# Musicdiscover

## Rapport final, 23 Novembre 2007

## Xavier Rodet, Ircam

## 1. Liste des équipes impliquées

- IRCAM, Institut de recherche et de Communication Acoustique/Musique, UMR9912
- LTCI : Laboratoire Traitement et Communication de l'Information, UMR 5141
- LIRIS, UMR 5205, Ecole Centrale de Lyon

## *2.* Liste des participants

- **IRCAM, Equipe Analyse-Synthèse des Sons**
  - Xavier Rodet, Professeur, Responsable de l'équipe Analyse-Synthèse, 20%
  - David Fenech, chercheur-développeur, 100%, financement MusicDiscover, de Février à Octobre 2007
  - Damien Tardieu, Doctorant depuis Janvier 2005, financement IRCAM et MusicDiscover, 20%
  - Chunghsin Yeh, Doctorant depuis Octobre 2003, financement MRT et MusicDiscover, 100%
  - Arié Livshin, Doctorant depuis Octobre 2002, financement EGIDE et MusicDiscover, 100%
  - Thomas Huebert, stagiaire, financement MusicDiscover
  - Philippe Bernat, stagiaire, financement MusicDiscover
  - Hagen Kaprykkowsky, stagiaire, financement MusicDiscover
  - Aurélien Marty, stagiaire, financement MusicDiscover
  - Emmanuel Vincent , Doctorant terminé en Décembre 2004, financement MRT et MusicDiscover
  - Bertrand Delezoide, Doctorant terminé en Avril 2006, financement CEA.
  - Geoffroy Peeters et Jérome Barthélémy : Consultants (Ircam)

- **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'etude 10% à partir d'octobre 2006
- Bertrand David, Maître de conférences, 15%
- Roland Badeau, Maître de conférences, 15%
- Gaël Richard, Professeur, 20%

- **LIRIS ECL**
  - Chen Liming, Pr, 20%
  - Dellandréa, Emmanuel, MCF, 30%
  - Aliaksandr Paradzinets, Doctorant, 10/2004, bourse EGIDE ACI
  - Zhongzhe Xiao, Doctorant, 10/2004, bourse EGIDE ACI
  - Hadi Harb, PostDoc, 10/2004 à 9/2005, ACI

## 3. Changements significatifs intervenus dans le projet

IRCAM :
Fin et soutenance de la thèse de E. Vincent, IRCAM, Décembre 2004.
Fin et soutenance de la thèse de B. Delezoide, IRCAM, Avril 2006.
Début de la thèse de D. Tardieu, Janvier 2005.
Pour des questions pratiques et de cohérence, le commencement de la tâche 4 a été déplacé à début 2007.
Engagement de David Fenech, chercheur développeur de Février à Octobre 2007 pour la tâche 4

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

LIRIS
Le contrat de post-doc d'Hadi Harb, LIRIS, a pris fin le 31 mars 2006.

## *4.* Résumé des principales avancées

**Tache 1. 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 portés 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.

**Tache 2. La Reconnaissance des instruments de musique et indexation** :

Le projet LIBSDB (rapport interne Ircam [Jacob05] Jacob, M., « Perspectives in the development of audio databases at SEL ») développe un système de base de données spécialisé dans le stockage et la gestion de descripteurs sonores, ainsi que des interfaces avec les environnements de développement utilisés à l'Ircam par les équipes scientifique. Le système supporte l'accès à des bases de données distribuées (locales ou distantes), et offre à l'utilisateur des paradigmes communs quelque soit l'interface utilisée (C, Matlab, Max/MSP, OpenMusic, Python…). Le développement d'une première version de cette librairie (code source C) et de son interface, a été finalisé en Décembre 2005, et a été mise à disposition des utilisateurs à l'Ircam. Cette librairie est utilisée par des tâches de MusicDiscover.

La reconnaissance des instruments entretenus a commencé par la mise en place de 5 bases de données de notes isolées, dont Studio En Ligne de l'Ircam, et la comparaison de différents algorithmes de classification. L'algorithme de classification utilisant un modèle

gaussien hiérarchique a été étendu au mélange de gaussiennes hiérarchique (gaussian mixture). L'extraction automatique des descripteurs [Peeters 04a] a été intégrée dans la Sound Palette On Line de l'Ircam. Un des principaux résultats est que les bases généralement ne couvrent qu'une faible portion des notes possibles d'un instrument, il faut donc compiler de nombreuses bases d'origines variées. Un des objectifs est de permettre la recherche et la manipulation d'un signal sonore par descripteurs de haut niveau et perceptifs [Tardieu04] et, par exemple, pour une aide à l'orchestration. Le système conçu dans la thèse de D. Tardieu à l'Ircam [Tardieu05] permet de spécifier un son cible et un ensemble d'instruments disponibles et détermine les combinaisons de sons instrumentaux imitant au mieux la cible. La connaissance des différents instruments est obtenue par analyse des bases qui contiennent une grande variété de modes de jeu et de hauteur. Des descripteurs extraits du signal et présentant des propriétés d'additivité mesurent la similarité perceptive entre deux sons. La recherche des meilleures combinaisons est accélérée par des restrictions de l'espace de recherche. La prise en compte de l'évolution du son au cours du temps est en cours d'étude.

Un autre objectif étudié dans la thèse de B. Delezoide à l'Ircam [Delezoide06c], en collaboration avec le CEA, est de permettre l'indexation de documents multimédia, segmentation temporelle de la structure et classification du contenu. Notre attention a porté sur l'étude de concepts dit « granulaires » de l'audio (musique, parole) et de l'image [Delezoide05a, 05b]. Nous avons montré que la fusion par réseau Bayésien et SVM de l'information fournie par ces concepts et plusieurs descripteurs numériques bas niveau améliore les résultats de segmentation et de classification multimédia des lieux. L'étude de modèles de fusion des concepts et des descripteurs visuels et auditifs nous a permis d'utiliser les relations de corrélation qu'ils entretiennent et le modèle de fusion le plus approprié [Delezoide06a, 06b].

Une base de données commune de *solos* de nombreux instruments a été créée en coopération entre l'Ircam et le LTCI, pour l'apprentissage, la reconnaissance et l'évaluation. Afin d'éviter des problèmes de copyright, chaque solo est une extrait de 30 secondes. La reconnaissance a été développée pour les solos [Livshin04a,b]. Il est possible de reconnaître des instruments solos en temps réel avec un taux suffisant pour de nombreuses applications. La reconnaissance dans les mélanges polyphoniques à l'Ircam est menée en collaboration avec la détection de F0s multiples (Cf. Tâche 3) qui permet de séparer les instruments harmoniques avant la reconnaissance. Une grande base de données des sons harmoniques isolés a été rassemblée à partir de 13 bases de données différentes. On a montré que le taux de reconnaissance des instruments avec uniquement les partiels harmoniques est en moyenne seulement 4% inférieur à celui des sons entiers [Livshin06a]. Il apparaît aussi que le résiduel non harmonique (" bruit ») des sons isolés [Livshin06b] permet la reconnaissance des instruments mais avec un taux plus faible que les partiels harmoniques. Ce travail a donné lieu à plusieurs publications et a constitué une large part de la thèse de doctorat de A. Livshin qui sera soutenue le 12 Décembre 2007 . Une autre approche de reconnaissance des instruments dans des mélanges polyphoniques, en parallèle avec la séparation de sources, a été testée avec succès dans la thèse de E. Vincent [Vincent04] (Cf. Tâche 3).

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.

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 utilise 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).

Enfin des algorithmes de détection des classes parole/musique/mélange ont été développés dans la thèse de B. Delezoide [Delezoide06c] et évalués par G. Peeters.

**Tache 3. La Séparation de sources :**

L'estimation de fréquences fondamentales multiples, dites F0s, est reliée aux tâches Reconnaissance des instruments et Séparation de sources [C. Yeh04a, 04b]. L'algorithme développé est fondé sur une estimation adaptive du niveau bruit [Yeh06b], et une analyse des

partiels dans le domaine fréquentiel et leur regroupement par F0 ce qui permet la séparation des sources [C. Yeh05, 06a, 06b]. Une base de donnée construite pour évaluer l'estimation de F0s est publiée sur le site MusicDiscover avec des fichiers MIDI de la base de données *RWC Classical* alignés aux enregistrements. Les résultats de cet algorithme se comparent plutôt favorablement avec ceux de A. Klapuri. Cet algorithme est utilisé pour la séparation de sources et pour la reconnaissance d'instruments.

La thèse de E. Vincent, [Vincent04] à l'Ircam propose une méthode de séparation de sources fondée sur des modèles statistiques des instruments et l'Analyse en Sous-Espaces Indépendants, et formulée comme un problème d'estimation bayésienne. Les critères bayésiens exacts ont été approchés pour mettre en place des algorithmes d'estimation rapides [Vincent05a, 05b]. Une approche de la séparation des instruments harmoniques est mené en collaboration avec la détection de F0s multiples et la reconnaissance des instruments (Cf. ces paragraphes). Les travaux en 2007 ont continué sur l'estimation du nombre de F0s, et le suivi des trajectoires des F0s [Yeh06a]. Quatre ensembles d'échantillons polyphoniques ont été préparés pour évaluer notre système, en mixant des échantillons monophoniques de différentes bases de données, RWC, Université McGill, Université Iowa et IRCAM. Nous avons créé une base de données de musique synthétique en utilisant un "sampler" et des échantillons monophoniques des instruments de musique [Yeh07]. En 2007, les algorithmes actuels d'estimation de F0s ont été évalués et comparés pour la première fois grâce à MIREX (Music Information Retrieval Evaluation eXchange) 2007. Nous avons participé dans la première tâche: l'estimation de F0s trame par trame. Notre système a été classé 2$^{ème}$, très près du meilleur.

Un autre 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 examples 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).

**Tache 4. Description sémantique structurée :**

Dans cette tâche, le but est de construire une structure plus élaborée en utilisant conjointement les diverses informations extraites du signal, rythme, fondamentaux, instrumentation, timbre, genre et style, harmonie, etc.. La structure temporelle d'un enregistrement de musique est fournie par les algorithmes de G. Peeters (Ircam). La structure apparaît alors comme un graphe décrivant des séquences d'événements à divers horizons temporels et des relations entre états au cours du temps et à différents niveaux. Pour intégrer ensuite toutes les informations extraites dans les autres tâches de MusicDiscover, un interface de navigation a été développé dans MusicDiscover. Il permet de synthétiser les différentes extractions faites sur les musiques, de les représenter graphiquement, et de naviguer interactivement afin d'explorer en direct ces différents paramètres.

L'interface de navigation MusicDiscover.swf a été développée en Flash9 / Actionscript 3 sous l'environnement de programmation FlashDevelop. Etant écrite en flash, on peut donc l'utiliser sous diverses plateformes (mac, pc, linux, etc) sous la forme d'un exécutable autonome ou bien depuis un navigateur internet de type firefox. Des boutons cliquables permettent d'afficher diverses représentations (voir ci-dessous), la timeline globale et zoomable en bas, et la section de recherche dans une base de données à droite.

L'ensemble des données extraites par divers outils est interfacé dans le navigateur flash en utilisant le format standard XML. Ce format a l'avantage de pouvoir être lu et parsé facilement par flash, et est en quelque sorte un format pivot entre différents outils. Le fichier _top.xml est associé à chaque titre musical et contient des références vers les différents sous fichiers XML. Le fichier _struct.xml décrit le morceau en structure musicale (différentes sections, correspondant à des zones de musique similaire). Les différentes descriptions permettent de choisir un niveau de finesse dans la décomposition du morceau en sous structures pour ajuster cette représentation. Ce fichier est lu par l'interface graphique en flash. Alors, chaque objet de structure est cliquable afin de permettre l'écoute de la section correspondante.

Le fichier _chords.xml contient les informations extraites de suite d'accords. Les accords sont représentés sous la forme d'un texte affiché dans une zone avec un début et une fin. On peut également cliquer sur chaque élément graphique pour entendre la section correspondant à chaque accord.

Le fichier _beat.xml représente le rythme sous la forme de temps. Le premier temps est distingué, par exemple le premier temps en gras et les 3 temps suivants en trait fin.

Le fichier _notes.xml contient l'extraction des informations de notes, via un algorithme de détection de FO multiples. La représentation graphique correspondante est sous forme classique de piano roll. Chaque note affichée est cliquable est permet d'écouter une note de piano de référence pour comparer l'extraction avec le moreau joué, en direct. Les informations de notes se superposent aux autres paramètres déjà évoqués ci-dessus (struct, beat).

Le fichier _drums.xml contient l'extraction sous forme d'éléments de batterie (BD : bass drum, LT : Low Tom, SD :Snare Drum, HH : High Hat). La représentation graphique correspondante reprend le paradigme visuel d'une programmation de patterns de boite à rythmes. Le nom de l'élément représenté est affiché lorsque l'on pointe sur l'objet à la souris

(ici : HH pour Hi Hat). La position des éléments suit un ordre logique de bas en haut (BD,LT,SD,HH) allant du grave vers l'aigu.
Le fichier _instr.xml contient les extraction automatiques d'instruments de musique détectés. Son style est comparable au fichier _drums.xml, sauf que les évènements ont aussi un temps de fin. Les instruments sont représentés sous la forme de mnémoniques comme po (piano), vc (violoncelle), etc,etc. Les blocs graphiques sont des objets cliquables (le nom de l'instrument apparaît en survolant l'objet avec la souris). Chaque ligne correspond à un instrument de musique détecté.
La matrice de similarité peut être fournie à l'interface sous deux formats : un fichier JPG ou bien un binaire SDIF matriciel. Elle corrèle bien avec les informations de structure affichées pour ce morceau.
Les morceaux de musique lisibles dans cette interface doivent être des fichiers encodés en mp3 , avec un *Bitrate* Variable ou non. La seule contrainte importante est l'obligation de les tagger avec des tags ID3 v2. Après chargement, les différentes informations contenues dans les XMLs présents sont représentées.

Les Bases de donnée de titres sont accessibles par l'interface qui offre une recherche par des critères comme auteur, titre, album, genre, humeur et retourne une liste de résultats. L'interface offre aussi une recherche par similarité **:** lorsqu'un titre est en train de jouer, cliquer sur *similar* permet de trouver des titres similaires sur la base de la tonalité, du timbre, du tempo et du rythme. Le résultat est une liste de fichier avec leur distance respective envers le fichier couramment sélectionné.

**Tache 5. Recherche de la musique par similarité :**

La recherche de la musique par la similarité est une fonctionnalité majeure dans les applications professionnelles ou pour amateurs. On peut être amené à rechercher les titres d'un même chanteur, les titres du même genre musical, d'un genre proche, de même type rythmique ou, plus délicat, la même pièce musicale interprétée par un orchestre ou un chanteur différent. Mais la difficulté ici est que la similarité perceptive musicale est sémantique et comporte une bonne part de subjectivité et qu'en conséquence elle est souvent différente des mesures acoustiques, objectives, que l'on peut définir sur le signal audio musical.

Dans ce projet nous avons étudié la similarité musicale en analysant le signal audio. La similarité musicale est la mesure qui permet à un sujet humain de qualifier de similaire ou non-similaire deux segments musicaux. Idéalement, une mesure de similarité musicale devrait être capable de qualifier comme très similaire l'ensemble des interprétations par différents auteurs de différents genres musicaux d'une même chanson.

La plupart des techniques existantes sont basée sur la modélisation des coefficients cepstrales. Les caractéristiques les plus connues de la famille des caractéristiques cepstrales

sont les Mel Frequency Cepstral Coefficients (MFCC) qui sont des coefficients de cepstre obtenus des spectres de fréquence filtrés par une banque de filtres conformément à l'échelle de MEL. L'échelle de MEL est une échelle reflétant la perception humaine avec des différentes précisions fréquentielles suivant les fréquences. Pour simplifier, l'échelle de MEL affirme que les hautes fréquences sont aperçues avec moins de précision que les basses fréquences. Cependant, si ce modèle est convenable pour l'analyse de la parole et de ce fait il est à la base de systèmes de reconnaissance automatique de la parole, il n'est pas nécessairement approprié pour l'analyse d'un signal sonore en général, et d'un signal musical en particulier.

La ligne conductrice de nos travaux est de compléter les techniques classiques basées sur MFCC avec des mesures de similarité sur des caractéristiques musicales. Nos principales contributions aux différentes sous-tâches dans le cadre du projet MusicDiscover sont les suivantes :

- ✓ Elaboration des algorithmes d'analyse sonore adaptés à la musique.

  Le signal musical est un signal portant des caractéristiques spécifiques. Il est quasi-stationnaire, les fréquences ont une échelle logarithmique qui rendent l'utilisation des techniques standard comme MFCC moins appropriées à l'analyse du signal musical. Une transformation inspirée par la transformation en ondelette permettant d'obtenir les échelles tempo-spectrales adaptées à la musique a été proposée. Cette transformation à l'échelle variable (Variable Resultion Tranform – VRT) est une transformation qui hérite des caractéristiques de transformation en ondelettes classique en améliorant la résolution fréquentielle dans la zone des hautes fréquences. Cette modification permet de mieux récupérer les harmoniques du signal.

- ✓ Développement d'un algorithme de détection de coups de batterie et des mesures de similarité.

  Ici, un algorithme de « beat detection » a été proposé. L'algorithme est basé sur la transformation VRT et les techniques de traitement d'images. Cet algorithme produit une courbe de probabilité de « beat ». Le traitement suivant construit à partir de cette courbe une nouvelle forme de représentation rythmique qui est l'histogramme de beat en 2D. L'algorithme permet également d'estimer le tempo précis d'une pièce musicale à partir de l'histogramme 2D. La comparaison de performance d'estimation du tempo avec les techniques existantes a montrée les résultats proches, voir meilleurs selon le cas.

- ✓ Elaboration d'algorithme de détection de notes et des mesures de similarité.

  Dans cette partie de travail, une approche de détection de notes à partir du signal audio a été propose. Cette approche utilise la transformation VRT comme celle qui est utilisée dans toutes les autres parties du traitement. L'algorithme de détection fait la modélisation du spectre afin de trouver les structures

harmoniques correspondantes aux notes. L'information sur les notes permet de construire les mesures de similarité mélodiques telles qu'un histogramme de succession de notes, tonalité et timbre.

- ✓ Application des algorithmes de similarité proposés

  Afin d'obtenir une mesure objective de performance des algorithmes proposés nous avons testé les applications suivantes :

  - Classification automatique de la musique en genres. Dans le cadre de ce test nous avons créé une base de données de référence pour la classification en genres. Cette base comprend environ 1800 titres musicaux classés en 6 genres. Cette base de référence mieux équilibrée par rapport aux bases existantes nous a permis de réaliser les tests de base pour vérifier la validité des mesures de similarité. La méthode que nous avons utilisée est un classificateur kNN appliqué sur les distances de similarité musicales. Nous avons démontré également, que une application des caractéristiques musicales en combinaison avec les caractéristiques basique spectrales permet d'améliorer les taux de classification à une manière significative.
  - Recherche des titres musicaux avec plusieurs interprétations. Ici, nous avons proposé d'injecter une liste de titres musicaux avec plusieurs interprétations dans une base de données de musique. Les essais suivants devaient produire des playlists similaires pour chacun des titres. Ensuite, les positions des autres interprétations du même titre ont été surveillées. Cette expérimentation a démontré les résultats corrects
  - Evaluation de composition de playlists avec avis d'utilisateurs (user feedback). Cette partie d'expérimentation consistait à collecter les votes d'utilisateurs afin d'évaluer la performance des mesures de similarité. L'idée de données « ground truth » était de proposer les playlists similaires avec insertion de titres de manière aléatoire. La suite de la procédure consistait à comparer les histogrammes de votes d'utilisateurs pour les « vraies » similarités et les similarités fausses. Cette expérimentation a montré une différence statistiquement significative entre les distributions de votes pour les vraies et les fausses similarités. Dans le cadre de ce travail, plusieurs méthodes de combinaison de distances ont été proposées.

**Taches 5. et 6. Recherche de la musique par similarité et Reconnaissance de titres musicaux :**

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 fournit 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.

Le LIRIS a proposé un classificateur de titres de musique en genre s'appuyant sur une architecture de comité d'experts où chaque expert utilise des caractéristiques spécifiques. Dans cette architecture des experts de classification par analyse acoustique, des experts de classification par analyse rythmique et des experts de classification par analyse textuelle coopèrent afin de fournir la classification globale de titres de musique. Chacun de ces experts produit une probabilité d'appartenance d'un titre de musique aux différentes classes, qui sont ici des genres musicaux. Chaque classificateur a un poids qui sera attribué par rapport à ses performances sur une base de développement. La somme des probabilités multipliées par les poids des experts constitue la sortie du classificateur général.

Par ailleurs, la classification en genre de la musique a été aussi l'objet des travaux en commun entre LIRIS ECL et ENST. En effet, des premières expérimentations sur un système de classification commun utilisant des descripteurs développés dans chacune de nos deux équipes semblent indiquer que des performances ont pu être améliorées. Les résultats expérimentaux une fois stabilisés feront l'objet d'une communication du projet.

Parallèlement à cette étude, une autre approche a été développée dans le but de réaliser la classification de titres musicaux en genres. Les caractéristiques extraites du signal audio reposent sur son codage selon une analyse temporelle, fréquentielle et temps-fréquence. Ces codages permettent alors de représenter le signal comme une séquence de motifs dont la distribution est analysée selon les lois de Zipf. Les caractéristiques extraites par application de ces lois représentent les entrées du classifieur qui est constitué d'un ensemble de perceptrons multi-couches organisés de manière hiérarchique de façon à respecter la taxonomie des classes.

Un autre aspect de notre travail concerne la définition de mesures de similarités musicales. Une première mesure est une similarité rythmique correspondant à une distance entre deux *beat histograms* (histogrammes des valeurs de périodes rythmiques). Deux autres mesures ont été développées, représentant une similarité mélodique et reposant sur la détection de notes obtenue par une technique inspirée de modèles harmoniques de périodes fondamentales. La première mesure de similarité mélodique repose sur les profils de notes (histogrammes) calculés sur l'ensemble du titre musical permettant d'en déduire la tonalité à partir de la distribution des notes. Deux titres musicaux sont alors comparés en calculant la similarité de leur profil de notes. La deuxième mesure de similarité est basée sur les histogrammes de succession de notes. Ainsi, le nombre d'occurrences des combinaisons de chaînes de trois notes est calculé de manière à en obtenir l'histogramme. Cet histogramme possède trois dimensions, chacune représentant une note de la chaîne. Deux titres musicaux sont alors comparés en calculant la similarité de leur histogramme de succession de notes.

**Rapports de recherche**

En complément du résumé des principales avancées figurant ci dessus, une description plus complète des travaux de recherche des trois équipes se trouve dans les rapports de recherche suivants :

**IRCAM** :

[C. Yeh 07a] Rapport final d'activités de recherche dans MusicDiscover, Reconnaissance de F0s multiples, C. Yeh, Novembre 2007

[A. Livshin 07a] Rapport final d'activités de recherche dans MusicDiscover, Reconnaissance des instruments, A. Livshin, Novembre 2007

[G. Peeters 07a] Rapport final d'activités de recherche dans MusicDiscover, Description sémantique structurée, G. Peeters, Novembre 200

**LCTI** :

[Richard 07] Rapport final d'activités de recherche dans MusicDiscover du LCTI, G. Richard, Novembre 2007

**LIRIS** :

[Chen 07] Rapport final d'activités de recherche dans MusicDiscover du LIRIS, L. Chen, Novembre 2007

## *5.* Réalisations obtenues dans le cadre du projet

- Développement d'un système de base de données spécialisé dans le stockage et la gestion de descripteurs sonores, ainsi que des interfaces avec des environnements de développement
- Mise en place de 5 bases de données de notes isolées
- Spécification, enregistrement et annotation d'une base de données complète de batterie ENST-Drums déjà diffusée à 10 universités dans le monde
- Développement d'un système d'aide à l'orchestration
- Contribution à la segmentation temporelle et à la classification du contenu de documents multimédia
- Constitution d'une base de données commune de *solos* de nombreux instruments
- Développement d'un algorithme de séparation de sources et de transcription d'enregistrements musicaux par modèles bayésiens
- Développement d'un algorithme d'estimation des F0s multiples
- Développement d'un algorithme de séparation de sources harmoniques fondés sur les F0s
- Développement d'algorithmes d'estimation de fréquences fondamentales multiples et de transcription par factorisation en matrices Non-négatives.

- Développement de plusieurs algorithmes d'estimation rythmique
- Développement d'algorithmes de reconnaissance des instruments de musique en contexte d' « échantillons »
- Développement d'algorithmes reconnaissance des instruments de musique en contexte mono-phonique (solos)
- Développement d'algorithmes reconnaissance des instruments de musique en contexte polyphonique
- Développement d'algorithmes d'extraction et de transcription de la piste de batterie à partir d'un signal musical polyphonique
- Spécification, enregistrement et annotation d'une base de données complète de batterie
- Constitution d'une base de donnée de référence de genres de musique. Genres « génériques » et sous-genres : Rock, Rap (HipHop, R&B), Jazz (blues), Classic, Dance (Disco, Electro, House), Metal (Hard Rock, Heavy Metal, Nu Metal). La base de donnée contient en total 1873 titres de 822 artistes soit 37480 secondes. A notre connaissance, c'est la première base de donnée de référence dans laquelle l'attribution des titres aux genres ne se fait pas d'une manière subjective par une personne mais tient compte d'un nombre important de retour utilisateurs.
- Etudes de reconnaissance du genre musical
- Développement de logiciels de traitement du signal musical
- Développement de logiciels de classification automatique en genre
- Développement de logiciels de navigation intelligente dans une base de données musicale
- Développement d'une description sémantique structurée dans un interface de représentation graphique, de navigation et de recherche en base de données.

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

- Réunion des équipes MusicDiscover le 17/11/2004 à L'Ircam, Paris
- Réunion des équipes MusicDiscover le 11/03/2005 au LIRIS, Lyon,
- Présentation d'un poster aux Journées ACI, NOvembre 2005
- Réunion des équipes MusicDiscover le 08/11/2005 à l'ENST, Paris,
- Réunion des équipes MusicDiscover le 24/10/2006 à l'Ircam, Paris,
- Présentation d'un poster aux Journées PARISTIC, Novembre 2006
- Réunion des équipes MusicDiscover le 05/06/2007 à l'IRIS, Lyon,
- Présentation d'un poster et d'une conférence aux Journées PARISTIC, Novembre 2007

## 7. Soutiens obtenus en liaison avec ce projet

**7.1. Postes chercheurs**

D. Tardieu, Doctorant, depuis Janvier 2005, financement sur projets IRCAM essentiellement (projet RIAM SAMPLE ORCHESTRATOR).

H. Papadopoulos, Doctorante, depuis Janvier 2007, financement sur projets IRCAM (projet RIAM ECOUTE).

**7.2. Postes ingénieurs**

D. Fenech , chercheur-développeur, financement MusicDiscover, de Février à Octobre 2007

**7.3. Contrats nationaux**

Le projet RIAM « Ecoute » a été obtenu par l'Ircam en 2006

Le projet RIAM « Sample Orchestrator» a été obtenu par l'Ircam en 2006

Le projet Quaero en cours de pré-financement 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**

**7.6. Contrats industriels**

- Le LIRIS-ECL a obtenu un contrat avec FTRD, débuté en Mai 2005 d'une durée de 12 mois, 100 K€
- L'Ircam a obtenu un contrat de licence avec la société MakeMusic (USA) sur la détection de F0 pour des logiciels pédagogiques (SmartMusic). Année 2005 et nouveaux développements en 2006, montant confidentiel
- L'Ircam a obtenu un contrat de licence avec la société MakeMusique, USA, détection de la hauteur du diapason montant confidentiel
- L'Ircam négocie un contrat de licence avec la société MakeMusique, USA, détection du tempo, des mesures et des temps, montant confidentiel

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

- Contact de l'Ircam avec Queen Mary University of London (E. Vincent)
- Contact du LIRIS-ECL avec le Pr. Dou Weibei, Department of Electronic Engineering, Tsinghua University, Chine

## 9. Publications obtenues dans le cadre du projet

- **Thèses**

[Delezoide06c] Delezoide, Bertrand, Modèles d'indexation multimédia pour la description automatique de films de cinéma, Université Paris 6, Avril 2006
[Vincent04] E. Vincent, Modèles d'instruments pour la séparation de sources et la transcription d'enregistrements musicaux, Thèse, Université Université Paris-6, Décembre 2004
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.
Livshin, A., Décembre 2007. "IDENTIFICATION AUTOMATIQUE DES INSTRUMENTS DE MUSIQUE", thèse de doctorat de l'Université Pierre et Marie Curie, soutenance prévue le 12 Décembre 2007.
Aliaksandr Paradzinets, « Variable resolution transform based music feature extraction and their application for music information retrieval », Thèse de Doctorat de l'ECL, Déc. 2007.
Zhongzhe Xiao, « Classification of emotions in audio signal », thèse de doctorat de l'ECL, Jan. 2008.

- **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).

Hadi Harb, Liming Chen, “A General Audio Semantic Classifier based on human perception motivated model”, Multimedia Tools and Applications, Eds. Springer Netherlands, ISSN 1380-7501 (Print) 1573-7721 (Online), http://dx.doi.org/10.1007/s11042-007-0108-9, March 05, 2007

- **Conférences internationales**

[Delezoide05a] Delezoide, Bertrand, Multimedia classification of movie shots using low-level and semantic features, ACM Multimedia 2005, Singapour, Novembre 2005
[Delezoide05b] Delezoide, Bertrand, Hierarchical film segmentation using audio and visual similarity, ICME 2005, Amsterdam, Juillet 2005
[Delezoide06a] Delezoide, Bertrand, Multimedia movie segmentation using low-level and semantic features, proposed at ACM Multimedia 2006, Santa Brabara, CA, USA, Octobre 2006
[Delezoide06b] Delezoide, Bertrand, Multimedia movie segmentation using low-level and semantic features, proposed at Aximedia 2006, Leeds, UK, Décembre 2006
[Livshin04a] Livshin, A., Rodet, X., « Instrument Recognition Beyond Separate Notes - Indexing Continuous Recordings », *ICMC 2004*, Miami, 2004
[Livshin04b] Livshin, A., Rodet, X. (2004b). *Indexing Continuous Recordings*, Proc. 7’th international conference on Digital Audio Effects (DAFx 2004), pp. 222-227, Naples, Italy.
[Livshin06a] A. Livshin, X. Rodet « The importance of the non-harmonic residual for automatic musical instrument recognition of pitched instruments », Audio Engineering Society (AES) Convention 120, Paris, France 2006
[Livshin06b] A. Livshin, X. Rodet « The Significance of the Non-Harmonic “Noise” Versus the Harmonic Series for Musical Instrument Recognition », *submitted to* 7’th International Symposium on Musical Information Retrieval (ISMIR), Victoria, Canada 2006
[Tardieu04] Tardieu, D., « Synthèse et transformation sonore par descripteurs de haut-niveau », Université Aix Marseille II, 2004. [DEA ATIAM]
[Tardieu05a] Tardieu, D., « Rapport d'avancement de thèse : aide à l'orchestration », 2005
[Tardieu06a] Tardieu, D., Carpentier, G., Assayag, G., Rodet, X., Saint-James, E., ´ IMITATIVE AND GENERATIVE ORCHESTRATIONS USING PRE-ANALYSED SOUNDS DATABASES ª, SMC, Marseille, 2006
[Vincent05a] E. Vincent. Musical source separation using time-frequency source priors. IEEE Transactions on Speech and Audio Processing, special issue on Statistical and Perceptual Audio Processing, 2005.
[Vincent05b] E. Vincent, R. Gribonval and C. Févotte. Performance measurement in blind audio source separation. IEEE Transactions on Speech and Audio Processing, 2005.
Pennsylvvania, USA, 2005.
[C. Yeh04a] C. Yeh and A. Röbel, "*A new score function for joint evaluation of multiple F0 hypothesis*", Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04),pp. 234-239, Naples, 2004.

[C. Yeh04b] C. Yeh and A. Röbel, "*Physical principles driven joint evaluation of multiple F0 hypotheses*", Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA'04, pp. , Jeju, 2004.
[C. Yeh05] C. Yeh, A. Röbel, and X.Rodet, "*Multiple fundamental frequency estimation of polyphonic music signals*", IEEE ICASSP, pp. 225-228 (Vol. III), Philadelphia,
[C. Yeh06a] C. Yeh and A. Röbel, "*Adaptive noise level estimation*", Workshop on Computer Music and Audio Technology (WOCMAT'06), Taipei, 2006.
[Rodet06] H. Kaprykowsky and X. Rodet, "*Globally Optimal Short-Time Dynamic Time Warping Application to Score to Audio Alignment*", IEEE ICASSP, Toulouse, France, 2006.
[C. Yeh06b] C. Yeh, A. Röbel, and X.Rodet, "*Multiple F0 tracking in solo recordings of monodic instruments*", 120th AES Convention, Paris, France, 2006.
[Yeh07] C. Yeh , N. Bogaards and A. Röbel: Synthesized polyphonic music database with verifiable ground truth for multiple F0 estimation, 8th International Conference on Music Information Retrieval (ISMIR'07), Vienna, Austria, 2007.
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

A Paradzinets, O. Kotov, H Harb, L. Chen., Continuous Wavelet-like Transform Based Music Similarity Features for Intelligent Music Navigtion. *Proceedings of CBMI07, Bordeaux, France.* 2007.

Kotov O., Paradzinets A., Bovbel E. Musical Genre Classification using Modified Wavelet-like Features and Support Vector Machines. *Proceedings of EuroIMSA, Chamonix, France.* 2007

A.Paradzinets, H.Harb, L.Chen, Use of Continuous Wavelet-Like Transform in Automated Music Transcription, *14th European Signal Processing Conference (**EUSIPCO06**)* (2006), Florence, Italy

Zhongzhe Xiao, Emmanuel Dellandrea, Weibei Dou, Liming Chen, "Two-stage Classification of Emotional Speech," International Conference on Digital Telecommunications (ICDT'06), p. 32-37, August 29 - 31, 2006, Cap Esterel, Côte d'Azur, France

Zhongzhe Xiao, Emmanuel Dellandrea, Weibei Dou and Liming Chen, "Features Extraction and Selection for Emotional Speech Classification", Proc. of IEEE Conference on Advanced Video and Signal based Surveillance (AVSS 2005), p411-416, Como, Italy, September 15-16, 2005

V.Parshin, A. Paradzinets, L.Chen, "Multimodal Data Fusion for Video Scene Segmentation", 8th Int. Conference on Visual Information Systems (VISUAL 2005: 279-289), Lecture Notes in Computer Science, Springer-Verlag GmbH , ISSN: 0302-9743, Volume 3736 / 2006, Editors: Stéphane Bres, Robert Laurini, ISBN: 3-540-30488-6, Amsterdam, The Netherlands, July 5, 2005, pp. 279 – 28

- **Journaux nationaux**

- **Conférences nationales**

  Présentation d'un poster aux Journées ACI, NOvembre 2005

Présentation d'un poster aux Journées PARISTIC, Novembre 2006
Présentation d'un poster et d'une conférence aux Journées PARISTIC, Novembre 2007