconnaissance vocale, commande vocale. - Effets de traitement de la voix : réduction de bruit, annulation d'écho, traitement d'antenne. - Réalité virtuelle, augmentée et mixte. - Programmation et développement informatique : Max/MSP, C/C++/C#. - Moteurs de jeux vidéo : Unity, Unreal Engine, Wwise, FMod, etc. - Logiciels d'édition audio : Audacity, Adobe Audition, etc. - Publications scientifiques et dépôts de brevet. - Maitriser parfaitement l'anglais. - Faire preuve de curiosité intellectuelle.

Autres offres

<https://nahimic.welcomekit.co/>

<https://www.welcometothejungle.co/companies/nahimic/jobs>

Références

- [1] I. GOODFELLOW, Y. BENGIO et A. COURVILLE. *Deep Learning*. 2016.
- [2] N. TAKAHASHI et al. « PhaseNet : Discretized Phase Modeling with Deep Neural Networks for Audio Source Separation ». In : *Interspeech 2018*. Interspeech 2018. ISCA, 2 sept. 2018, p. 2713-2717.
- [3] Z.-Q. WANG, X. ZHANG et D. WANG. « Robust TDOA Estimation Based on Time-Frequency Masking and Deep Neural Networks ». In : *Interspeech 2018*. Interspeech 2018. ISCA, 2 sept. 2018, p. 322-326.