GENERATING STRUCTURED DRUM PATTERN USING VARIATIONAL AUTOENCODER AND SELF-SIMILARITY MATRIX

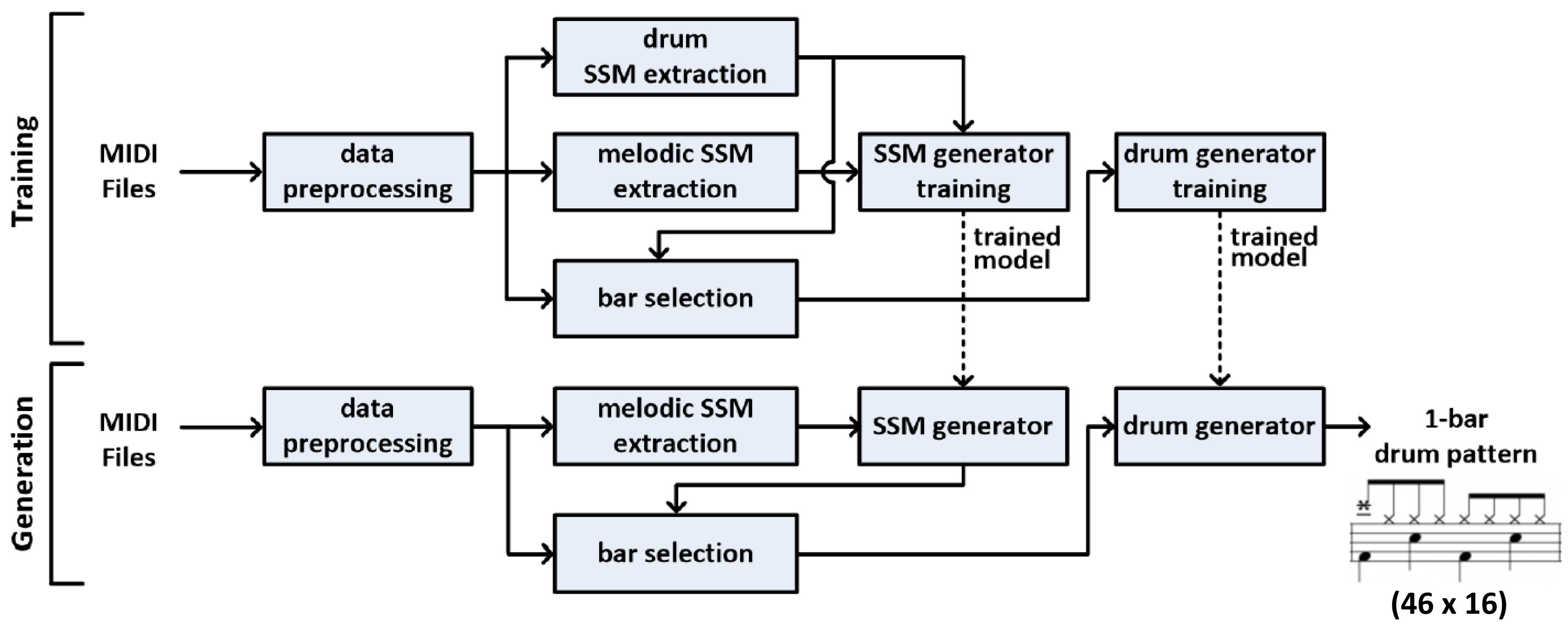
I-Chieh Wei[1], Chih-Wei Wu[2] and Li Su[1]

[1]Institute of Information Science, Academia Sinica, Taiwan, and,[2]Netflix, Inc., USA

Summary

The challenge of automatic music generation is to generate patterns that are both **rhythmically** and **harmonically** cohesive throughout the piece. In this work, we focus on **conditional music generation** which aims to **generate rhythmically compatible percussive track** given the melodic tracks. we proposed a model that incorporates long term structure into the music generation process. The subjective evaluation results suggest the effectiveness of the proposed method.

System overview

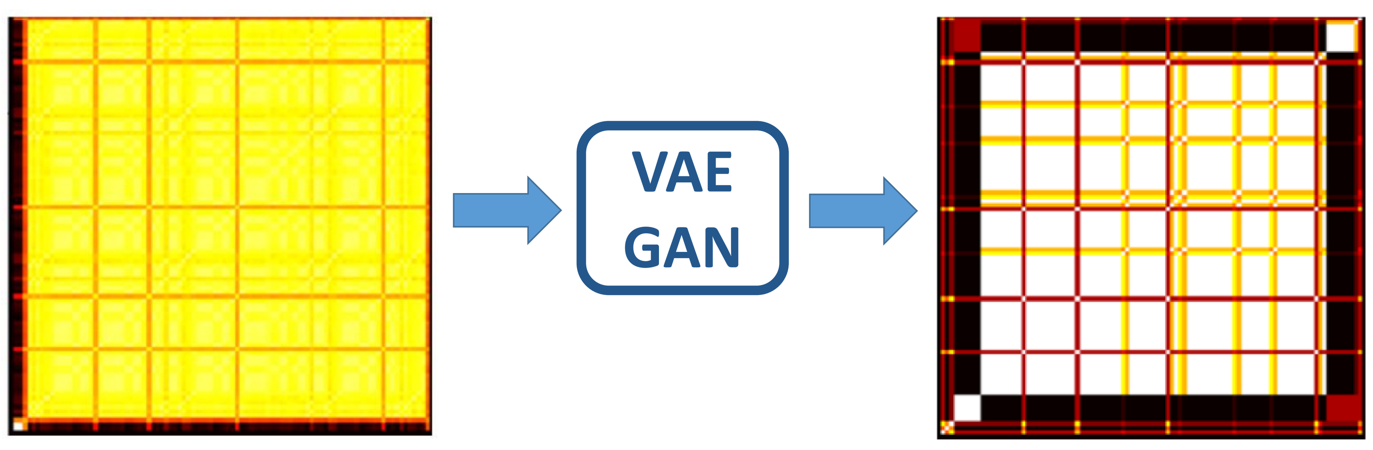


- **Drum SSM generator**

  Melodic and drum SSM generally share a similar structure. (SSM example is shown below). This similarity makes the inferencing possible through a predictive model. A **VAE-GAN** model is trained in this work to infer a drum SSM from a melodic SSM input. The SSM generator is trained by minimizing the loss function:

$$\mathcal{L}_{ssm} = -\mathrm{E}_{v \sim q_s(z_s|s_m)}[\log p_s(s_d|z_s)] + \mathrm{KL}(q_s(z_s|s_m)\|p(z_s)) + \mathcal{L}_D$$

SSM example of "Can't Buy Me Love" by Beatles



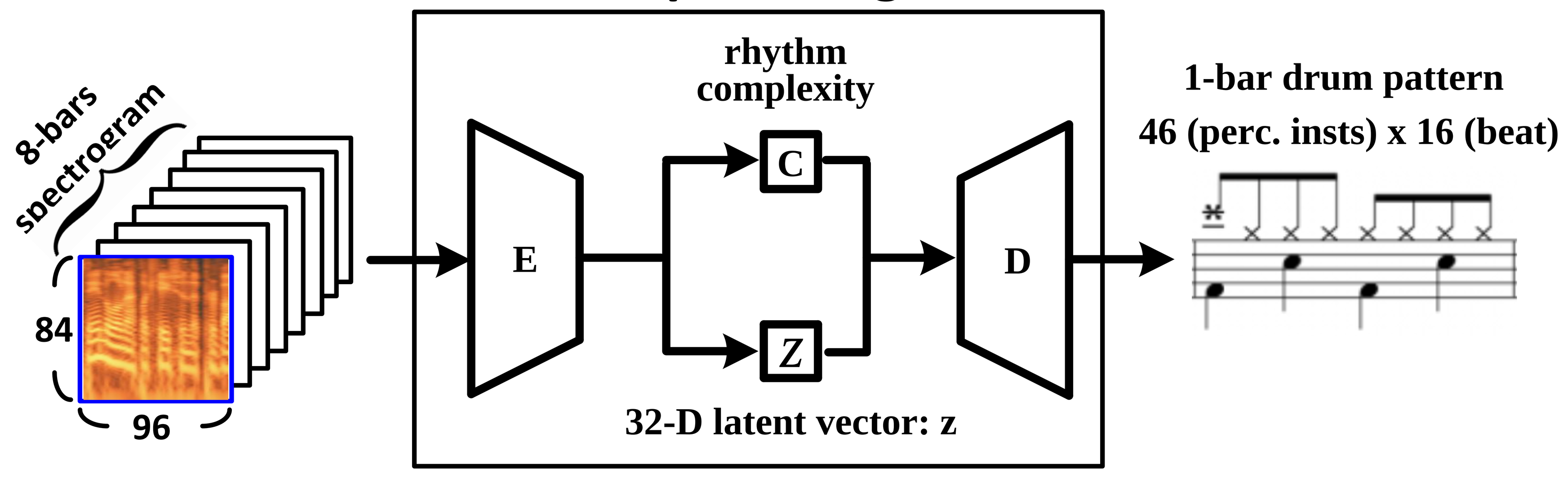
melodic-tracks SSM (256 x 256) **drum-track SSM (256 x 256)**

- **Drum pattern generator**

  Selected 8-bars spectrograms are feed into a VAE-based drum pattern generator as input. The model is trained to generate a single-bar drum pattern output for current bar. The following loss function is minimized during training stage:

$$\mathcal{L}_{drum} = -\mathrm{E}_{z \sim q(z|y)}[\log p(x|z)] + \mathrm{KL}(q(z|y)\|p(z)) + r(c;\hat{c})$$

Drum VAE pattern generator



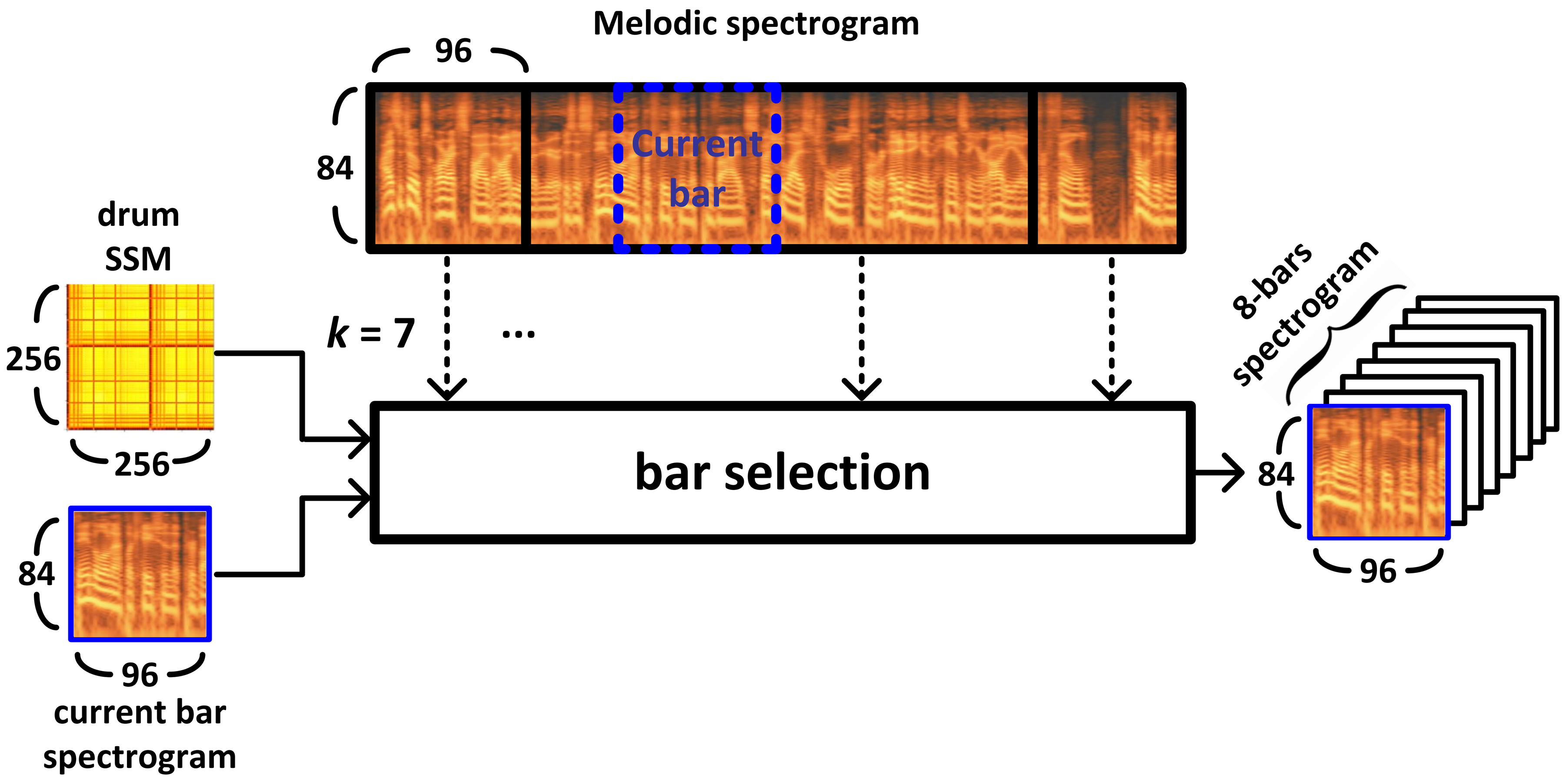
- **Dataset**

  In this work, we use the **L**akh **P**ianoroll **D**ataset (**LPD-5**), LPD-5 contains 21,425 songs and each song has five tracks (piano, guitar, string, bass, and drums) extracted from the original MIDI.

Code and listening example:

https://sma1033.github.io/drum_generation_with_ssm

Key idea — bar selection via SSM



- **Select reverent bars**

  k-nearest bars are located for every bar-level spectrogram according to the given song's drum SSM. These bars are identified by finding the ***k*-smallest distance** in every column of the drum SSM.

- **8-bars spectrogram feature**

  Together with the current bar spectrogram, the 8-bars (7+1) spectrogram are weighted then stacked into an eight-channel feature representation, which will be used as the input to the subsequent drum pattern generator.

Evaluation

- **Compare four methods**

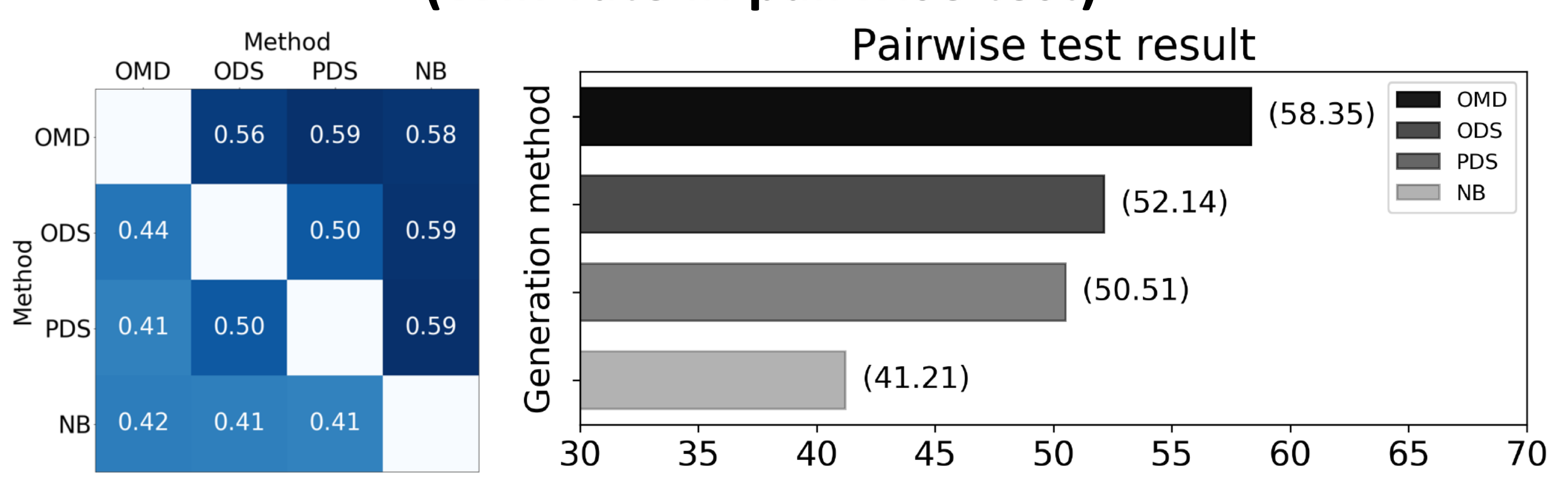
  We evaluate the drum tracks collected from the four methods:

  - ***OMD*** – **O**riginal **M**IDI **D**rums are the drum patterns predefined in the source MIDI files.
  - ***ODS*** – **O**riginal **D**rum **S**SM is the proposed model except that the drum SSM used for bar selection is computed from ***OMD***.
  - ***PDS*** – **P**redicted **D**rum **S**SM is the model we proposed.
  - ***NB*** – **N**eighboring **B**ars is the baseline model that uses no structural information for generation.

- **Pairwise comparison**

  Pairwise performed for drum tracks from the four methods. The result shows the effectiveness of our proposed model.

(Win-rate in pairwise test)



- **Subjective evaluation from musicians**

  Descriptive feedback from three musician is collected, such as, **ODS** - good distinction between sections; NB - unstructured yet unexpected.