Hierarchical variational temporal learning for dynamic musical audio synthesis

Apprentissage temporel variationnel hiérarchique pour la synthèse audio musicale dynamique

Mots clés : Array

Directeur de thèse : Jean-Pierre Briot
Co-encadrant(s) :
Unité de recherche : Sciences et Technologies de la Musique et du Son
Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
Domaine scientifique principal : Sciences et technologies de l'information et de la communication

Résumé du projet de recherche (Langue 1)

Generative systems are machine-learning models whose training is based on two simultaneous optimization tasks. The first is to build a latent space, that provides a low-dimensional representation of the data, eventually subject to various regularizations and constraints. The second is the reconstruction of the original data through the sampling of this latent space [1]. These systems are very promising because their space is a high-level, “over-compressed” representation that can be used as an intermediate space for several tasks, such as visualization, measurements, or classification. However, one of the most prevalent problem of ML algorithms applied to musical creativity is that they only process a single temporal scale or at best a finite set of small scales. The goal of this project is to work on an approach able to process multiple temporal granularities through a hierarchical multi-scale processing [2, 3]. Hence, the main goal will be to develop a recursive form of learning by iteratively learning increasingly temporally complex signals. A first approach towards this idea is to first learn a variational latent space on small chunks of audio (or audio grains) directly from the raw audio and iteratively build latent spaces for more long and complex audio samples [4]. The goal of this PhD will be to both provide a generative system based on raw audio, but also to evaluate its use and control for musical creativity. First, most current waveform methods are heavily computationally expensive because they fail on local feature extraction [5, 6]. Indeed, most of these methods are deterministic, and thus do not alleviate the problem with a probabilistic approach that could allow the system to catch meaningful features, but also do not provide intermediary levels of knowledge, that could be useful for feature extraction or generation interface. In this project, we thus aim to rather enforce the system to catch local properties at a small scale coming from unsupervised learning, to extract useful representation of this local structure to then catch useful temporal information.

Informations complémentaires (Langue 1)