



Détection de voix de synthèse.

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
- 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). Une des missions de l'équipe R&D est d'explorer de nouveaux concepts en vue des futurs produits d'A-Volute.

Problématique

La récente progression du réalisme des moteurs de synthèse de voix (Cortana, Siri, Alexa ...) ouvre la voie à des nouvelles interactions entre l'homme et la machine. Certaines grandes entreprises ont par exemple déjà annoncé le lancement à grandes échelles la prise de rdv téléphoniques pour des assistants virtuels¹. De même, la synthèse des émotions dans les voix de synthèse² ou encore le *morphing* de voix³ vont devenir des technologies matures dans les années à venir.

Sujet

Une première étape de ce sujet de stage est de faire une large revue bibliographique des méthodes envisageables pour la création d'un détecteur de voix de synthèse. Une introduction du domaine dit de l'"anti-spoofing" peut être trouvée dans [6]. Ce domaine de recherche s'attache à développer des contre-mesures pour sécuriser la vérification automatique du locuteur. C'est un domaine proche de celui de la reconnaissance du locuteur. Des pistes de réflexion pour les futures implémentations pourront aussi être trouvées dans la littérature du traitement de signal tenant compte ou informé par la phase [2, 7].

Des campagnes d'évaluation⁴ relatives à ces aspects sont organisées tout les deux ans sur différentes tâches. ASVSpooof2015 [4] s'est d'abord concentré sur l'évaluation des capacités des systèmes à s'adapter à de nouvelles menaces d'usurpation. Un résumé des méthodes soumises à ce challenge et leur résultats est disponible [5]. ASVSpooof2017 [1] s'est ensuite focalisé sur la menace potentiellement la plus répandue qui consiste à utiliser une enceinte lors de l'usurpation. Il est à noter qu'en février 2019, le challenge adoptera la métrique t-DCF [3] et les tâches ne sont pas encore dévoilées. Par ailleurs, chaque challenge fournit des bases de données d'apprentissage et d'évaluation qui pourront servir comme données de travail.

Une fois identifié le type de méthode le plus adapté à notre cas d'usage (critère de performance de détection, de temps réel et de simplicité d'intégration logicielle), une deuxième tâche sera d'implémenter une ou plusieurs solutions de l'état de l'art. Enfin, à partir de cette base, des améliorations seront à apporter sur les critères cités ci-dessus ou sur l'extension au cas multi-canal de la méthode choisie.

1. <https://www.koreus.com/video/google-assistant-rdv-coiffeur.html>
2. <http://recherche.ircam.fr/anasyn/emotionrecognition/main.php>
3. <https://modulate.ai/>
4. <http://www.asvspooof.org/>



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

Autres offres

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

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

Références

- [1] H. DELGADO et al. « ASVspoof 2017 Version 2.0 : meta-data analysis and baseline enhancements ». In : *Odyssey 2018 The Speaker and Language Recognition Workshop*. Odyssey 2018 The Speaker and Language Recognition Workshop. ISCA, 26 juin 2018, p. 296-303.
- [2] H. ERDOGAN et al. « Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks ». In : *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ICASSP 2015 - 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). South Brisbane, Queensland, Australia : IEEE, avr. 2015, p. 708-712.
- [3] T. KINNUNEN et al. « t-DCF : a Detection Cost Function for the Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification ». In : *Odyssey 2018 The Speaker and Language Recognition Workshop*. Odyssey 2018 The Speaker and Language Recognition Workshop. ISCA, 26 juin 2018, p. 312-319.
- [4] Z. WU et al. « ASVspoof 2015 : Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan ». In : (), p. 5.
- [5] Z. WU et al. « ASVspoof : The Automatic Speaker Verification Spoofing and Countermeasures Challenge ». In : *IEEE Journal of Selected Topics in Signal Processing* 11.4 (juin 2017), p. 588-604.
- [6] Z. WU et al. « Spoofing and countermeasures for speaker verification : A survey ». In : *Speech Communication* 66 (fév. 2015), p. 130-153.
- [7] N. ZHENG et X.-L. ZHANG. « Phase-Aware Speech Enhancement Based on Deep Neural Networks ». In : *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27.1 (jan. 2019), p. 63-76.