

Proposition de sujet d'IMPACT, Ecole Centrale de Lille, option DAD

Fonctions de coût subjectives pour la génération automatique de karaoké.

Sujet proposé par Nathan SOUVIRAA-LABASTIE (A-Volute / Villeneuve d'Ascq)

Encadrants

- Nathan SOUVIRAA-LABASTIE, ingénieur R&D (A-Volute) nathan.souviraa-labastie@a-volute.com
- Raphaël GREFF, directeur R&D (A-Volute) raphael.greff@a-volute.com
- Antoine LIUTKUS, chargé de recherche (inria) antoine.liutkus@inria.fr

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 tient ce projet IMPACT.

Sujet

Problématique

La séparation de sources consiste à extraire les différentes sources sonores présentes dans un signal audio, par exemple en estimant leurs distributions fréquentielles. Beaucoup de techniques de séparation existent (et notamment récemment des approches basées sur le deep learning) mais la plupart cherchent à optimiser un critère objectif, par exemple la divergence d'ITAKURA-SAITO ou encore des métriques énergétiques comme le rapport signal à distorsion ou artefact [4]. Cependant, ces métriques rendent peu compte de la qualité perçue par l'auditeur alors que celle-ci devrait être l'objectif d'un tel algorithme. Par exemple, la plupart des techniques de séparation de sources audio font appel à une étape de masquage temps fréquence, et dans de nombreux cas, cette étape induit des artefacts (gazouillis) perceptibles à l'oreille mais non pris en compte par les critères objectifs. De plus, l'application visée étant un générateur automatique de karaoké, la qualité perceptive finale est primordiale pour A-Volute.

Certaines approches automatiques d'évaluation subjectives¹ existent mais sont pour l'instant minoritaires [1], [2]. En particulier, l'algorithme de prédiction des notes subjectives développé dans [2]² qui est pourtant l'état de l'art du domaine est particulièrement lent et ne peut pas être utilisé comme fonction de coût au cours de l'apprentissage d'un algorithme de séparation de sources car son utilisation ferait augmenter de près de 7000% le temps d'apprentissage (ce qui est rédhibitoire lorsque l'on sait que le temps d'un apprentissage complet sur cette tâche est de l'ordre de plusieurs semaines).

Approche

Dans de précédents travaux, A-Volute a réussi à développer un algorithme *end-to-end* de prédiction des notes subjectives ayant les mêmes performances que [2] mais avec un impact normalement minime

1. Généralement, l'évaluation subjective des résultats de séparation de sources audio est effectuée de façon non automatique c'est à dire par un humain, e.g., par le biais de test MUSHRA où l'évaluation consiste à attribuer une note entre 0 et 100 à chaque extrait audio.

2. Approche par calcul de *features* fréquentielles suivit d'un MLP à une couche de 8 neurones

sur le temps d'apprentissage. L'objectif de cet IMPACT est de démontrer que cette approche améliore les résultats de séparation en terme de perception, *i.e.* subjectivement. Il faudra dans un premier temps convertir l'algorithme de Keras à Pytorch, vérifier que les mêmes performances de prédictions sont atteintes, afin d'ensuite s'en servir comme fonction de coût au sein de l'apprentissage d'un algorithme de l'état de l'art [3]³ (lui même en pytorch).

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] F.-R. STOTER et al. « Open-Unmix - A Reference Implementation for Music Source Separation ». In : *Journal of Open Source Software* (2019).
- [4] 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.

3. <https://github.com/sigsep/open-unmix-pytorch>