



Séparation de sources audio - Postprocessing des résultats sur critères subjectifs et psychoacoustiques.

Sujet proposé par Nathan SOUVIRAA-LABASTIE (A-Volute / Villeneuve d'Ascq)
Stage de 6 mois niveau master 2, métropole lilloise, année 2019

Encadrants

- Nathan SOUVIRAA-LABASTIE, ingénieur R&D (A-Volute) nathan.souviraa-labastie@a-volute.com
- Damien GRANGER, ingénieur R&D (A-Volute) damien.granger@a-volute.com
- Maxime BAELDE, 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 industriel

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 de master.

Sujet

Problématique

La séparation de sources consiste à extraire les différentes sources sonores présentes dans un signal audio, notamment en estimant leurs distributions fréquentielles. Beaucoup de techniques de séparation existent mais la plupart cherchent à optimiser un critère objectif, par exemple la divergence d'ITAKURA-SAITO. La plupart des techniques de séparation de sources audio font ensuite appel à une étape de masquage temps fréquence. Dans de nombreux cas, cette étape induit des artefacts (gazouillis) perceptibles à l'oreille mais non pris en compte par les critères objectifs. De même, les résultats (signaux séparés) sont eux aussi très souvent évalués de façon objective, par exemple par des métriques comme le rapport signal à distorsion ou artefact [5]. Certaines approches d'évaluation subjectives existent mais sont pour l'instant minoritaires [1], [2]

Approche

Le choix de l'approche peut dépendre des aspirations de l'étudiant. Une première approche peut porter sur un traitement temporel du problème, par exemple avec une correction de phase sur les signaux de résultats. Une deuxième approche peut porter sur le post-traitement des coefficients du masque temp-fréquence, i.e. avant de repasser dans le domaine temporel. D'autres approches sont envisageables comme la modification de l'étape "classique" d'*overlap and add* pour mieux prendre en compte les aspects subjectifs. Par ailleurs des pistes de réflexion peuvent être trouvées dans [4] et l'utilisation d'algorithmes issus du *Deep Learning* [3] est aussi envisageable.

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. - Maîtriser parfaitement l'anglais. - Faire preuve de curiosité intellectuelle.

Autres offres

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

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

Références

- [1] M. CARTWRIGHT, B. PARDO et G. J. MYSORE. « Crowdsourced pairwise-comparison for source separation evaluation ». In : p. 5.
- [2] V. EMIYA et al. « Subjective and Objective Quality Assessment of Audio Source Separation ». In : *IEEE Transactions on Audio, Speech, and Language Processing* 19.7 (sept. 2011), p. 2046-2057.
- [3] I. GOODFELLOW, Y. BENGIO et A. COURVILLE. *Deep Learning*. 2016.
- [4] A. M. KREME, V. EMIYA et C. CHAUX. « Phase reconstruction for time-frequency inpainting ». In : *International conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*. Guildford, United Kingdom, juil. 2018.
- [5] E. VINCENT, R. GRIBONVAL et C. FEVOTTE. « Performance measurement in blind audio source separation ». In : *IEEE Transactions on Audio, Speech, and Language Processing* 14.4 (juil. 2006). 5*, p. 1462-1469.