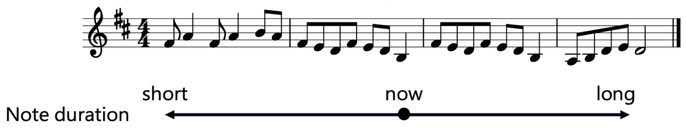
Attributes-Aware Deep Music Transformation

Lisa Kawai[1], Philippe Esling[2], Tatsuya Harada[1,3]
[1] The University of Tokyo, [2] IRCAM, [3] RIKEN

ISMIR MTL2020

Research Objective

Continuous control of musical attributes is important.
→ To transform music with a model that
1. can control **how much** it changes the musical attributes



2. is applicable to **any set of musical annotations**
→ No need to change the model structure or implementation for new attributes

Musical Attributes

Annotations from **jSymbolic** [McKay & Fujinaga, 2006]
- Calculates **musical statistics** (e.g. pitch, melody, chord, and rhythm)
- Selected 12 interpretable features, which are **continuous** values

Experimental Settings

Dataset: **Nottingham** dataset [Foxley, 2011] (monophonic)
Input representation: piano-roll with {rest, continue} tokens
Baselines: Naïve VAE (MusicVAE [Simon+, 2018]) and GLSR-VAE [Hadjeres+, 2017]
Network: Composed of Gated Recurrent Unit

Project Page and Contact

Other results with audio samples are available at →

Email: kawai@mi.t.u-tokyo.ac.jp

Method

- VAE-based model with attribute conditioning to decoder
- Adversarial discriminator to force $\boldsymbol{z}$ to abstract from $\boldsymbol{a}$
- Usually $\boldsymbol{a}$ is **binary** [Lample+, 2017] $(\boldsymbol{a}, 1-\boldsymbol{a} \in \{0,1\})$

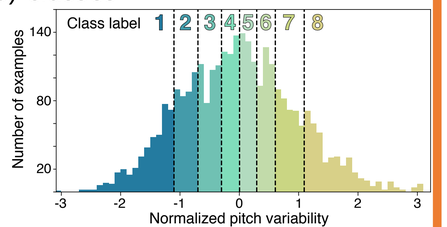
⟷ We have **<u>continuous annotations</u>** of attributes which are needed to achieve continuous control.

How did we solve this problem?

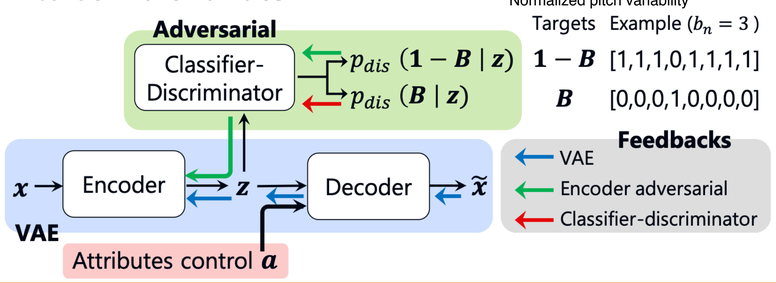
1. **Quantize** data into K(=8) classes

e.g.

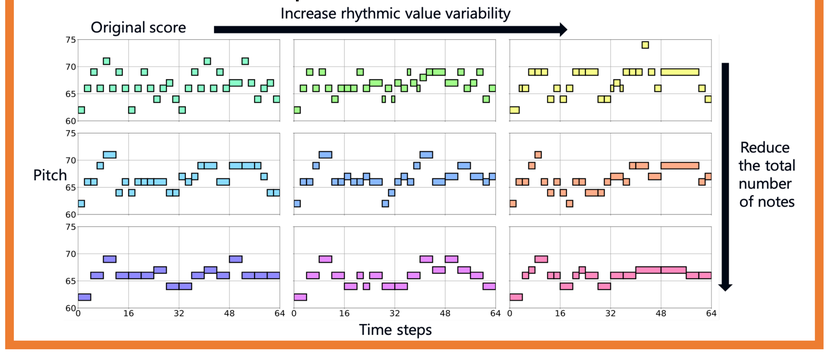
$\boldsymbol{a}$		$\boldsymbol{b}$
-1.0	→	2
-0.5	→	3
1.0	→	7



2. Extend the discriminator to be **multivariate**

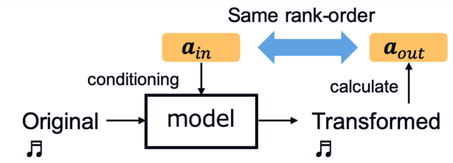


Generated Samples



Results

Spearman's rank-order **correlation coefficient**



attribute	Naive	GLSR	Ours
total number of notes	0.973	0.975	**0.981**
pitch variability	0.807	-	**0.938**
rhythmic value variability	0.830	-	**0.938**
pitch kurtosis	0.528	-	**0.723**
pitch skewness	0.366	-	**0.492**
prevalence of most common rhythmic value	0.795	-	**0.851**
average note duration	0.968	-	**0.983**
note density variability	0.677	-	**0.855**
amount of arpeggiation	0.126	-	**0.386**
chromatic motion	0.284	-	**0.622**
direction of melodic motion	0.428	-	**0.702**
average interval spanned by melodic arcs	0.262	-	**0.523**
chord duration	0.920	-	**0.944**

Our model outperforms the others on all the attributes.

Reconstruction accuracy

Our proposed model does not deteriorate the reconstruction quality.

	Naive	GLSR	Ours
NLL	2.269	**1.002**	1.679
accuracy	0.759	**0.808**	0.790

Chord consistency: Cosine similarity of <u>chroma feature</u>

12-dim vector that shows duration ratio of pitch classes

Our method outperforms the baselines.
GLSR-VAE is better for $\delta \approx 0$, but the cosine similarity significantly drops for $\delta > 0$.