

# Rapport final - LTCI

## Musicdiscover

### 1. Liste des équipes impliquées

- LTCI : Laboratoire Traitement et Communication de l'Information /UMR 5141

### 2. Liste des participants

Pour les doctorants, préciser la date de début de thèse et le type de financement, pour les post-doctorants, préciser la date de début de travail dans le projet et le type de financement. Les statuts doivent être précis par exemple Maître de Conférence

- **LTCI : groupe Audio Acoustique et Ondes (AAO)**
  - Olivier Gillet, doctorant (date de début de thèse: décembre 2003, financement BDI CNRS), 30% ; fin de thèse juillet/2007
  - Valentin Emiya, doctorant, (date de début de thèse: octobre 2004, financement MNRT), 30%
  - Miguel Alonso, doctorant, (date de début de thèse: octobre 2002, financement SFERE), puis post-doctorant 30% jusqu'en avril 2007
  - Slim Essid, doctorant (date de début de thèse: octobre 2002, financement MNRT puis sur contrat Musicdiscover), 50%, puis Postdoctorant (financement par projet MusicDiscover jusqu'en mars 2006), puis Ingénieur d'étude 10% à partir d'octobre 2006
  - Bertrand David, Maître de conférences, 15%
  - Roland Badeau, Maître de conférences, 15%
  - Gaël Richard, Professeur, 20%

### 3. Changements significatifs intervenus dans le projet

L'équipe du LTCI impliquée dans le projet a été très stable.

### 4. Résumé des principales avancées

**L'Analyse Rythmique et détection de ruptures :** Plusieurs travaux ont été menés dans le domaine de l'analyse rythmique et la détection de ruptures. On s'est tout d'abord 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. Ces méthodes possèdent un intérêt particulier pour plusieurs aspects de ce projet (extraction rythmique, séparation de sources et reconnaissance des instruments de musique en particulier). Les travaux menés dans ce cadre ont notamment porté sur le développement de techniques adaptatives de plus faible complexité.

Concernant l'analyse rythmique, un premier aspect a visé l'amélioration des performances en basant la détection du rythme sur une fonction de détection des ruptures du signal. 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. Finalement, plusieurs approches de fusion de décisions ont été utilisées pour optimiser la robustesse et les performances des détecteurs de périodicités des analyseurs rythmiques. Pour une plus grande efficacité, certains modules parmi les plus complexes ont été optimisés et implémentés en langage C/C++ et placés sur le réflecteur interne du projet (détecteur de transitoires).

Ce travail sur l'analyse rythmique et la détection de ruptures a donné lieu à des évaluations internationales. En effet, le LTCI a participé à deux reprises (Juillet 2005 et Juillet 2007) aux campagnes d'évaluation internationales en Recherche d'Information Musicale (MIREX). Nos algorithmes se sont classés premier en 2005 et troisième en 2007.

Par ailleurs, les résultats de ce travail ont été largement diffusés (1 article de conférences et 3 articles de revue) et ont constitué l'essentiel des résultats obtenus par M. Alonso pour sa thèse de doctorat (soutenue le 13 novembre 2006) et son Post-Doc qui a suivi.

**La 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. 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 travail sur la transcription proprement dite de signaux percussifs a été poursuivi tout au long du projet et a notamment été appliquée au cas plus complexe des signaux polyphoniques. Pour cela, une importante base de données a été enregistrée (la base ENST-Drums) dont une large part est rendue publique pour permettre une meilleure évaluation comparative des algorithmes proposés dans la littérature. A ce jour, la base ENST-Drums est déjà utilisée par 10 universités ou laboratoires de recherche dont 2 nord-américaines. Un système complet de transcription de la piste de batterie de signaux polyphoniques a ainsi été construit et intègre plusieurs innovations majeures. Ce travail a constitué une large part de la thèse de doctorat d'Olivier Gillet et a donné lieu à de nombreuses publications.

Une autre direction de travail consiste à exploiter la détection de ruptures pour la reconnaissance des instruments. Un premier travail dans cette direction a déjà permis de montrer la pertinence d'une approche exploitant deux classifieurs (un pour la partie tenue des sons et un autre pour les attaques) mais a également mis en lumière l'importance d'un meilleur couplage entre les deux classifieurs. 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 publications.

Enfin, sur un axe indexation une nouvelle méthode pour l'analyse de fréquences fondamentales multiples a été développée. Cette méthode est basée sur le principe du maximum de vraisemblance. Pour chaque composante, un modèle sinusoïdal est utilisé couple à une modélisation à moyenne ajustée (MA) du bruit et un modèle autorégressive (AR) de l'enveloppe des partiels. Un estimateur de la fréquence fondamentale, dans le cas d'une fréquence fondamentale unique, a été obtenu par l'application d'un principe de maximum de vraisemblance pondéré. L'estimateur de fréquences fondamentales multiples est une

extension directe de cette méthode. Un algorithme spécifique au piano a ensuite été dérivé (e.g. qui prend en compte l'inharmonicité des partiels pour cet instrument).

**La 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 de séparation est traité d'une part en proposant une nouvelle approche basée sur une décomposition harmonique/bruit et un masquage spectro/temporel et en améliorant une méthode existante de séparation qui exploite le filtrage de Wiener. L'efficacité de la séparation est démontrée à l'aide de critères d'évaluation objectifs couramment utilisés dans ce domaine. De nombreux exemples de séparation sont par ailleurs proposés pour une évaluation subjective des différentes méthodes de séparation développées (<http://www.enst.fr/~gillet/ENST-drums/separation/>).

Des travaux ont été initiés en cours de projet pour améliorer la transcription des signaux polyphoniques. Cette nouvelle approche qui exploite une « factorisation en matrices positives » du signal est à la frontière du domaine de la séparation de sources mais obtient des résultats particulièrement prometteurs pour la transcription. Un enrichissement des premières approches en lui imposant des contraintes d'harmonicité des sources décomposées ont permis d'encore améliorer les résultats et ont permis d'obtenir de très bons résultats à la récente campagne d'évaluation internationale (MIREX 2007).

**La Reconnaissance du genre musical :** Une collaboration active est conduite entre le LTCI et le LIRIS sur une tâche de classification de titres musicaux en genre. L'objectif est ici de comparer les systèmes de classification des deux partenaires pour aller vers une solution qui fusionne les deux systèmes en un nouveau qui permettent des performances supérieures. Le LIRIS a fourni sa base de données sonores au LTCI ainsi que ses fichiers de *scoring* au format *xml*. Les performances moyennes des deux systèmes sont comparables. Des performances supérieures à chacun des systèmes sont obtenues grâce à une stratégie de fusion adéquate qui tirent le meilleur profit des deux systèmes.

## 5. Réalisations obtenues dans le cadre du projet

### Contributions du LTCI

- Développement de plusieurs algorithmes d'estimation rythmique
- Développement d'algorithmes de reconnaissance des instruments de musique en contexte mono-phonique et polyphonique.
- Développement d'algorithmes d'extraction et de transcription de la piste de batterie à partir d'un signal musical polyphonique
- Développement d'algorithmes d'estimation de fréquences fondamentales multiples et de transcription par factorisation en matrices Non-négatives.
- Spécification, enregistrement et annotation d'une base de données complète de batterie ENST-Drums déjà diffusée à 10 universités dans le monde
- Etudes de reconnaissance du genre musical

## 6. Réunions et Conférences organisées dans le cadre du projet

Lister aussi les réunions propres au projet

## **7. Soutiens obtenus en liaison avec ce projet**

### **7.1. Postes chercheurs**

Délégations, détachements, doctorants, post-doctorants (donner l'année de début, l'année de fin et le type du financement)

### **7.2. Postes ingénieurs**

Préciser CDD ou CDI, donner l'année de début, l'année de fin (si nécessaire) et le type du financement

### **7.3. Contrats nationaux**

Le projet Quaero actuellement en cours de discussion sera un prolongement naturel de Musicdiscover pour deux des trois partenaires.

### **7.4. Contrats européens**

Le réseau d'excellence Kspace qui a débuté en janvier 2006, quoique centré sur l'analyse sémantique de documents multimedia reprend certains aspects du projet Musicdiscover.

### **7.5. Contrats internationaux hors CEE**

année début, année fin, type de financement, montant

### **7.6. Contrats industriels**

année début, année fin, type de financement, montant

## ***Contacts internationaux dans le cadre de ce projet***

- Nom Prénom, Département ou Laboratoire, Université, Pays
- .....

## **8. Publications obtenues dans le cadre du projet**

- Thèses

M. Alonso, « Extraction d'Information Rythmique à Partir d'Enregistrements Musicaux », Thèse de Doctorat de l'ENST, nov. 2006.

S. Essid, « Classification automatique des signaux audio-fréquences: reconnaissance des instruments de musique, thèse de doctorat de l'Université Pierre et Marie Curie, Dec. 2005.

O. Gillet, « Transcription des signaux percussifs. Application à l'analyse de scènes musicales audiovisuelles », thèse de doctorat de l'ENST, Juin 2007.

- Journaux internationaux

R. Badeau, G. Richard, B. David, "Fast and stable YAST algorithm for principal and minor subspace tracking," submitted to IEEE Transactions on Signal Processing.

M. Alonso, G. Richard et B. David, "Accurate tempo estimation based on harmonic+noise decomposition", EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 82795, 14 pages, 2007.

M. Alonso, G. Richard et B. David, "Tempo Estimation for Audio Recordings", accepted for publication in the Journal of New Music Research.

S. Essid, G. Richard and B. David, Instrument Recognition in Polyphonic Music Based on Automatic Taxonomies, IEEE Transactions on Speech, Audio and Language Processing, Volume 14, Issue 1, Jan. 2006 Page(s):68 – 80

S. Essid, G. Richard and B. David, “Musical Instrument Recognition by pairwise classification strategies”, IEEE Transactions on Speech, Audio and Language Processing, Volume 14, Issue 4, July 2006 Page(s):1401 - 1412.

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

O. Gillet, G. Richard, “Transcription and Separation of Drum Signals from Polyphonic Music”, accepté dans les IEEE Trans. On Audio, Speech and Language Processing (publication début 2008).

- Conférences internationales

M. Alonso, G. Richard et B. David, “Extracting Note Onsets from Musical Recordings”, International Conference on Multimedia and Expo (IEEE-ICME'05), Amsterdam, The Netherlands, July 2005.

M. Alonso, G. Richard et B. David, “Accurate tempo estimation based on harmonic+noise decomposition”, EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 82795, 14 pages, 2007.

R. Badeau, B. David et G. Richard, “YAST Algorithm for Minor Subspace Tracking”, International Conference on Acoustics, Speech, and Signal Processing ICASSP'06, Toulouse, France, 15-19 mai 2006, vol. III, pp. 552-555

B. David , R. Badeau et G. Richard, “HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals”, International Conference on Acoustics, Speech, and Signal Processing ICASSP'06, Toulouse, France, 15-19 mai 2006, vol. III, pp. 45-48

R. Badeau, B. David et G. Richard, “Conjugate gradient algorithms for minor subspace analysis”, International Conference on Acoustics, Speech, and Signal Processing ICASSP'07, Honolulu, Hawaii, USA, 15-20 avril 2007, vol. III, pp. 1013-1016

O. Gillet et 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.

P. Leveau, S. Essid, G. Richard, L. Daudet, 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.

S. Essid, G. Richard, B. David, "Instrument recognition in polyphonic music," International Conference on Acoustics, Speech, and Signal Processing ICASSP'05, Philadelphia, USA, March 2005.

O. Gillet, G. Richard, "Drum Track Transcription of Polyphonic Music Using Noise Subspace Projection", International Conference on Music Information Retrieval (ISMIR), London, Great-Britain, Sept. 2005.

S. Essid, G. Richard, B. David, "Inferring Efficient Hierarchical Taxonomies for MIR Tasks: Application to Musical Instruments", International Conference on Music Information Retrieval (ISMIR), London, Great-Britain, Sept. 2005.

O. Gillet, G. Richard, "ENST-Drums: an extensive audio-visual database for drum signals processing". In Proc of 7th International Conference on Music Information Retrieval, ISMIR 2006, Oct. 2006, Victoria, Canada.

V. Emiya, R. Badeau et B. David, "Multipitch estimation of inharmonic sounds in colored noise", 10th International Conference on Digital Audio Effects DAFx-07, Bordeaux, France, 10-15 septembre 2007, pp. 93-98

O. Gillet, G. Richard, "Extraction and Remixing of Drum Tracks from Polyphonic Music Signals". IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA'05, New Paltz, USA.

N. Bertin, R. Badeau et Gaël Richard, "Blind signal decompositions for automatic transcription of polyphonic music: NMF and K-SVD on the benchmark, International Conference on Acoustics, Speech, and Signal Processing ICASSP'07, Honolulu, Hawaii, USA, 15-20 avril 2007, vol. I, pp. 65-68

- Journaux nationaux
- Conférences nationales