

Deep symbolic learning of multiple temporal granularities for musical orchestration.

Résumé

1. **Directeurs**
Florent JACQUEMARD, Geoffroy PEETERS
2. **Encadrant**
Philippe ESLING
3. **Candidat**
Léopold CRESTEL
4. **Mails**
philippe.esling@ircam.fr, florent.jacquemard@ircam.fr, geoffroy.peeters@ircam.fr
5. **Laboratoire**
Institut de Recherche et Coordination Acoustique / Musique (IRCAM)
6. **Axe.1**
“Le développement des services numériques pour l'accès à la connaissance et à l'information, le traitement des données numériques.”
7. **Axe.2**
“La compréhension de l'humain dans ses dimensions cognitives, neurophysiologiques et biomécaniques”

Deep symbolic learning of multiple temporal granularities for musical orchestration.

Overview

Orchestration is the subtle art of writing musical pieces for orchestras. Hence, it requires combining instrumental properties in order to reach ideas of timbre coming from the blending and combination of individual spectral properties. The paramount challenge in this field is that there has never been any systematic research in music analysis and theory on orchestration, a gap this project would start to address. Orchestration practice involves solving complex problems implicating vast knowledge of the properties of musical instruments and how they operate. In contemporary classical music where timbre has become a primary structuring force, computer-based tools are being developed in order to help composers solve problems of finding instrumental combinations that could fulfill the desired musical goals. The topic of orchestration encompasses open issues from auditory perception, music analysis and composition, signal processing, computer science and combinatorics [2]. Integration of systematic music analyses based on perceptual principles and experiments into computer-aided orchestration platforms would revolutionize creative and pedagogical practice. However, the problem of orchestration is a highly combinatorial NP-problem. Indeed, if we solely consider the range of notes, dynamics, chords and playing modes provided by a single musical instrument, we can foresee the infinite number of combinations provided by an orchestra [3].

Recently, the field of *deep learning* [4] has witnessed a tremendous interest amongst the machine learning community. Its goal is to train connexionist architectures for artificial intelligence similar to those constructed in the well-known neural networks. However, the typical feed-forward perceptrons are usually bound to small depth (number of layers) because of the *gradient diffusion* problem that lead to inefficient learning in deeper networks. By relying on greedy layer-wise training algorithms [5], it is possible to train each layer of the networks independently and in an unsupervised manner. This leads to various types of architectures such as *deep belief networks* that can display an extensive number of layers [6]. This leads to deep neural architectures of representations similar to those found in the brain for perceptual task such as visual or auditory processing. This approach appears as a valid candidate to learn models of AI that could bypass the problem of the *curse of dimensionality*, but also to learn automatically higher-level abstractions for both symbolic and temporal datas. Even though these principles are undergoing intensive research in signal processing communities, its implications on symbolic representations of music are yet to be explored, a research direction that we call *deep symbolic learning*. A first step towards such a system would be to study the learning behavior of connexionist architectures with regards to orchestral composition. These researches could be enhanced by current knowledge in perceptual effects of orchestration led by a partner laboratory in McGill university. Furthermore, orchestration lies at the crossroads of signal (sound) and symbolism (musical score) but also thrives between micro-temporal evolution and macro-temporal articulations. Hence, this project is aimed to explore neuroscientifically-informed machine intelligence systems for symbolic data with the non-determinism related to the exploration of a highly combinatorial space (because of the quasi-infinite possibilities offered by the orchestral colors) but also centered around temporal problematics and their inherent variable granularities [1]. The goal of this PhD project is to provide an approach that could help in translating the intent of a composer in the process of orchestration. Hence, the main idea is to first learn the inherent structures that co-exist between different musical elements (relationships inside the symbolic knowledge of musical scores, between different signals but also between the signal and the score). Then, based on the learned connexionist architectures of representation, the system could propose some re-orchestration and original improvisations. That way, this system could provide an interaction between the spectrum of sounds and their evolution in time.

Project and aims

The proposed project aims to transform orchestration analysis and practice through a unique opportunity to bring together for the first time knowledge from music theory, perceptual psychology, neuroscience, digital signal processing, and artificial intelligence in such a way as to provide a scientific analysis and theory of the subject with transformative implications for the practice and teaching of music composition, orchestration, music psychology, and music theory. The main goals within the time frame of this proposal are to develop novel algorithms of *deep symbolic learning* centered on orchestration through an extensive validation by perceptual experiments and to apply the knowledge and tools acquired to the development of radically new orchestration pedagogies based on

combined symbolic manipulation and real-time auditory experience. These activities will be coordinated to form the foundations of a scientific theory of orchestration.

The main idea of this project is first to develop a library embedding several paradigms of deep learning, by normalizing various implementations found in the literature into a comprehensive open-source package. This will allow to automatize the testing of different neural architectures and various perceptual assumptions derived from current knowledge in auditory perception. By refining the parametrization and testing the different architectures, the library could be trained and taught on the database of symbolic orchestral effects currently gathered by a partner laboratory of auditory perception in McGill university. Then, the goal of this PhD will be to extract perceptually meaningful temporal motifs from data related to multiple instruments playing simultaneously, and try to relate both sources of informations. Hence, it introduces hard questions : what representations to associate to this temporal information (symbolic, signal, hybrid) ? What sort of relations should be uncovered between different sources ? How to model these relationships and how to perform reasoning based on this model ? Finally, interesting phenomenon might be more or less visible, depending on the temporal granularity of the representation. Hence, an interesting aspect of this work would be to study multi-resolution representations to model time series with different granularity levels. These steps are intended to uncover relationships that exist between the signal and symbolic properties of musical lines and their combinations. However, the learned connexionist architecture and its weight can easily be inverted to lead to a generative model. Hence, these analyses could lead to computer-aided orchestration and improvisation systems that should be tested perceptually in selected cases against recordings of a real orchestra to determine the adequacy of the proposed solutions and playability of the results by human performers. Hence, the integration of learning algorithms in this framework could allow their assessments in musical frameworks. The main scientific challenges can be delineated as follows :

1. *Develop the architecture of a deep learning library in C++ language*, merging several approaches from open-source libraries in a modular way
2. *Analyze symbolic and numeric transformations for multivariate time series*, including the multiple granularities approaches in order to refine the deep symbolic learning.
3. *Analyze the efficiency of various connexionist architectures and approaches* applied to multi-source and multivariate symbolic and numeric time series of musical orchestration.
4. *Compare the learning results to the perceptual studies* in terms of meaningful motifs for orchestral effects.
5. *Invert the learning algorithms architecture* in order to devise a generative model for computer-aided orchestration and improvisation
6. *Assess the library in musical composition and creation processes*

Outcomes

The student is expected to perform the first exploratory work on deep symbolic learning in order to produce publications in high impact-factor journals and also usable softwares with a direct application to perceptual psychology, time series analysis and musical orchestration. Regarding existing softwares, the student will be able to benefit from the wealth and international recognition of softwares produced by IRCAM in the last two decades.

References

- [1] Esling P. and Agon C. "Time series data mining", *ACM Computing Surveys*, vol.45, no.1, 2012.
- [2] Esling P. and Agon C. "Taking all advices in time series classification for musical orchestration" *in prep.*
- [3] Esling P. and Agon C. "Multiobjective time series matching and classification for audio retrieval" *IEEE Transactions on Speech Audio and Language Processing*, vol. 21, no. 10, 2013.
- [4] Bengio, Y. "Learning Deep Architectures for AI", *Foundations and Trends in Machine Learning*, vol.2, no.1, pp. 1-127
- [5] Bengio, Lamblin, Popovici & Larochelle "Greedy Layer-Wise Training of Deep Networks", *NIPS'2006*
- [6] Hinton, Osindero & Teh "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, 2006