ScorePerformer: Expressive Piano Performance Rendering with Fine-Grained Control

Ilya Borovik[1] and Vladimir Viro[2]
[1]Skolkovo Institute of Science and Technology, Russia, ilya.borovik@skoltech.ru
[2]Peachnote GmbH, Germany, vladimir@peachnote.de

ISMIR 2023

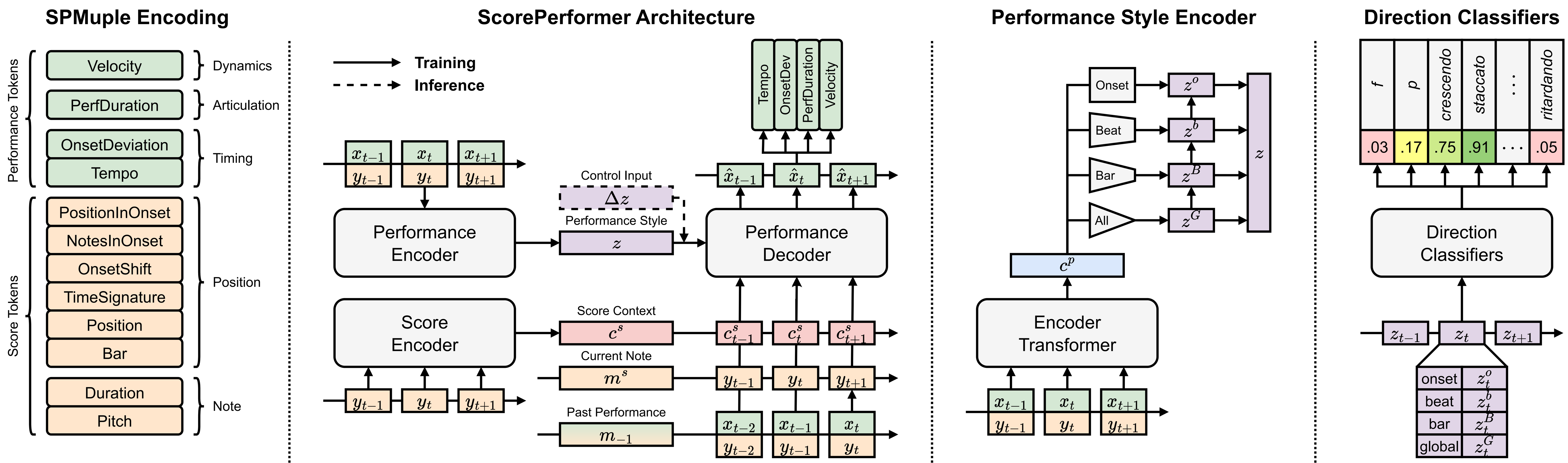
Motivation

1. effective control of musical instruments often requires considerable expertise, instrument training, and physical ability
2. develop a model that allows to perform musical works interactively using an intuitive interface, e.g. with musical or natural language-based control

Solution

ScorePerformer, a controllable piano performance rendering deep learning model that:

1. combines transformers and hierarchical MMD-VAE style encoding heads for encoding performance styles at the global, bar, beat, and onset levels
2. provides musical language-driven manipulation over the learned performance style space through a set of trained style embedding to performance direction classifiers
3. utilizes a tokenized encoding for aligned score and performance music (SPMuple) with a smooth and efficient local window tempo computation function



Latent Style Space

1. "Label" the latent style space using the performance direction classifiers
2. Compute per-direction control vectors in the style space that move the performance towards the markings

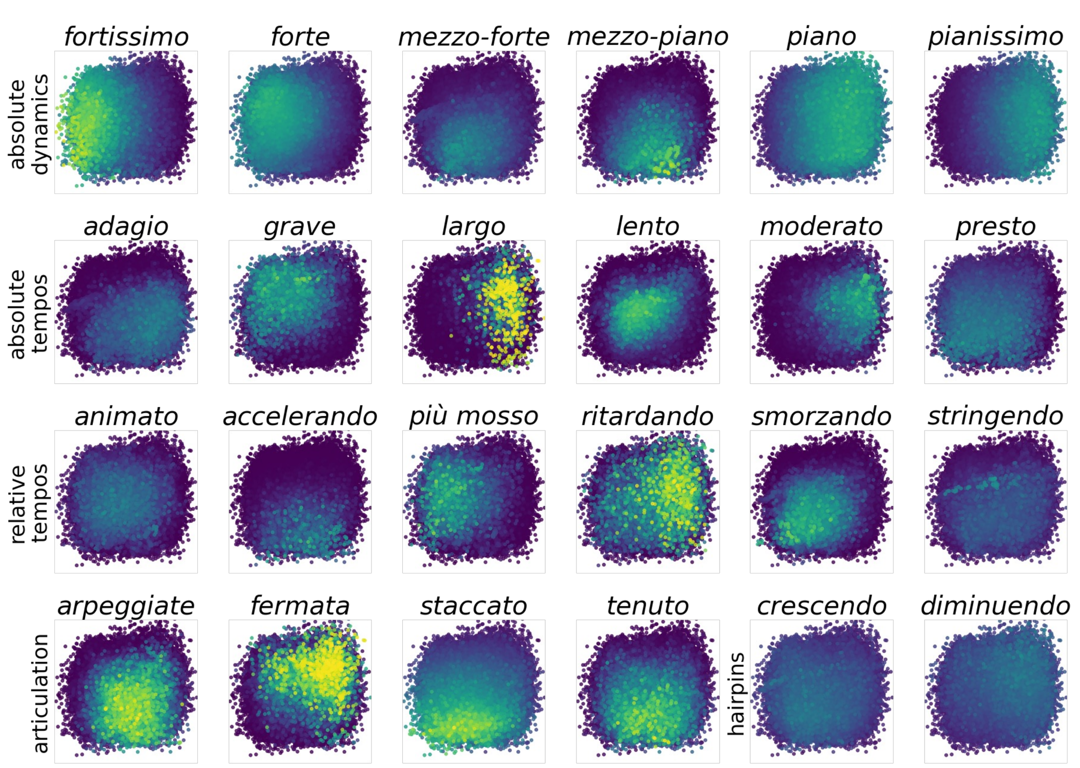


Figure 1. Two principal components of the style embeddings classified by the direction marking classifiers.

Performance Rendering Control

1. Sample random style or delta style embeddings
2. Use per-direction control embeddings, explicitly or by mapping natural language inputs:
   - "play softer from here" → "more piano"
   - "now gradually gain momentum" → "switch to accelerando"
   - "play notes with detached articulation" → "perform staccato"

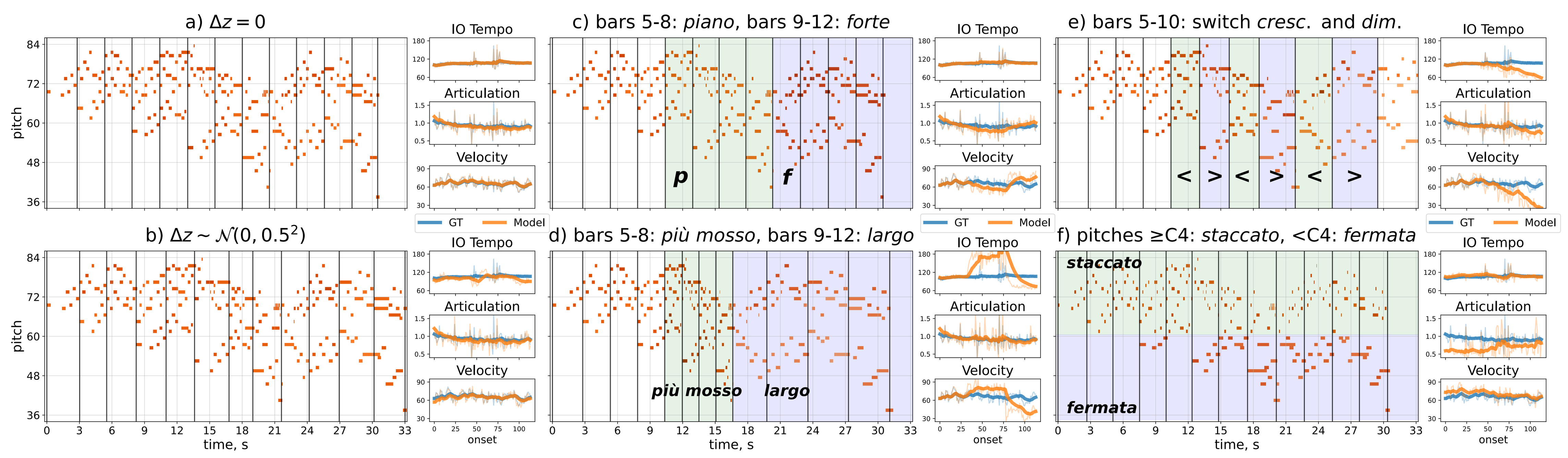


Figure 2. Pianorolls and performance features (inter-onset tempo, articulation, and velocity) for the first 12 musical bars of Bach's "Fugue No.19 in A major", rendered by ScorePerformer with unconditional or conditional style control. The title of each plot indicates the form of the control input. Colored areas highlight the regions with the applied control.

Tempo Functions

	Error ↓			Correlation ↑		
Tempo	IOI	OD	PD	IOI	OD	PD
Bar	0.140	0.012	0.063	0.650	0.361	0.837
Beat	0.116	0.009	0.066	0.727	0.406	0.854
Onset	0.124	0.011	0.056	0.709	0.339	0.890
Window	**0.090**	**0.008**	**0.048**	**0.901**	**0.538**	**0.907**

Table 1. Evaluation of local tempo functions in SPMuple on performances generated with unaltered style embeddings. IOI – inter-onset interval, OD – onset deviation, PD – performed duration, Vel – velocity.

Latent Hierarchies

G	B	b	o	z	IOI	OD	PD	Vel
32	**20**	**8**	**4**	**64**	**0.901**	**0.538**	**0.907**	**0.943**
32	20	12	✗	64	0.464	0.194	0.739	0.861
32	32	✗	✗	64	0.417	0.067	0.722	0.812
64	✗	✗	✗	64	0.327	0.066	0.658	0.576
32	20	8	✗	60	0.410	0.065	0.764	0.847
32	20	✗	4	56	0.842	0.224	0.881	0.857
32	✗	8	4	44	0.863	0.386	0.886	0.913
✗	20	8	4	32	0.890	0.485	0.904	0.939

Table 2. Latent hierarchy combinations. Correlation between real and generated performances. G – global, B – bar, b – beat, o – onset, and z – total latent dimensions.

Ablation Study

	IOI	OD	PD	Vel
ScorePerformer	**0.901**	**0.538**	**0.907**	0.943
w/o Score Encoder	0.885	0.526	0.889	**0.951**
w/o input seq. m^s	0.844	0.422	0.895	0.925
w/o SALN	0.871	0.469	0.920	0.930
w/o in-out emb. tie	**0.901**	0.459	0.873	**0.951**
w/o Continuous Tokens	0.576	0.116	0.747	0.561

Table 3. Evaluation of model configurations using the correlation between ground truth and generated performances.