



Génération automatique de version Karaoké.

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

Encadrants

- Nathan SOUVIRAÀ-LABASTIE, ingénieur R&D (A-Volute) nathan.souviraalabastie@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

Une version Karaoké d'un morceau de musique est une version où la voix du chanteur n'est plus présente dans le morceau. Généralement une telle version de la musique est présentée avec un sous-titrage des paroles permettant à l'utilisateur de chanter au rythme du morceau "instrumental".

La plupart du temps, ces versions Karaoké sont générées ("masterisées") à la main par un ingénieur du son. Les sociétés de divertissement ont déjà des bases de données bien fournies de ce genre de contenu. Cependant, elles ne peuvent pas faire face à la quantité de morceaux créés chaque jour notamment par les musiciens amateurs et doivent se concentrer sur les morceaux les plus connus. Ainsi, un outil de génération automatique de version Karaoké permettrait au grand public d'accéder à une base de données potentiellement infinie de version Karaoké. De même, dans le cas d'un contenu écouté en streaming un outil automatique (et temps réel) serait également nécessaire.

Approche

Une première tâche sera de mettre en oeuvre une méthode de l'état de l'art de séparation de sources audio. La méthode sera à discuter selon les aspirations de l'étudiant (ML [2], empirique type REPET [1]). Une base de donnée et les résultats d'une campagnes d'évaluation de séparation de musique est disponible dans [3].

Une deuxième tâche sera de générer un sous-titrage synchronisé avec le morceaux de musique. Pour la majorité des morceaux, les paroles sont disponibles en ligne en version texte brute. Il s'agirait alors de synchroniser et d'afficher cette version sur la base d'une comparaison avec la version speech-to-text de la source de voix qui a été retirée. Des compétences ou expériences en Natural Language Processing (NLP) ou en traitement de données symboliques seraient un plus.

Le sujet pourra être étendu sur une thèse de doctorat dont la mission consistera en la continuité de ce projet.



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] Z. RAFII et al. « An Overview of Lead and Accompaniment Separation in Music ». In : *IEEE TRANSACTIONS ON AUDIO* (), p. 31.
- [2] N. TAKAHASHI et al. « Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Detection ». In : *arXiv :1604.07160 [cs]* (25 avr. 2016).
- [3] D. WARD et al. « SISEC 2018 : state of the art in musical audio source separation - Subjective selection of the best algorithm ». In : *4th Workshop on Intelligent Music Production, WIMP 2018*. University of Huddersfield, sept. 2018.