

# Transcription automatique de musique sur partition

## Mots clés :

- **Directeur de thèse** : BERTRAND DAVID
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire Traitement et Communication de l'Information
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

## Résumé du projet de recherche (Langue 1)

La transcription automatique de musique constitue l'un des sujets importants du traitement du signal audio et musical et a reçu une attention soutenue dans la dernière décennie. Cependant, d'une part la transcription de signaux polyphoniques complexes, mêlant des instruments de nature diverse reste un problème difficile et d'autre part, la plupart des systèmes cherchent à obtenir une forme symbolique de la transcription telle que la propose le langage MIDI. L'objectif de cette thèse est d'aborder la transcription (sous forme de partition) de la musique enregistrée. Pour cela, le système doit inférer des paramètres de haut niveau tel que la métrique, la position des premiers temps ou encore la tonalité. Une partie intéressante de ce travail concernera le cas particulier de la partition de Jazz, qui se présente sous forme d'une mélodie surmontée d'un chiffrage de l'harmonie. Un premier travail dans cette direction a déjà été effectué en couplant un détecteur de mélodie [Durrieu10] avec un reconnaissseur automatique d'accords. Les premiers résultats ont été très encourageants mais il est apparu très clairement que les partitions de jazz obtenues étaient difficiles à interpréter pour un musicien [Weil2009]. En effet, les erreurs de chaque sous système (extraction de mélodie, détection d'accords, conversion MIDI vers partition) se cumulent ce qui justifie une approche plus intégrée du problème. Les applications sont potentiellement nombreuses, de la formation musicale à la publication en passant par la recherche automatique de documents à partir de requêtes audio. L'état de l'art La détection de hauteurs multiples et plus généralement la transcription audio vers symbolique est un sujet d'intérêt fort de la communauté scientifique audio comme en témoignent les nombreuses compétitions MIREX sur les sujets afférents (détection du tempo, détection multipitch) et le nombre d'articles récents qui traitent du sujet [Papadopoulos11][Peeters11][Emiya10][Fuentes12][Reis12] A Telecom ParisTech, plusieurs thèses ont eu lieu dans le domaine (V. Emiya 2008, B. Fuentes 2013), ou avec des applications proches (N. Bertin 2008, J-L. Durrieu 2010, Rigaud en cours) ce qui fournit un terreau propice pour cette étude. La thèse de C. Joder (2011) a d'autre part utilisé les CRF pour l'alignement de l'audio avec la partition, ce qui devrait constituer une piste intéressante pour ce travail. Approches Nous proposons deux approches : -\* une approche en deux étapes : audio vers midi puis midi vers partition. Cette approche pourrait permettre d'obtenir un algorithme de référence combinant de la détection multipitch performante et de l'estimation de rythme (tempo, barres de mesure). Cette approche s'appuie sur des traitements de type bottom-up qui utilisent des représentations mi-niveau extraites à partir du signal brut telles que la décomposition en matrice non-négatives pour la détection de hauteur multiples ou la fonction de détection pour le rythme (qui met en valeur les accents rythmiques). -\* une approche intégrée probabilisée (top-down). En suivant par exemple les cadres bayésiens tels qu'a pu initialement les mettre en place T. Cemgil [Cemgil06]. Par ailleurs, dans un cadre précis, contextualisé pour un certain genre de musique, de type d'instrumentation ou de courant musical, nous nous proposons d'apprendre des modèles de langages tels que les N-grams. Les outils de modélisations statistique envisagés sont notamment les Conditional Random Fields [Joder11] ou les Modèles de Markov Cachés [Emiya10, Papadopoulos11].

## Résumé du projet de recherche (Langue 2)

Les principales difficultés de cette tâche tiennent au recouvrement des différents événements dans le plan temps-fréquence (pour la détection des hauteurs multiples), à la détection de la fin des notes (si la détection des attaques est un sujet très étudié, celle des extinctions l'est beaucoup moins) et à l'indétermination intrinsèque entre les variations de tempo et la valeur rythmique des notes (le problème est mal posé dans ce cas). Enfin le sujet étant peu traité dans la littérature, il sera nécessaire de constituer des bases de données appropriées, notamment en ce qui concerne la transcription sous forme de partition de jazz.

## Informations complémentaires (Langue 1)

Nous avons des contacts réguliers avec l'équipe de Ali Taylan Cemgil, {associate professor} au {Dept. of Computer Engineering} à Bogazici University, Istanbul, Turkey. T. Cemgil est un chercheur reconnu de la communauté, spécialiste des champs de markov et des modèles bayésiens appliqués à la musique.

## Informations complémentaires (Langue 2)

{{Aptitudes recherchées}} Nous recherchons un candidat capable de maîtriser à la fois les aspects techniques parfois avancés de ce sujet de thèse (outils de traitement du signal, modélisation bayésienne, modèles statistiques) et ses aspects pratiques (développement sous Matlab et/ou Python, simulations numériques, évaluation sur des bases de données). Il aura en outre une solide formation musicale. Une culture et goût pour les aspects musicologiques seront les bienvenus. {{Equipe d'accueil}} L'étudiant en thèse sera localisé à Télécom ParisTech, département de Traitement du Signal et des Images (TSI), groupe Audio Acoustique et Ondes (AAO), sous la responsabilité de Gaël Richard et Bertrand David. {{Déroulement de la thèse}} Une réunion d'avancement hebdomadaire ou bi-hebdomadaire est usuelle dans notre équipe. A la fin des 6 premiers mois, le candidat rédigera un premier rapport bibliographique et participera à un séminaire pendant lequel il présentera un état de l'art et ses premiers résultats et implémentations. A la fin de la première année, le candidat rédigera un article qu'il soumettra à une conférence internationale. L'objectif en termes de publications à l'issue de la thèse est de parvenir à la soumission d'au moins un article de revue et de 3 à 4 articles de conférences internationales. Au bout de 18 mois, le candidat rédigera un court rapport résumant ses contributions à mi-parcours en les positionnant par rapport à l'état de l'art. Ce rapport devra également contenir une description des travaux envisagés pour le reste de la thèse. Ce document (d'une quinzaine de pages) servira de base pour l'évaluation à mi-parcours. Pour cette évaluation, le candidat présentera oralement ses travaux devant les encadrants et au moins une personne extérieure pouvant être l'un des membres de son futur jury de thèse. {{Valorisation}} Il sera aussi probablement demandé au doctorant de participer à des projets de recherche nationaux et internationaux dans lesquels est impliquée notre équipe. {{Bibliographie}} [Cemgil06] {A generative model for music transcription Audio}, Cemgil, A. T.; Kappen, H. J. & Barber, D., Speech, and Language Processing, IEEE Transactions on, IEEE, 2006, 14, 679-694 [Miyamoto07] {Probabilistic Approach to Automatic Music Transcription from Audio Signals}. Miyamoto, K. ; Kameoka, H. ; Takeda, H. ; Nishimoto, T. ; Sagayama, S. Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on [Weil2009] {Automatic Generation of Lead Sheets from Polyphonic Music Signals} J. Weil, J.-L. Durrieu, G. Richard et Thomas Sikora, Proc. of ISMIR 2009, Kobe, Japan, 2009. [Emiya10] {Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle} Emiya, V. ; Badeau, R. ; David, B. Audio, Speech, and Language Processing, IEEE Transactions on Volume: 18 , Issue: 6, 2010 [Joder11] {A Conditional Random Field Framework for Robust and Scalable Audio-to-Score Matching} Joder, C. ; Essid, S. ; Richard, G. Audio, Speech, and Language Processing, IEEE Transactions on Volume: 19 , Issue: 8 [Fuentes12] {Probabilistic model for main melody extraction using Constant-Q transform} Fuentes, B. ; Liutkus, A. ; Badeau, R. ; Richard, G. Acoustics, Speech and Signal Processing (ICASSP), 2012 [Reis12] {Automatic Transcription of Polyphonic Piano Music Using Genetic Algorithms, Adaptive Spectral Envelope Modeling, and Dynamic Noise Level Estimation} Reis, G. ; Fernandez de Vega, F. ; Ferreira, A. Audio, Speech, and Language Processing, IEEE Transactions on