Sujet de thèse - campagne 2017-2018

Laboratoire : Institut d’Alembert
Etablissement de rattachement : Université Pierre et Marie Curie, Paris, France
Titre de la thèse : Modèles de Markov Cachés pour l’analyse musicologique automatique des improvisations musicales.
Directeur de thèse : Prof Olivier ADAM
Mail de contact : olivier.adam@upmc.fr
Codirection éventuelle : DR Grégory NUEL
Collaborations dans le cadre de la thèse : LPMA Université Pierre et Marie Curie
Rattachement à un programme : non
Cotutelle envisagée : non
Si oui avec quelle université & quel laboratoire : non
Le sujet peut-il être publié sur le site web de l’ED SMAER : oui

Résumé du sujet :

L’objectif principal de la thèse est de développer des outils mathématiques pour l’analyse musicale dans le but de décrire la structure des morceaux et les styles de jeu des musiciens (motifs mélodiques, tonalité, modes). Ce travail visera à caractériser les degrés de liberté des musiciens notamment lors de séances d’improvisation, en particulier pour des musiques traditionnelles dont on sait que les codes culturels pourraient finalement s’opposer à cette notion de musique improvisée.
Après une étude bibliographique sur les différentes méthodes d’extraction automatique, nous nous intéresserons à comment introduire de l’information musicale a priori, notamment via le PLCA déjà utilisé au laboratoire. Nous nous focaliserons également sur les motifs récurrents, via le recours à des HMM.
Sujet développé

Currently, most international MIR research programs focus on “euro-genetic” music, with a domination of the classical piano repertoire. This music is highly standardized, both in the characteristics of its language (i.e. temporal organization of notes) and of its instruments. In this PhD, we are interested in less standardized music, including musical improvisation and “non-eurogenetic” orally transmitted musical repertoires. In particular, idiomatic characteristics encountered in these repertoires, such as the absence of written supports, the non-standardization in the instrument-making, or unusual musicological codes (e.g., non-tempered scale, micro-rhythms), would bring complex and original challenges to the MIR community. In this PhD project, following our research efforts in recent years, we will work on the musical repertoires of different plucked-string traditional instruments: the marovany of Madagascar, the mvet of Cameroun, the N’Goni of Mali, the sitar of India, the GuQin and Pipa zithers of China, but also euro-genetic instruments like piano and folk guitars to relate our studies to usual AMT frameworks. This ensemble of musical repertoires propose different degrees of improvisation. For example, the marovany repertoires typically consist of original melodic motives, which represent an important culture patrimony of Madagascar, heritated from generation to generation. From these motives, each musician proposes its own interpretation, with personal variations and improvisations around these motives. We will then see how classical analytical framework to generate prior knowledge can be transposed in spite of the differences between these repertoires, and how it can be adapted to new musical features. This work should then better understand how such a framework can be generalized to different instrument repertoires.

Objectives

The fundamental PhD objective is to develop computational tools of musical analysis to organize and structure audio recordings from musical improvisations and orally transmitted repertoires. In more practical terms, we wish to develop automatic tool of music analysis to assist musicologists in their work. AMT systems fit well this goal as they allow to retrieve the minimal information from complex audio so as to reconstitute the underlaying music score, and could be used in a variety of user scenarios such as searching and organizing music collections with barely any human labor. Higher-level analysis will also be considered, so as to provide direct insights into a musical piece structure through musical concepts such as melodic motive and tonality/mode. In definitive, this PhD project aims to investigate how state-of-the art probabilistic and statistical methods can assist musicological experts in their analysis of specific musical repertoires. A focus will be put on musical improvisation, drawing methods from our past analysis of orally transmitted repertoires, which will be seen as more coded and prescripted improvisations, in comparison to more integral and free improvisations as in certain styles of jazz (e.g. free jazz).

Objective #1 : Developing robust ready-to-use repertoire-specific AMT systems

This first objective aims to provide robust transcriptions of a musical repertoire, in close collaboration with music experts (ethnomusicologists, musicians), who only them have the authority to validate the quality of a transcription automatically generated. The original instrument repertoires of this PhD present complex musical features (e.g., the GuQin zither and sitar repertoires contains many harmonic notes, vibrato, glissandi and sympathetic resonances), which have not yet been studied in Music Information Retrieval, and will present interesting challenges for both signal processing and ethnomusicology disciplines in terms of information representation and extraction. As in Dorian Cazau’s PhD project, a model-based approach (also called blackboard approach in literature) will be favoured in this PhD, with the characterization and incorporation of repertoire-specific multi-dimensional prior knowledge into statistical frameworks modeling the generative process of musical signals.

A first step of this PhD thesis will then be to identify the pertinent acoustic information depending on the traditional musics we would like to focus on. This information could include both timbre (i.e. modeling of the generic “morphological” features of the sound related to the physics of an instrument, e.g. intermodulation, sympathetic resonances, inharmonicity) and musicological (e.g. harmonic transition, playing dynamics, tempo and rhythm) classes. Inter-disciplinary collaboration is crucial here, to define the most pertinent and discriminative musical information, which should be the most usefull in computational systems of automatic transcription.

Objective #2 : Higher-level structure segmentation of a musical piece

In this second objective, we will work on a motive-based structure segmentation of a musical piece. Both as a tool towards the understanding of a piece harmonic structure, but also as an input knowledge able to inform retroactively a
Ecole doctorale SMAER
Sciences Mécaniques, Acoustique, Electronique, Robotique

Acoustic modeling, through statistical methods like PLCA for example, has been widely studied. To overcome the signal, a simple thresholding is eventually performed to make the decision on which events are actually played or not.

Methods based on sophisticated Hidden Markov Models (De Fonzo et al., 2007; Koller, D. and Friedman, N., 2009) (ex: envelopes, together with a simple musical key model). In this PhD, we will adopt this modeling approach, drawing from pitch modeling, using basic musicological models on music language (Raczynski et al., 2013, 2014). Ryynanen and limitations of current state-of-the-art multiple pitch transcription models, some researchers have worked on a symbolic view where every posterior distribution is computed through belief propagation. In the particular case of HMM, belief propagation reduces itself to the classical forward/backward algorithms. For the statistical inference, the basic approach will be the maximum likelihood, either directly by numerical optimization (ex: gradient descent, Newton-Raphson, etc.) or mixed with the (generalized) Expectation-Maximization algorithm.

One of the substantive objectives of this project is then to try understanding how musical objects of motives can be used optimally musicological analysis on one side, and in automatic indexing and classification tasks on the other side.

Although AMT is a challenging task which will probably require much more years of fundamental research to provide satisfying results, we really believe that a low-level conservative form of AMT methods could be derived to serve successfully this task of structural segmentation based on repetitive motives. Second, to tackle the tricky musicological question of what constitutes and contributes to similarity of music, which is typically addressed in musicology by studying musical structures on the symbolic level (i.e. in a score), we will draw upon string-based pattern matching methods to develop comparisons tools able to include the multi-dimensionality aspect of a motive. The goal of this technique is to break down a global similarity score from into its different music-related components (e.g. motif salience, rhythmical pattern, harmonic progression).

Objective #3 : Keep on developing the PSIFAMT sound database
To compare performance of different AMT algorithms, it is necessary to test these methods on simulated audio signals and real music. However, the international community of scientists working on Music Information Retrieval has mainly focused on classical piano. Academic and commercial music have received the most attention from researchers, while non-eurogenetic repertoires have been totally left outside. Then, one part of this PhD thesis will be dedicated to collect new music recordings and to enlarge our dataset with songs from professional musicians.

This large music dataset will be dedicated to plucked-string instruments, including widely traditional ones, are multifold:
- Making available a new sound database for AMT evaluation in the MIR community, which is now indispensable to test the robustness of current algorithms to music diversity, and face them to more complex musical features;
- Plucked-string instruments present a great diversity in terms of timbre (e.g. inharmonicity, envelop spectrum variability, temporal profile modulation), playing styles and effects (e.g. different string excitation modes, palm muting, glissandi, vibrato, tremolo, harmonic notes);
- Put the attention on non-eurogenetic traditional repertoires, encompassing the greatest diversity of musicological codes (e.g. micro-rhythm, tuning based on non-equal temperament, modal scale) and of musical practice, and complexity in music encoding.

Methods
Most Multi Pitch Estimation methods are based on an acoustic modelling of note events. A salience score is then estimated for each of this event, according to the similarity degree between their respective models and the unknown signal. A simple thresholding is eventually performed to make the decision on which events are actually played or not. Acoustic modeling, through statistical methods like PLCA for example, has been widely studied. To overcome the limitations of current state-of-the-art multiple pitch transcription models, some researchers have worked on a symbolic pitch modeling, using basic musicological models on music language (Raczynski et al., 2013, 2014). Ryynanen and Klapuri (2005) proposed a melody transcription method that uses a Hidden Markov Model (HMM) to model note envelopes, together with a simple musical key model. In this PhD, we will adopt this modeling approach, drawing from methods based on sophisticated Hidden Markov Models (De Fonzo et al., 2007; Koller, D. and Friedman, N., 2009) (ex: constrained models, several observed sequences, non standard parametric distributions, several layers of hidden sequences, motifs, etc.). In order to deal with these models, we will adopt a Bayesian network point (Jensen, 1996) of view where every posterior distribution is computed through belief propagation. In the particular case of HMM, belief propagation reduces itself to the classical forward/backward algorithms. For the statistical inference, the basic approach will be the maximum likelihood, either directly by numerical optimization (ex: gradient descent, Newton-Raphson, etc.) or mixed with the (generalized) Expectation-Maximization algorithm.
References
Cazau, D.; Chemillier, M. & Adam, O. Information retrieval of marovany zither music with an original optical-based system Proceedings of DAFx 2013, Maynooth, Ireland, 2013, 1-6
Mysore, G. J. A non-negative framework for joint modeling of spectral structure and temporal dynamics in sound mixtures Stanford University, USA, 2010
Raczynski, S., Vincent, E. and Sagayama, S. Dynamic Bayesian networks for symbolic polyphonic pitch modeling IEEE Transactions on Audio, Speech and Language Processing, Institute of Electrical and Electronics Engineers, 2013, 21, 1830-1840
Ryynanen, M. Probabilistic Modelling of Note Events in the Transcription of Monophonic Melodies Tampere University, 2004