# ATIAM 2018 - ML Project
# Latent sequencing for dynamic musical patterns

Axel Chemla–Romeu-Santos[1,2] , Philippe Esling[1*]
[1] Università Degli Studie di Milano, Laboratorio d'Informatica Musicale (LIM)
[2] Institut de Recherche et Coordination Acoustique Musique (IRCAM)
UPMC - CNRS UMR 9912 - 1, Place Igor Stravinsky, F-75004 Paris
{chemla, esling}@ircam.fr

November 16, 2018

**Abstract**

*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. 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. The main goal of this project is to use variational models in both audio and symbolic worlds, and to make them interact to have a end-to-end, full and controllable instrument.

## 1 Introduction

Among recent generative systems found in the literature, two have had a large success in the machine learning community. First, the *variational auto-encoder* (VAE) is based on a two-stage inference/generation procedure that showed great generalization properties and good reconstruction abilities despite of its light structure. VAEs are based on a twofold procedure : first an *encoding* pass, that projects input data to an abstract space called the *latent space*, and a *decoding* pass, that gives back the corresponding data from the latent position. Latent space can thus be understood as a *representation* for the learned database, that can be navigated freely to generate coherent data with the original data space.

We propose here to develop two different procedures : first, we will use this framework in the symbolic world by developing a rhythmic pattern generator,

*https://esling.github.io/

where we can generate stochastic patterns from the interaction with this latent space. Then, we will use it to generate the triggered sounds, in a way that this system is compatible with the variational pattern generator.

**Rhythm generation.** Here, we first propose to train variational auto-encoders on *sequencer patterns* to extract a "rhythm" space, that can be navigated freely to interpolate between different patterns in real-time. First, this requires the construction of a dataset. Then, you will have to construct the Variational Bayes algorithm, and proceed to the training. Introduction of any label information can be involved during the training to provide a label-sensitive latent space.

**Sound Generation.** The aim here is to learn the sounds that will be triggered by the latent rhythm generator. In the same way than before, you will have to construct a dataset (can be different from drums). Thus, we here generate sounds by sampling one position in latent space. As this variational synthesizer has to be controlled by the pattern generator, you will have to think about a way to make this synthesizer controllable.

**Control and refinement.** You knows thus have latent sequencer with a variational synthesizer. Now that the main core is developed, how could you improve the interaction of a user with this synthesizer (additional knobs, additional options, randomizations, you are free about almost anything...!) ?