PianoTree VAE: Structured Representation Learning for Polyphonic Music

ISMIR 2020 Poster Presentation

Ziyu Wang[1], Yiyi Zhang[2], Yixiao Zhang[1], Junyan Jiang[1], Ruihan Yang[1], Junbo Zhao[3], Gus Xia[1]

[1]Music X Lab, NYU Shanghai, [2]Center for Data Science, New York University, [3]Computer Science Department, Zhejiang University

{ziyu.wang, yz2092, yixiao.zhang, jj2731, ry649, j.zhao, gxia}@nyu.edu

Key Idea: Tree-structured VAE Built upon Polyphonic Musical Syntax

- A novel **data structure** for polyphonic music.
- A VAE **model structure** having a one-to-one correspondence with the data structure.
- Experiments show the proposed **structure** yields better latent space for generation purpose.
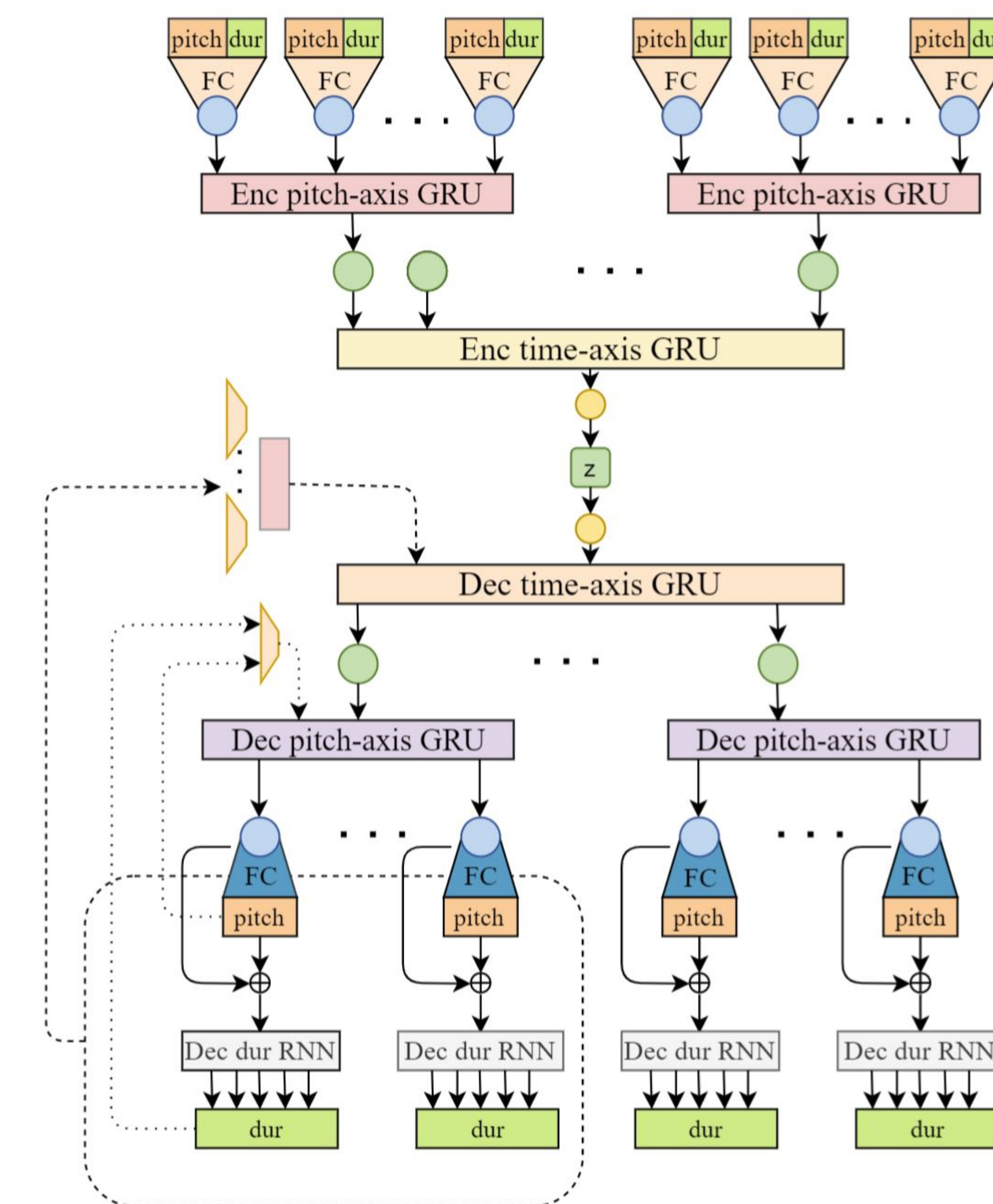
Data Structure

Deep Structure

Surface Structure
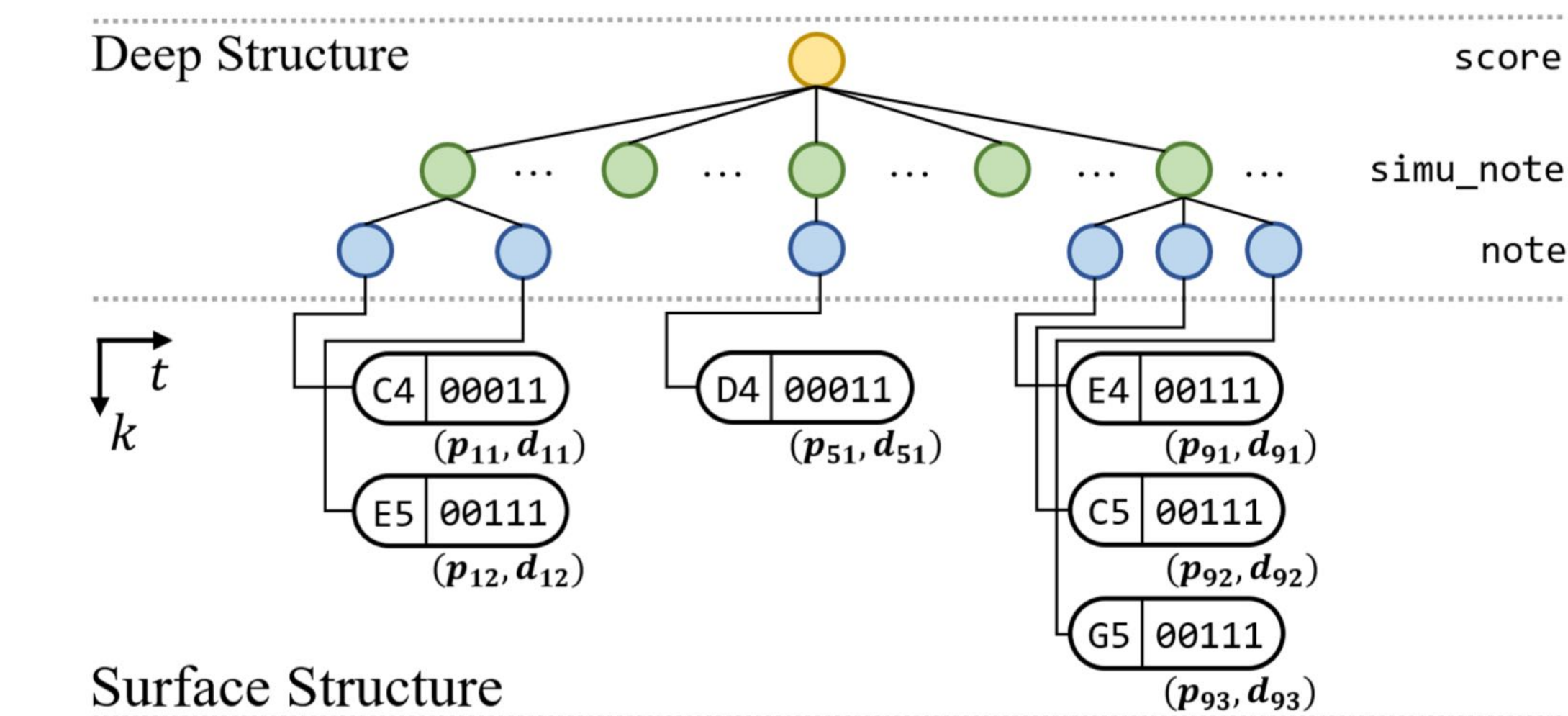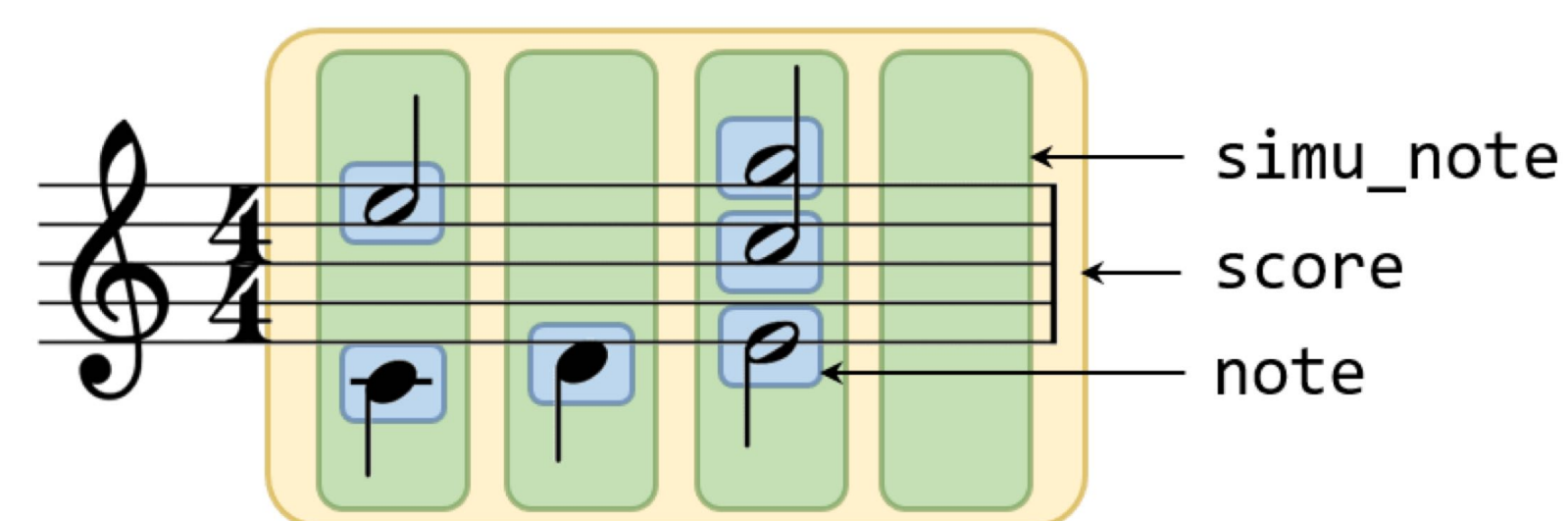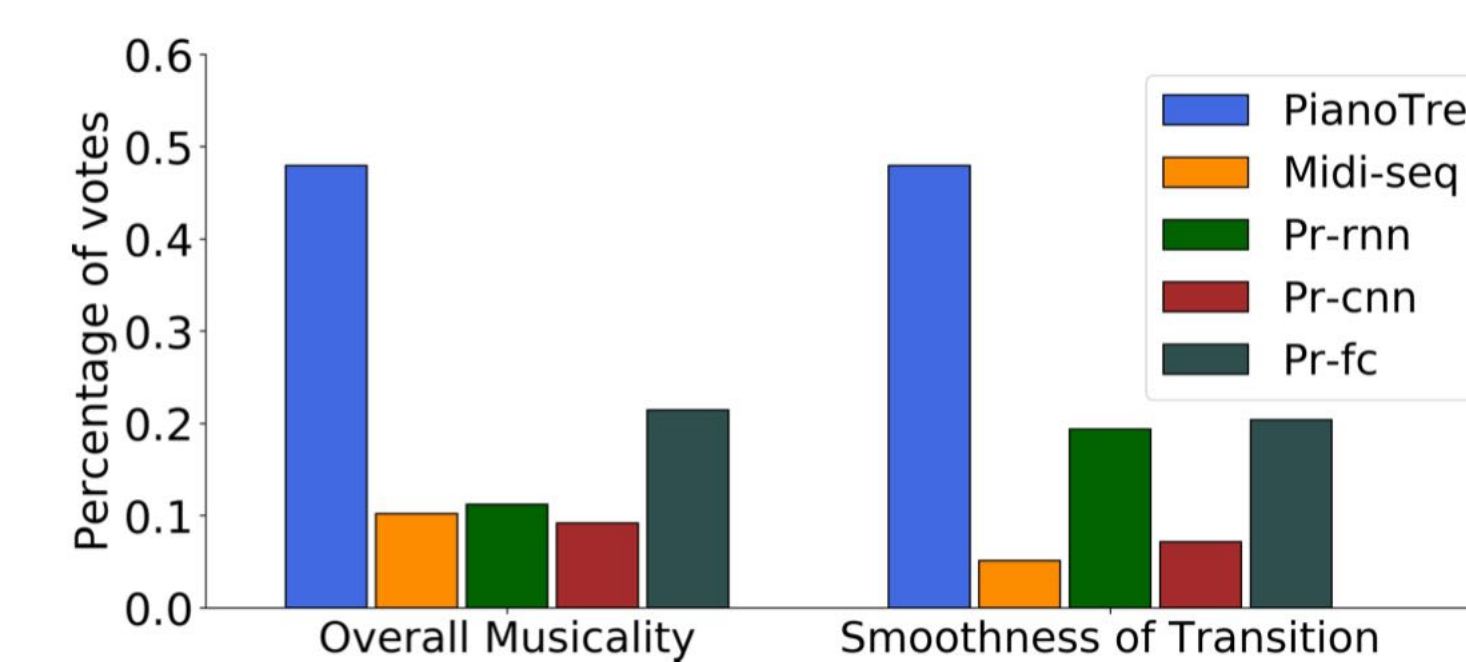
Source Code & MIDI Demo

Model Structure

Experiments

• Reconstruction

Models	PianoTree	midi-seq	pr-rnn	pr-cnn	pr-fc
Onset Precision	**0.9558**	0.8929	0.9533	0.9386	0.9211
Onset Recall	**0.9532**	0.6883	0.9270	0.8818	0.8827
Onset F1	**0.9545**	0.7774	0.9399	0.9093	0.9015
Duration Precision	**0.9908**	0.3826	0.9777	0.9757	0.9688
Duration Recall	0.9830	**0.9899**	0.9891	0.9796	0.9743
Duration F1	**0.9869**	0.5519	0.9834	0.9777	0.9715

• Latent Space Interpolation

• Latent Space Visualization

a) The distribution of notes with different pitches.
b) Overview: Helix of notes with different pitches and durations.
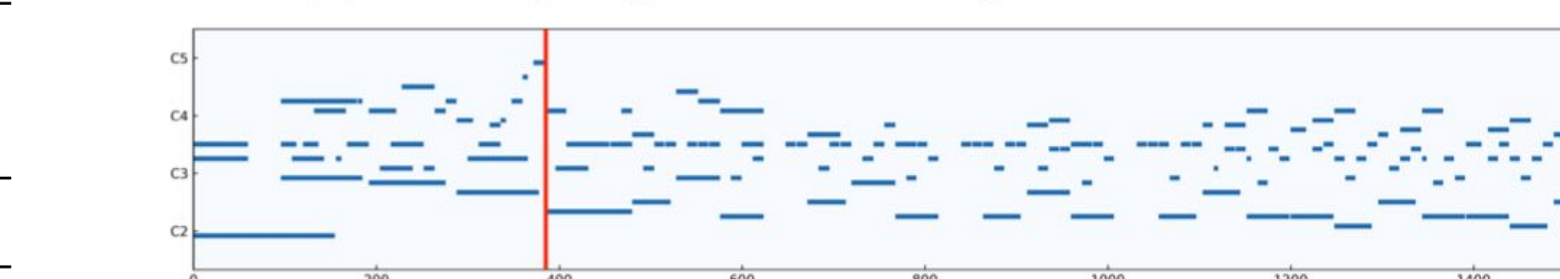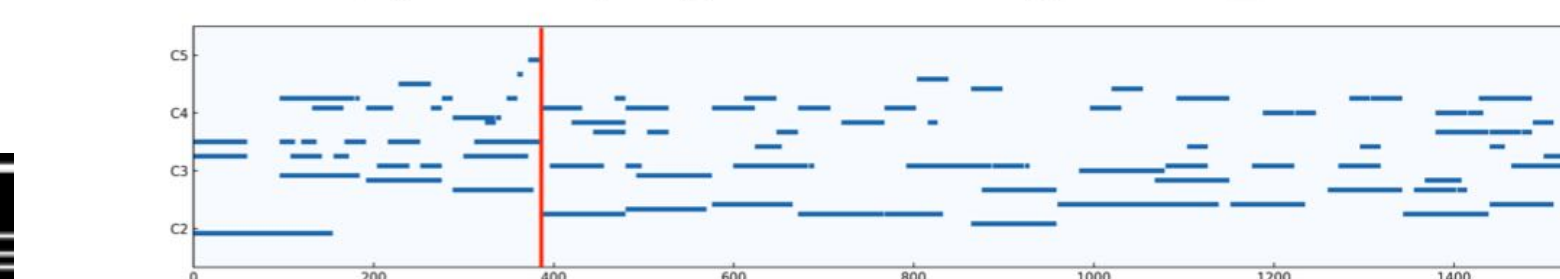c) The distribution of notes with different durations and the same pitch.
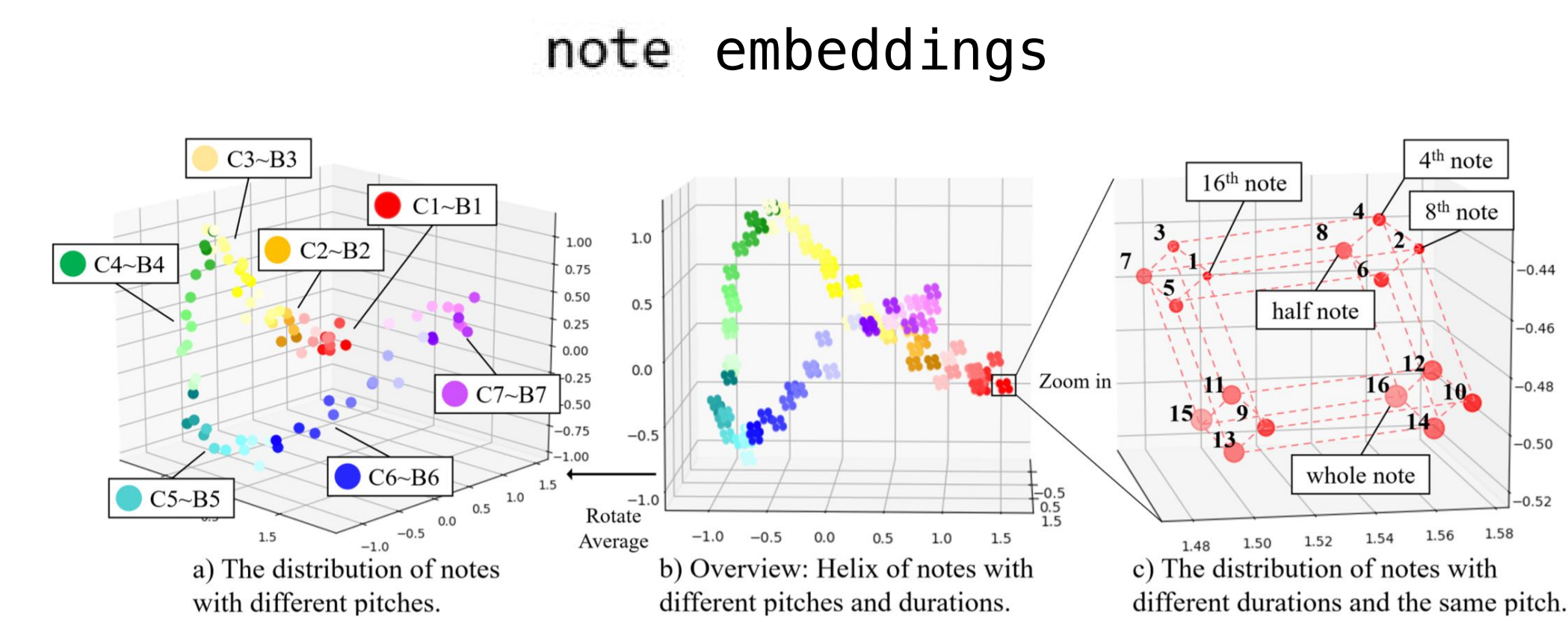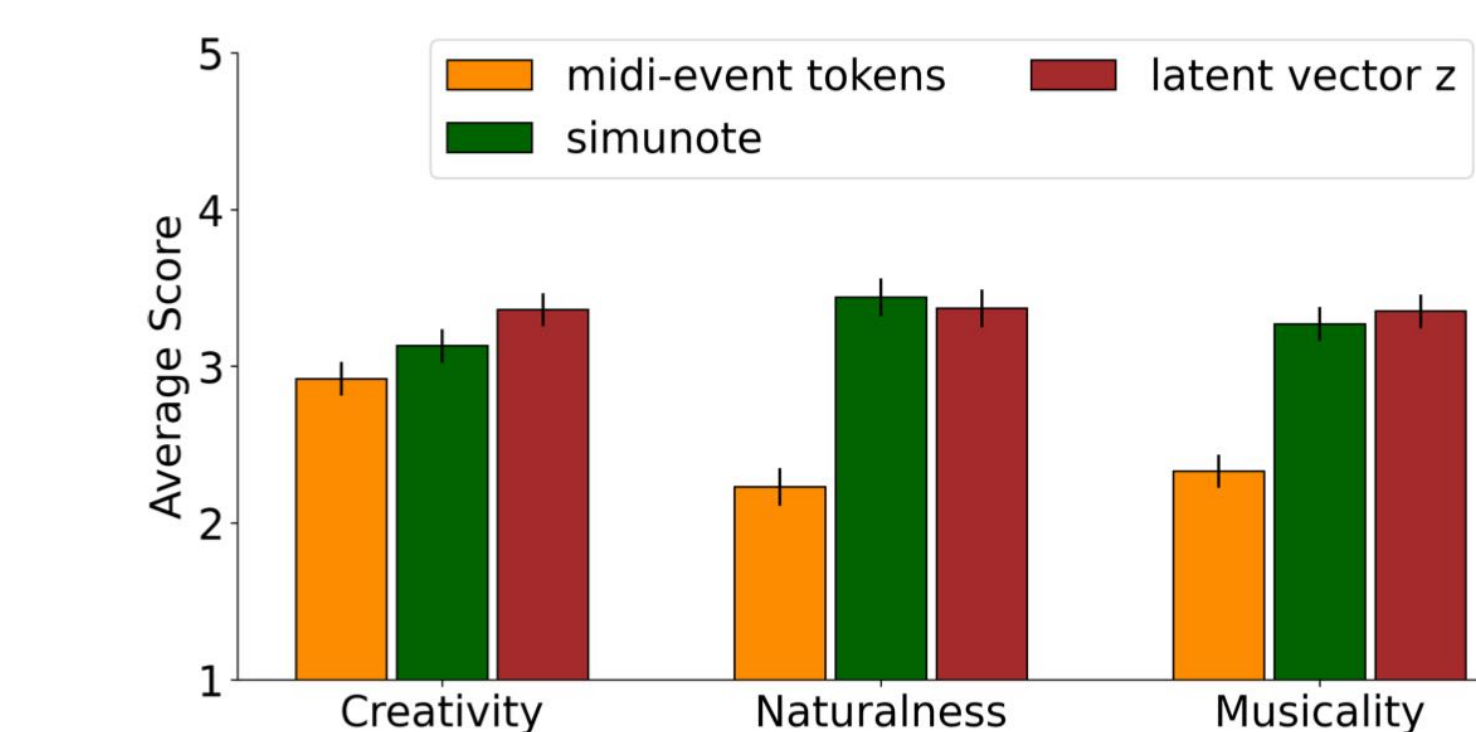
• Downstream Music Generation
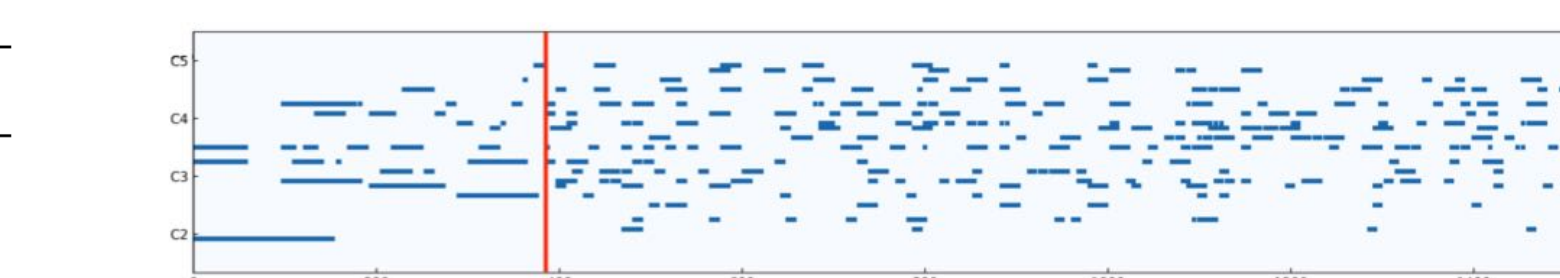
(a) A sample generated using midi-event tokens.

(b) A sample generated using `simu_note`.

(c) A sample generated using latent vector *z*.

`simu_note` embeddings

a) The relative relationship of different major chords.
b) Ring structure of chords with different degree scales in C major.