

I. MusicDiscover : rapport d'avancement LIRIS ECL (T0+6)

A. Avancements

1. Tache 5. Recherche de la musique par similarité acoustique

La recherche de la musique par similarité acoustique nécessite avant tout une définition. En effet, la similarité acoustique peut être obtenue à partir des caractéristiques classiques du signal sonore et des mesures de similarité convenables appliquées à celles-ci. Néanmoins, le problème est d'aboutir à des mesures de similarités qui correspondent à des aspects de similarité aperçus par des sujets humains. Dans le cadre du projet, nous nous concentrons sur les similarités aperçues par des non-musiciens.

La définition de la similarité acoustique affecte directement les méthodologies et critères d'évaluation des différents algorithmes et caractéristiques qui peuvent être proposés. Nous avons particulièrement travaillé sur des définitions de similarité qui répondent à deux critères :

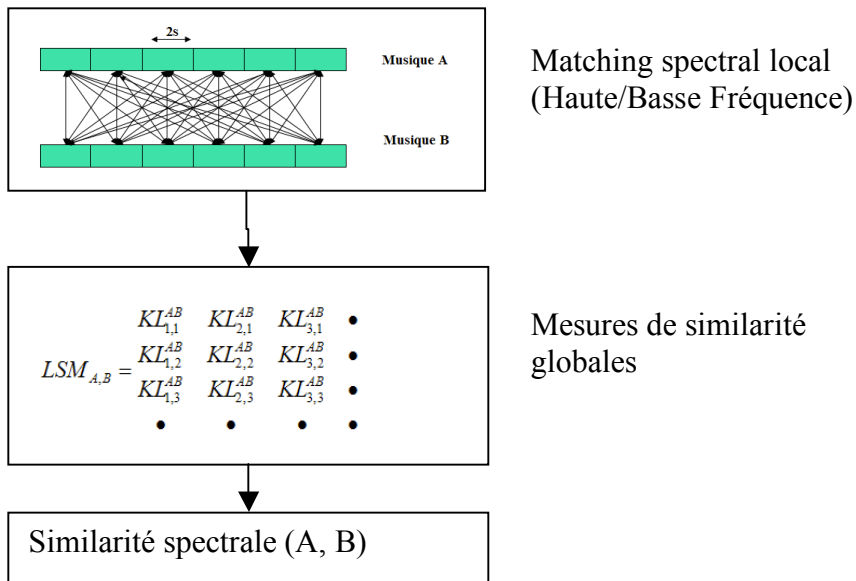
- 1- Une similarité correspondante aux producteurs, distributeurs et consommateurs non-musiciens de la musique populaire.
- 2- Une similarité qui permet la mise au point des critères d'évaluation qui engendrent le moins de subjectivité.

Nous avons donc défini la similarité dans ce cadre par :

- 1- Les titres qui correspondent aux mêmes genres, comme ils sont définis dans la tâche 7, sont considérés comme similaires.
- 2- Les titres provenant de plusieurs interprétations d'une même mélodie sont considérés comme similaires.
- 3- Les titres provenant des réinterprétations en plusieurs langues du même œuvre sont considérés comme similaires.

Des premiers travaux ont été menés sur une mesure de similarité spectrale. Ces travaux sont basés sur la modélisation du spectre par une Modélisation Gaussienne Incrémentale (MGI) et la distance de Kullback-Leibler comme mesure de similarité. L'algorithme consiste en

- 1- une estimation de similarités spectrales dites locales, pour des segments de 1 à 2 secondes
- 2- un regroupement des similarités locales pour les basses et hautes fréquences dans des matrices de similarité locales
- 3- estimation de la similarité globale en analysant les matrices de similarité locales



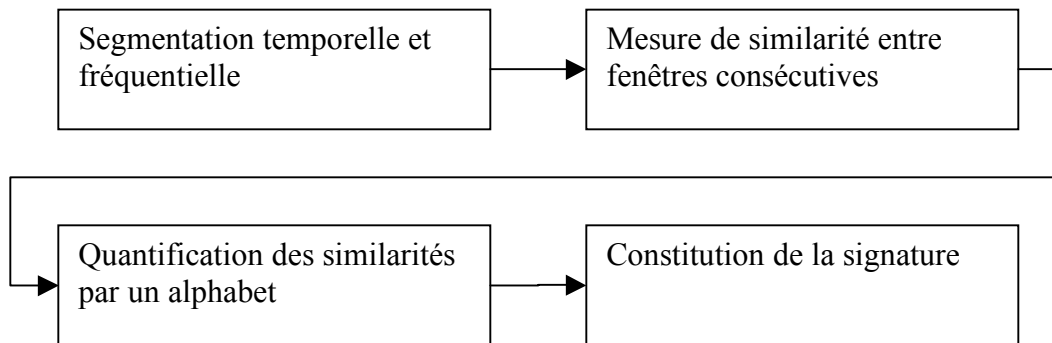
Nous travaillons actuellement sur deux directions. La première est la modélisation des phénomènes basiques de la perception auditive humaine, tel que les courbes isotoniques et les effets de masque, dans le cadre de la modélisation MGI. La deuxième est dans l'application de la distance de Kullback-Leibler sur les ratios de bandes de fréquence afin de prendre en compte une caractéristique du timbre.

2. Tache 6. Reconnaissance de titres musicaux

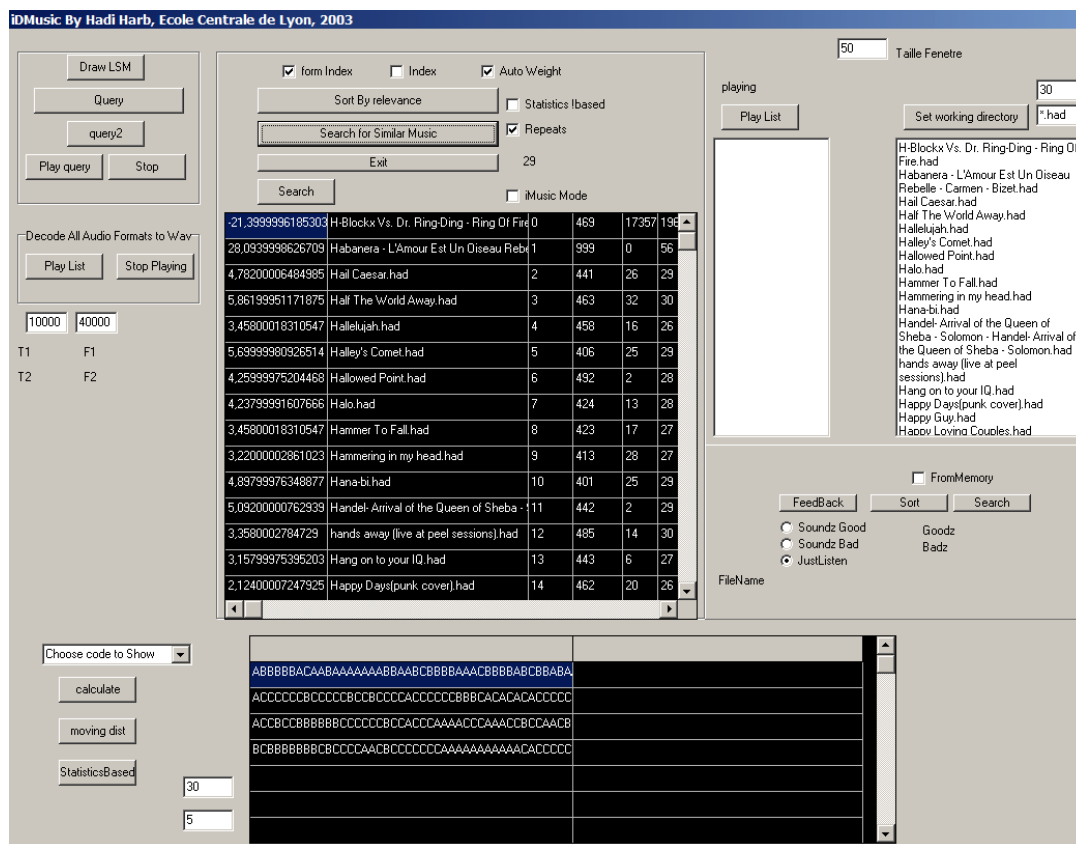
L'objectif de l'identification de musique dans le cadre du projet est d'intégrer un identifiant unique à chaque titre. Un identifiant obtenu à partir d'une signature du signal sonore aura l'avantage d'être unique et de ne pas tenir compte des erreurs possibles des métadonnées (nom du titre, artiste) ou des multiples versions du même titre. Dans un premier temps nous avons testé l'utilisation des probabilités d'appartenance à des classes de musique prédéfinies obtenues par un classificateur généraliste. Les premières expérimentations ont montré que ce type d'identifiant n'est pas suffisamment discriminant pour une base de donnée de plusieurs milliers de titres musicaux. Cet identifiant a cependant l'avantage d'être très compacte. En conséquence nous avons travaillé sur l'adaptation de l'ADN sonore que nous avons proposée ultérieurement à l'identification des titres dans le cadre du projet. L'ADN sonore, une chaîne de caractères, consiste en une modélisation de la structure temporelle et fréquentielle du signal sonore.

L'algorithme développé pour l'extraction de l'ADN sonore consiste en

- 1- une segmentation temporelle et fréquentielle du spectre du signal
- 2- estimation de similarité entre les fenêtres consécutives
- 3- codage des valeurs de similarité par des caractères
- 4- constitution de la signature à partir des caractères



Dans ce cadre nous avons développé une interface graphique servant d'un moteur d'extraction et de recherche de signature.



Un effort important a été par ailleurs consacré pour étudier la durée minimale du signal sur laquelle l'ADN sonore doit porter afin d'obtenir une discrimination suffisante dans une grande base de donnée. Les expérimentations d'identification que nous avons menées dans une base de 60 000 titres suggèrent que la durée de 5 secondes est suffisante.

Même si la taille actuelle de l'ADN sonore, qui constitue une clé pour les titres de musique, est réduite, il faut aboutir à une sorte de hachage qui permet une recherche

rapide. Nous étudierons ainsi la possibilité de modélisation de l'ADN sonore par des Modèles de Markov Cachés.

3. Tache 7. Classification des titres musicaux en genre et en humeur

Un premier travail dans le cadre de classification de la musique en genre a été concentré sur la définition des genres. Un ensemble de 6 genres basiques qui sont utilisés sur les plateformes de téléchargement en ligne a été défini. Les genres sont : Classique, Rock, Dance, Jazz, Metal, Rap. Dans cette définition nous avons évité les sous-genres dans la mesure où ces genres sont moins intuitifs pour les utilisateurs tout en étant différents d'une plateforme de téléchargement à une autre. Nous avons constitué une base de donnée de 2000 segments de 20 secondes de musique qui couvre les 6 genres définis. Cette base servira pour les évaluations des algorithmes de classification et d'estimation de similarité acoustique.

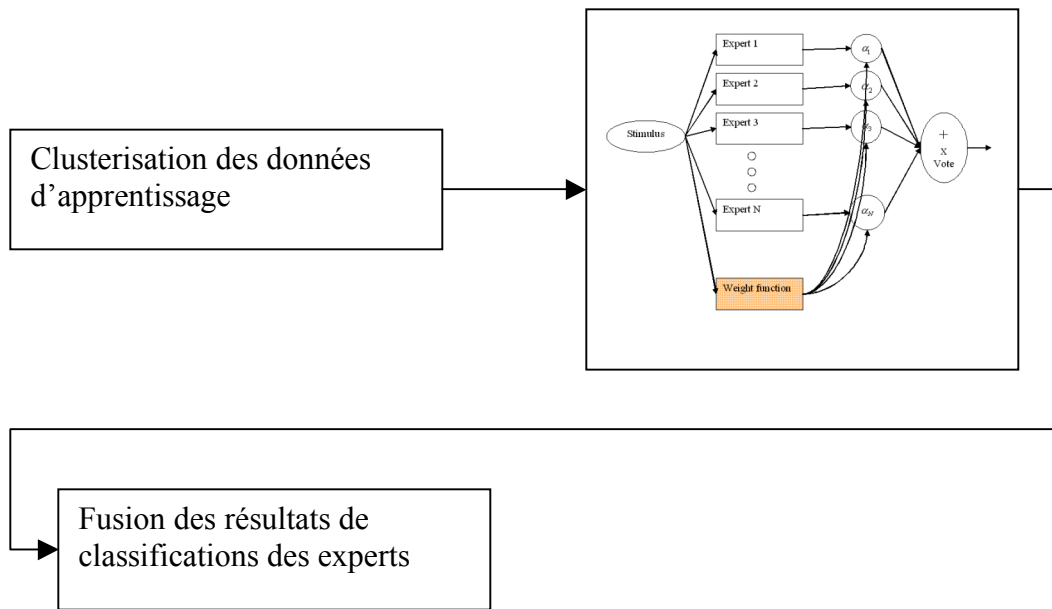
La base est constituée de

Rock : 476 fichiers
Rap : 497 fichiers
Classic : 109 fichiers
Dance : 319 fichiers
Metal : 431 fichiers
Jazz : 281 fichiers

Tous les fichiers sont d'une durée de 30 secondes en format mp3, 96 Kbps, mono, 44KHz. Les fichiers ont été capturés des webradios diffusants sur Internet (répertoriées Shoutcast). Les genres ont été renseignés à partir des genres des webradios et confirmés par la base [www.freedb.org].

La description est en XML, un fichier XML par genre et un fichier XML global. Les fichiers sont organisés par répertoires pour chaque genre. Le titre, l'artiste et le genre de chaque fichier sont renseignés. Nous envisageons le partage de cette base d'échantillons sonores entre les partenaires musicDiscover.

Nos premiers travaux portent sur l'utilisation des caractéristiques MGI généralistes et une approche multi-experts. Les premiers résultats sont encourageants notamment sur l'approche multi-experts. Nous étudierons l'effet de l'utilisation des modèles issus de la perception auditive humain dans le cadre des caractéristiques MGI.



D'un autre coté, nous avons développé une approche orientée vers le rythme de la musique pour la classification en genre. Nous avons proposé un beat histogram à 2 dimensions qui est une continuité du beat histogram classique connu dans la littérature. L'avantage du beat histogram à 2D réside dans son indépendance du choix des seuils pour la détection des beats. Nous avons ainsi proposé une mesure de similarité entre deux beat histograms 2D. Lorsque cette mesure de similarité a été appliquée au problème de classification en genre, une amélioration de performances par rapport au beat histogram classique a été observée.

La combinaison du classificateur en genre basé sur le beat histogram 2D et celui basé sur les MGI-multi-experts fera l'objet de nos travaux futurs.

