

Question : Le domaine fréquentiel des applications pour la voix et des applications pour la musique sont différents. Ceci pourrait avoir un impact sur notre modèle.

Réponse : Oui : Le domaine de la musique requiert une fidélité vers des plus hautes fréquences que la voix. Cependant l'oreille humaine localise difficilement les sons de hautes fréquences, ce n'est donc pas rédhibitoire à l'utilisation des modèles pour la voix.

Organisation : Création d'un google sheets pour la décision de l'organisation de la prise de son et des microphones : https://docs.google.com/spreadsheets/d/1y-oqidjZmegUV0tEP882KaG_TSIUaYdtXZx79op04gk/edit?usp=sharing . Aussi, voir : <https://lossenderosstudio.com/article.php?subject=17> .

Proposition : création d'un dictionnaire de notes par instrument ?

Conclusion : Oui. Utilisable en données d'entraînement. *Décider des répétitions à faire : 15 fois la même note ou varier le timbre / jeu?* Utilisable pour les algorithmes (M)NMF.

Question : Est-ce possible, avec ce dictionnaire de retrouver la réponse de la salle au rayonnement de chaque instrument ? Est-ce possible d'utiliser un dictionnaire de notes pour imiter un sweep et obtenir une réponse de la salle approchée pour le rayonnement et timbre spécifique de l'instrument ? **Question pour Fabre.**

Organisation : Préplug les micros avant l'enregistrement.

Question : Modèle demucs : Séparation multicanale, stéréo ou monocanale?

Réponse : Stéréo seulement. Il y a peut-être moyen de "tricher" et de l'insérer malgré tout dans notre approche.

Remarque : L'algorithme FastMNMF est particulièrement bon dans le cas surdéterminé, rivalisant avec voir dépassant les performances des modèles neuronaux. Possibilité de réimplémenter ces modèles avec des framework de deep learning pour faire du finetuning. Voir <https://github.com/sekiguchi92/SoundSourceSeparation> .

Remarque : Abondance de bases de données speech, qui permettent de faire de l'apprentissage supervisé. Les données bruitées sont plus adaptées pour le non-supervisé, en raison de la définition variable de "bruit" (blanc, réverbération, ambiance...).

Proposition : Utiliser FastMNMF2 pour apprendre des activations H vis-à-vis d'un dictionnaire de notes W , initialisation via une séparation initiale réalisée par Demucs. Limites de Demucs? Limites de FastMNMF2 ?

Remarque : Ne pas oublier la partie évaluation subjective ! Il suffit pour une personne de s'y pencher 2 jours.

Remarque : Enregistrements lointains difficiles à séparer.

Remarque : Pour gagner en efficacité à appliquer les algos, il faudrait jouer un peu avec.

Remarque : Papier de "Samy" : *SEMI-SUPERVISED MULTICHANNEL SPEECH ENHANCEMENT BASED ON FASTMNMF WITH A SINGLE-CHANNEL DIFFUSION MODEL*.