



UE 5AI03
Traitement Avancé du Son
Audio Labs

N. Obin

Contexte

L'analyse de scène sonore représente la capacité d'un être humain appréhender l'environnement dans lequel il évolue, pour le comprendre et interagir avec cet environnement. Les évolutions des savoirs et des technologies permettent aujourd'hui d'envisager d'automatiser cette appréhension de l'environnement sonore : l'analyse de scène auditive computationnelle (CASA).



L'analyse de scène auditive computationnelle se décompose en : localisation de sources sonores, la séparation de sources sonores, et l'extraction d'information à partir de sources sonores. L'extraction d'information à partir de sources sonores recouvre une palette extrêmement large de technologies : depuis les informations centrées sur l'humain (reconnaissance du sexe et de l'âge, état émotionnel, reconnaissance de la parole) aux informations de son environnement. L'analyse de scène auditive computationnelle prolifère aujourd'hui dans de nombreuses applications : depuis la robotique (localisation, séparation, extraction) jusque plus généralement à l'indexation de contenu audio multi-média (films, musique, web, etc...)

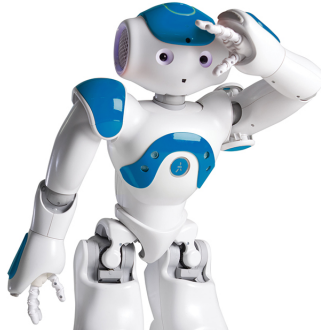
L'objectif des séances de laboratoire audio est d'explorer une thématique, depuis les aspects théoriques jusqu'à leur mise en œuvre pratique. Concrètement, les 3 séances seront centrées sur l'un des laboratoires :

- audio lab 1 : localisation de sources sonores (binaurale et antennerie) ;
- audio lab 2 : séparation de source sonores par décomposition en matrices non-négatives (NMF) ;
- audio lab 3 : indexation de contenu audio pour la reconnaissance de genre musicaux

Dans toute la suite, vous allez travailler en groupe sur l'un des laboratoires.

Ces séances donneront lieu à une évaluation continue et à une soutenance orale en fin de projet.

Audio Lab 1 : Localisation de Source Audio



L'objectif de ce laboratoire est de réaliser et d'explorer la localisation de sources sonores

Objectifs

1. Implémentation de la localisation de l'azimut d'une source par antennerie (formation de voies, PHAT, etc...)
2. Implémentation de la localisation de l'azimut d'une source binaurale (avec/sans utilisation des HRTF)

Matériel

Le matériel audio est composé de :

- Un ensemble de signaux synthétiques simulant la propagation d'une source à un azimut θ d'un réseau linéaire de microphones ou la simulation d'une écoute binaurale
- La base de données de HRTF du MIT
<http://sound.media.mit.edu/resources/KEMAR.html>

Méthodologie

La méthodologie consistera à :

- expérimenter et comparer la localisation d'une source sonore binaurale et par antennerie

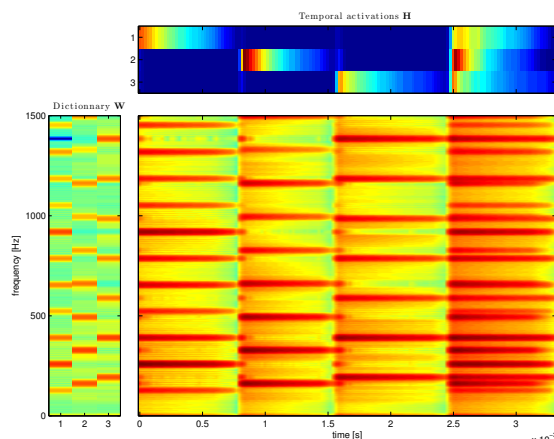
Pour comprendre et illustrer le fonctionnement de la localisation de source par formation de voies :

- représenter les caractéristiques temps-fréquence des signaux reçus à chaque microphone
- polariser un réseau de microphones dans une direction θ_0
- réaliser le filtrage des signaux après polarisation dans une direction θ_0
- estimer la position de la source à partir de polarisations successives

Requis

Calcul de la transformée de Fourier à court-terme directe

Audio Lab 2 : Séparation de Source Audio



L'objectif de ce laboratoire est d'explorer la séparation de sources audio par décomposition en matrices non-négative (NMF) pour réaliser la séparation des sources audio à partir des fichiers attachés ou d'extraits au choix.

Objectifs

1. Implémentation de la séparation de source audio par NMF non-supervisée
2. Implémentation de la séparation de source audio par NMF supervisée

Matériel

Le matériel audio est composé de :

- Un enregistrement d'un piano jouant une séquence do/mi/sol suivi d'un accord do/mi/sol
- Un morceau de musique mixé, et les pistes séparées

Méthodologie

La méthodologie consistera à :

- expérimenter la séparation de source non-supervisée sur un exemple synthétique, puis sur un exemple réel
- comparer la séparation de source non-supervisée et supervisée sur un exemple réel

Pour comprendre et illustrer le fonctionnement de la NMF :

- représenter l'évolution des dictionnaires et des activations au cours de la séparation
- représenter les dictionnaires obtenus avec/sans apprentissage
- explorer l'influence du nombre de bases du dictionnaire
- représenter l'évolution du coût au cours de la séparation
- réaliser les mesures de distorsion entre les signaux des sources audio originales et séparées

Requis

Calcul de la transformée de Fourier à court-terme directe et inverse.

Audio Lab 3 : Reconnaissance de Genres Musicaux



L'objectif de ce laboratoire est d'explorer l'indexation de genres musicaux à partir de l'apprentissage automatique de contenu audio.

Objectifs

1. Implémentation de l'indexation de genres musicaux par apprentissage automatique

Matériel

Le matériel audio est composé de :

- une collection d'enregistrements audio issus de vos bibliothèques musicales (4 genres suffisamment distincts, au moins 4 extraits par genre)
- L'utilisation de base de données de recherche, comme le GTZAN music genre collection <http://marsyas.info/downloads/datasets.html>

Méthodologie

La méthodologie consistera à :

- expérimenter la reconnaissance de genre musicaux, à partir d'un classifieur au choix
- comparer les scores de reconnaissance obtenus en fonction des paramètres du classifieur

Pour comprendre et illustrer le fonctionnement de la reconnaissance :

- créer une base de données, en partitionnant la base en : entraînement, développement, évaluation
- représenter les caractéristiques temps-fréquence des extraits audio pour chaque genre
- mesurer les scores de reconnaissance obtenus, en fonction des paramètres du classifieur
- comparer les scores de reconnaissance pour différentes représentations temps-fréquence (TFCT, MFCC, etc...)

Requis

Calcul de la transformée de Fourier à court-terme, Mel-frequency Cepstral Coefficients (MFCC)