

P11. Détection de musique dans un signal sonore

Année 2015

Encadrants techniques: Dominique PASTOR (dpt. SC), Gouenou COATRIEUX (dpt. ITI)
Encadrants gestion de projet: Rolland FLEURY, Grégoire MERCIER. **Partenaire extérieur:** Jean-Marc HANSER, Tykomz.

Mots clés : Traitement du signal, reconnaissance vocale, détection de musique, banque de sons ester, segmentation audio, indexation d'un fichier audio, classification

Résumé : Dans le cadre de ses travaux dans le domaine de la reconnaissance vocale, l'entreprise Tykomz souhaite acquérir un logiciel permettant la détection de musique dans un signal sonore. La solution développée passe par trois étapes: la segmentation automatique et/ou uniforme du signal, l'extraction de paramètres temporels et/ou fréquentiels sur chaque segment et enfin la classification des segments.

1. Présentation et contexte du projet

En présence de fond sonore, les logiciels de reconnaissance vocale peuvent s'activer de manière inadéquate, même lorsqu'il n'y a pas de voix à reconnaître, par la présence de jingle ou de musique. Dès lors, la détection de musique dans un signal sonore devient une problématique cruciale pour la phase d'apprentissage. Ce projet s'inscrit dans ce cadre en proposant une méthode permettant de détecter les portions de musique dans un signal sonore et en permettre l'indexation.

2. Méthodologie développée pour aboutir

Compte tenu de l'ampleur du sujet, le groupe s'est organisé en mode projet. Après la définition des rôles et pour des soucis d'efficacité dans la gestion de projet, le projet est subdivisé en phases, lesquelles seront composées de tâches élémentaires. Ainsi, une première phase d'état de l'art émerge dès le départ, ce qui a permis au groupe de se familiariser avec la problématique et d'aborder des pistes cohérentes pour sa résolution. Ensuite, les phases de développement technique, d'évaluation des performances et de documentation viendront clôturer le projet.

3. Développement des différentes tâches et principaux résultats

3.1 Etat de l'art :

Etape déterminante du projet, l'état de l'art, qui a débuté par une étude bibliographique, suivie par un recueil d'expertise et une approche expérimentale, a permis d'identifier les trois étages du logiciel à implémenter (voir figure ci-après).

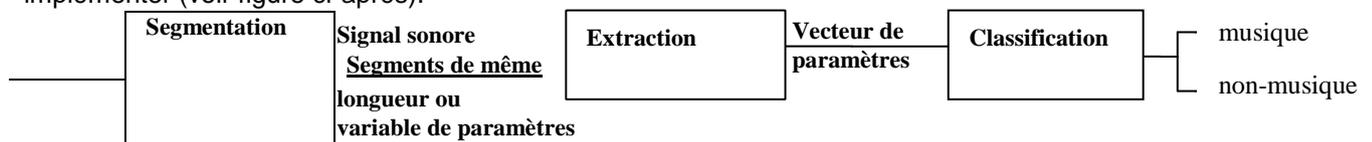


Figure 1 : Les trois étapes du traitement algorithmique

3.2 Développement

La figure 2 ci-après résume quelques résultats obtenus à l'issue de la phase de développement. La figure (a) représente un signal segmenté automatiquement en cherchant les zones de rupture des paramètres d'un modèle probabiliste (loi gamma). La musique conserve les paramètres de cette loi, ce qui donne des segments beaucoup plus longs que ceux de la parole. La figure (b) représente la modulation de vraisemblance qui montre de faibles variations pour les portions de musique contrairement aux portions de parole du fait des zones de

transition voisées-non-voisées. La figure (c) montre une méthode de classification classique utilisant les seuils. Enfin, la figure (d) représente une visualisation d'un segment audio dans un espace tridimensionnel avec en abscisse la modulation d'entropie, en ordonnée la longueur des segments et en cote la modulation de vraisemblance.

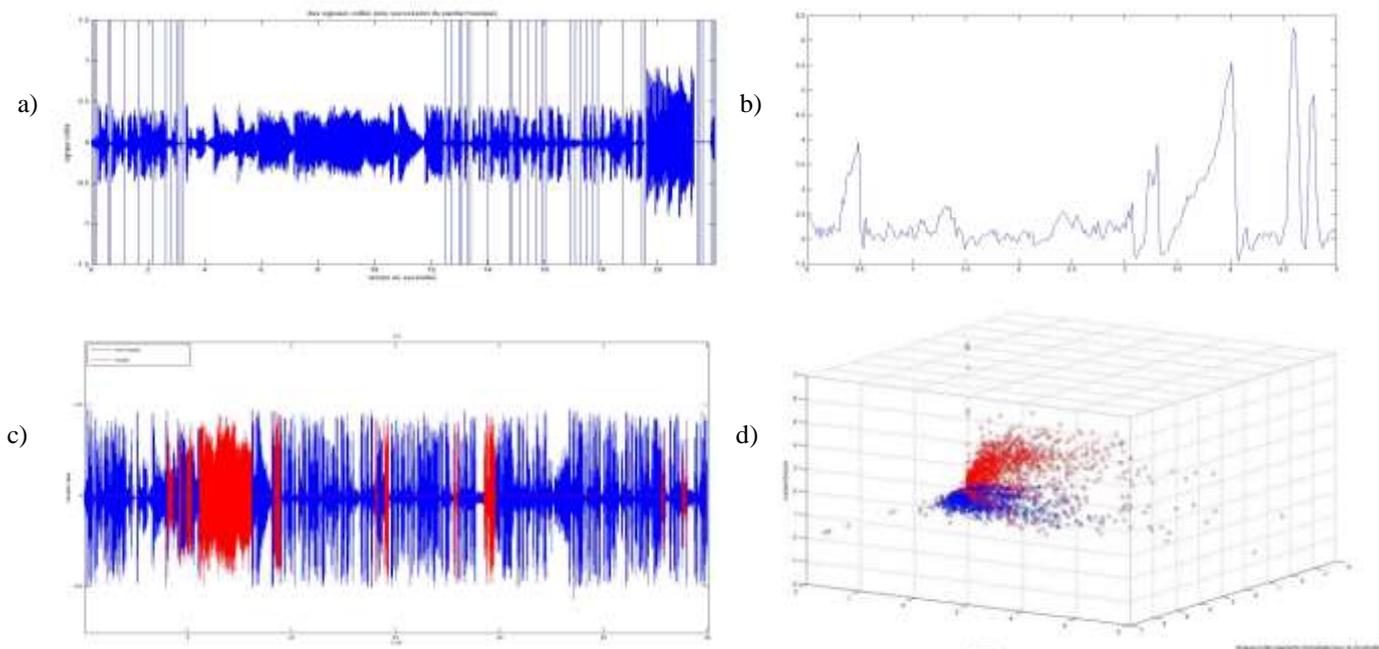


Figure 2 : les trois étapes de l'algorithme de détection. a) segmentation automatique, b) modulation de vraisemblance, c) classification par la méthode des seuils d) visualisation d'un segment audio dans un espace tridimensionnel (en rouge : musique et en bleu : non-musique).

3.3 Évaluation de performances et documentation technique :

Il s'agit donc d'évaluer les performances des méthodes étudiées ou proposées lors de l'état de l'art et tout le long du projet. Le développement de l'algorithme a été fait de sorte à permettre les tests de plusieurs méthodes ou combinaisons de méthodes ayant pour but la détection de musique. De plus, chaque méthode peut être exécutée en utilisant différents paramètres, induisant indubitablement des résultats différents. Il convient donc de choisir un jeu de test constitué de signaux dont on connaît la nature à chaque instant (i.e. où l'on aura situées les sections de musique), à choisir un ensemble de réglage, et à lancer l'algorithme plusieurs fois en faisant varier un paramètre. En comparant les résultats obtenus avec ce que l'on connaît des signaux, nous sommes en mesure de proposer un paramètre optimal pour le jeu de test dans les conditions données.

Afin de rendre l'exploitation de l'algorithme et des résultats des tests plus aisée pour Tykomz, l'évaluation des performances sera accompagnée d'une documentation technique sur le fonctionnement et l'utilisation de l'algorithme.

4. Conclusions et perspectives.

Ce projet, qui a pour but l'étude de méthodes pour la détection de musique dans un signal sonore, s'appuie donc au départ sur un état de l'art, lequel a permis d'effectuer des choix pertinents pour arriver à un algorithme permettant d'évaluer plusieurs approches dans la mesure où il a été conçu de sorte à rendre possible les tests sur d'autres méthodes qui pourront être implémentées dans le futur comme l'utilisation d'autres algorithmes de segmentation, d'autres descripteurs ou d'autres méthodes de classification. Enfin, ces travaux pourront probablement permettre à Tykomz de décider de l'intérêt d'utiliser un tel algorithme en phase d'apprentissage, ou peut-être leur ouvrir de nouvelles perspectives qui exploiterait la détection de musique.

Bibliographie:

- [1] Julien PINQUIER « Indexation sonore : recherche de composantes primaires pour une structuration audiovisuelle ». Thèse de doctorat en informatique. Université Toulouse III – Paul Sabatier, 2004, 201 p.
- [2] Abdullah I. Al-Shoshan « Speech and Music Classification and Separation: A Review ». Journal of King Saud University, Engineering Science:19(1);95-133, 2006
- [3] Zhong-hua Fu, Jhing-Fa Wang « Robust Features for Effective Speech and Music Discrimination » ROCLING. 2008

- [4] Martin R. Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors[C]//Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. IEEE, 2002, 1: I-253-I-256.
- [5] Yizhar Lavner, Dima Ruinskiy, « A decision-tree-based algorithm for speech/music classification and segmentation », EURASIP Journal on Audio, Speech, and Music Processing, Article ID 239892, 2009, 14 p.
- [6] Gholampour I, Nayebi K. « A new fast algorithm for automatic segmentation of continuous speech[C] » //ICSLP. 1998.