

Rapport d'avancement - MusicDiscover

version LTCI

Tache 1. Analyse rythmique et détection de ruptures

Plusieurs travaux ont été menés dans le domaine de l'analyse rythmique et la détection de ruptures. Les premiers travaux sur les décompositions harmoniques / bruit avaient déjà permis de montrer leur potentiel pour l'analyse rythmique [Alonso&al2003] (cette dernière étant effectuée sur la partie « bruit » qui conserve principalement les attaques des notes) mais comme on le verra, elles possèdent un intérêt particulier pour plusieurs aspects de ce projet (séparation de sources et reconnaissance des instruments de musique en particulier). Ainsi, les travaux concernant ces méthodes ont été poursuivis. On s'est notamment intéressé aux méthodes d'analyse spectrale dites à Haute Résolution (HR) qui s'affranchissent des limites de l'analyse de Fourier, en exploitant la structure particulière du modèle de signal sous-jacent qui représente le signal comme une somme de composantes sinusoïdales. Les travaux menés dans ce cadre ont notamment porté sur le développement de techniques de plus faible complexité, permettant l'estimation et la poursuite des paramètres tout en conservant la même résolution spectrale et sur la sélection automatique de l'ordre du modèle. Cependant, l'application de ces techniques à des signaux audio n'est pas immédiate, car le modèle de signal n'est généralement pas rigoureusement vérifié, et les efforts récents se sont donc portés sur la mise au point d'un certain nombre de pré-traitements destinés à améliorer la robustesse de l'analyse HR sur les signaux audio (on trouvera une synthèse de ces travaux dans [Badeau2005]).

Concernant l'analyse rythmique, deux aspects principaux ont été étudiés. Un premier aspect a visé l'amélioration des performances en basant la détection du rythme sur une fonction de détection des ruptures du signal. Ce premier travail a permis de proposer un nouvel estimateur du flux spectral et de calculer une fonction de détection de ruptures en exploitant des principes psycho-acoustiques et un filtrage non-linéaire de type filtrage médian [Alonso&al05]. Soulignons que cet algorithme a été soumis en juillet 2005 à évaluation dans le cadre de la seconde campagne d'évaluation internationale en Recherche d'Information Musicale (MIREX) qui aura lieu en septembre 2005.

Il est maintenant prévu de développer des approches hybrides (ou de fusion) qui permettraient de mieux exploiter les spécificités des techniques par flux spectral et des méthodes exploitant une décomposition harmonique/ bruit.

Un second aspect du travail a consisté à développer des méthodes de suivi du tempo au cours du temps et d'estimation du « tatum », qui peut-être considéré comme une subdivision du tempo principal. Une des approches qui a donné les résultats les plus prometteurs et qui permet notamment d'extraire un profil robuste de variation du tempo est basée sur l'utilisation de la programmation dynamique avec contraintes.

[Alonso&al03] M. Alonso, R. Badeau, B. David et G. Richard, Musical tempo estimation using noise subspace projections, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03), New Paltz, New York, 19-22 octobre 2003,

[Badeau, 2005] R. Badeau, « Méthodes à haute résolution pour l'estimation et le suivi de sinusoïdes modulées. Application aux signaux de musique », Thèse de doctorat de Télécom-Paris, soutenue le 18 avril 2005.

[Alonso&al05] Miguel Alonso, Gaël Richard and Bertrand David, "Extracting Note Onsets from Musical Recordings", *International Conference on Multimedia and Expo (ICME'05)*, Amsterdam, The Netherlands, July 2005.

Tache 2. Reconnaissance des instruments de musique et indexation

Une première étude a concerné plus particulièrement la reconnaissance (et la transcription) des instruments percussifs et notamment les signaux de batterie. L'étude de cette classe d'instruments permet par ailleurs de faire le lien entre les études d'analyse rythmique et les études de séparation de sources (extraction des signaux percussifs par séparation, v. tache 3). Initié en 2004, le travail sur la recherche de boucles de batterie par le contenu a été poursuivi et a permis d'obtenir un système performant de transcription automatique de boucles de batteries comportant divers post-traitements (réverbération, compression, utilisation de kits synthétiques,...). Le système complet développé permet également de formuler une requête à la voix par fredonnement d'onomatopées. La requête est alors transcrite à l'aide d'un algorithme classique de reconnaissance vocale par modèles de Markov cachés et est alignée avec les transcriptions des boucles de batterie stockées en mémoire. Les résultats d'une requête est représentée graphiquement et autorise une navigation par similarité dans la base de boucles de batterie. Ce travail a fait l'objet de deux publications récentes dont l'une a reçu le « *Best Paper Award* » [Gillet&al2005a, Gillet&al2005b].

Un autre travail intéressant a été mené en relation avec la tâche 1 et consiste à exploiter la détection de ruptures pour la reconnaissance des instruments. En effet, il est largement accepté que l'attaque d'un son est particulièrement importante pour la perception du timbre et donc pour la reconnaissance des instruments de musique. Hors, les approches classiques en reconnaissance des instruments de musique diluent souvent l'information portée par l'attaque avec l'information portée par la partie tenue du son (on pense ici notamment à toutes les approches statistiques qui exploitent des vecteurs de paramètres calculés sur des durées prédéterminées). Il apparaît alors pertinent de développer des classifieurs différents pour les parties d'attaque et pour les parties tenues. Un premier travail dans cette direction a déjà permis de montrer la pertinence de l'approche mais a également mis en lumière l'importance d'un meilleur couplage entre les deux classifieurs [Essid&al.2005].

Il est aussi important de souligner les premiers efforts en direction d'une reconnaissance d'instruments en contexte polyphonique (i.e. la reconnaissance d'ensemble d'instruments ou la reconnaissance de signaux de batterie en présence d'instruments harmoniques) et qui ont donné lieu à plusieurs soumission d'articles (ISMIR 2005 et IEEE-Waspaa2005).

[Essid&al.2005] S. Essid, P. Leveau, G. Richard, L. Daudet and B. David, "On the usefulness of differentiated transient/steady-state processing in machine recognition of musical instruments", *International Convention of the Audio Engineering Society (AES)*, Barcelona, Spain, May 2005.

[Gillet&al2005a] O. Gillet et G. Richard , “Drum loops retrieval from spoken queries”, *Journal of Intelligent Information Systems - Special issue on Intelligent Multimedia Applications*, vol. 24, n° 2/3, pp. 159-177, March 2005

[Gillet&al2005b] O. Gillet and G. Richard, “Indexing and querying drum loops databases”, *International workshop on Content Based on Multimedia and Indexing (CBMI’05)*, Riga, Latvia, June 2005. **Received the CBMI BEST PAPER Award**

Tache 3. Séparation de sources

Un premier aspect étudié dans cette tâche concerne la séparation de sources pour l’extraction de la piste de batterie. Ce problème apparaît en effet plus simple que la séparation d’instruments harmonique et présente plusieurs applications de grand intérêt pour l’indexation automatique ou la manipulation sonore (i.e. remixing). Le principe développé dans le cadre du projet exploite une analyse Haute Résolution (HR) pour obtenir une séparation du signal audio en deux composantes : une composante quasi-harmonique qui regroupe principalement les instruments harmoniques et une composante bruit qui regroupe une partie du signal de batterie. Le signal de batterie est ensuite reconstitué en associant au mieux une décomposition en banc de filtres et l’analyse haute-résolution. Une première étude limitée à la transcription des deux éléments principaux de la batterie a également montrée l’intérêt d’un tel traitement pour la transcription. Ce premier travail a donné lieu à la soumission de deux articles pour la prochaine conférence ISMIR2005 et le prestigieux workshop IEEE-Waspaa ([Gillet&al2005c], [Gillet&al2005d])

Notons que ce travail est le résultat d’une première exploitation de la base de données de signaux de batterie enregistrée au LTCI au début février 2005. Cette base est constituée d’enregistrements de trois musiciens batteurs professionnels chacun ayant joué sur sa propre batterie afin d’obtenir des signaux les plus variés possibles. Un quatrième batteur a été également partiellement enregistré afin de régler l’ensemble du dispositif d’enregistrement. Chaque batteur a enregistré des frappes isolées, des solos et des accompagnements de morceaux de type « minus one », où la piste batterie est absente. Ces derniers enregistrements permettent en particulier d’effectuer un mixage avec les enregistrements originaux des autres instrumentistes tout en possédant un enregistrement de la piste batterie seule. L’objectif de cette base de données est ainsi de pouvoir mener aussi bien des études de reconnaissance automatique des éléments de la batterie, que des études de séparation de la piste de batterie à partir d’un signal audio mixé. Cependant, une exploitation plus poussée de cette base demande à ce jour un important travail d’annotation.

[Gillet&al2005c] O. Gillet and G. Richard, “Extraction and remixing of Drum Tracks from Polyphonic Music Signals”, *soumis à IEEE-Waspa 2005, New Palts, USA*.

[Gillet&al2005c] O. Gillet and G. Richard, “Drum Track Transcription of Polyphonic Music Using Noise Subspace Projection”, *soumis à ISMIR 2005*