RhythGen: Controlling Rhythmic Complexity in Symbolic Music Generation for Serious Game Interventions

Utrecht University

Efraim Dahl, Peter van Kranenburg, Anja Volk

efraimcdahl@gmail.com, p.vankranenburg@uu.nl, a.volk@uu.nl

Aim

Finetune an LLM-based symbolic music generator for time varying control over syncopation.


Simple example (little syncopation)


Complex example (strong syncopation)

Motivation

Music Generation in Games: Automatic music generation introduce a large variety of music into a game.

With **time-varying** control, music can incorporate audible changes, to which a player reacts – application for Musical Attention Control Training MACT.

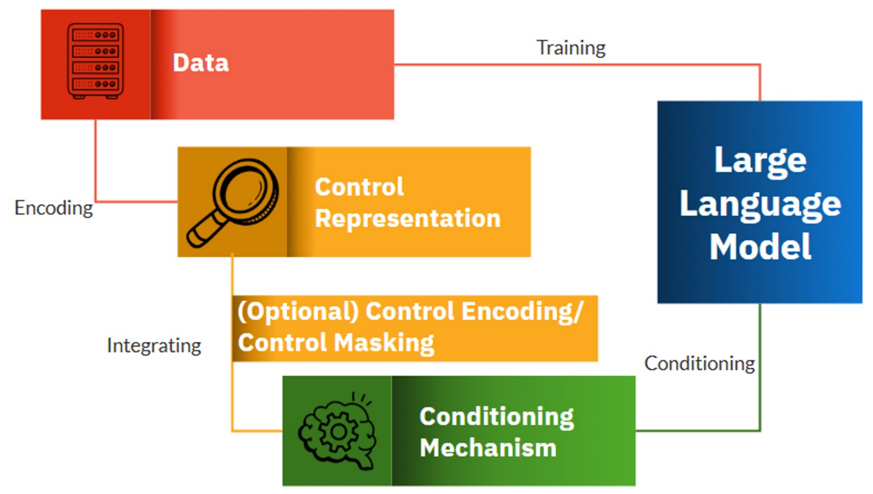
Control over **rhythmic complexity/syncopation** makes music more/less challenging for interaction.

Technical Considerations

Large Language Models (LLMs) form the state of the art for automatic music generation, but expensive to train.

Rhythmic complexity/syncopation is difficult to capture.

Method



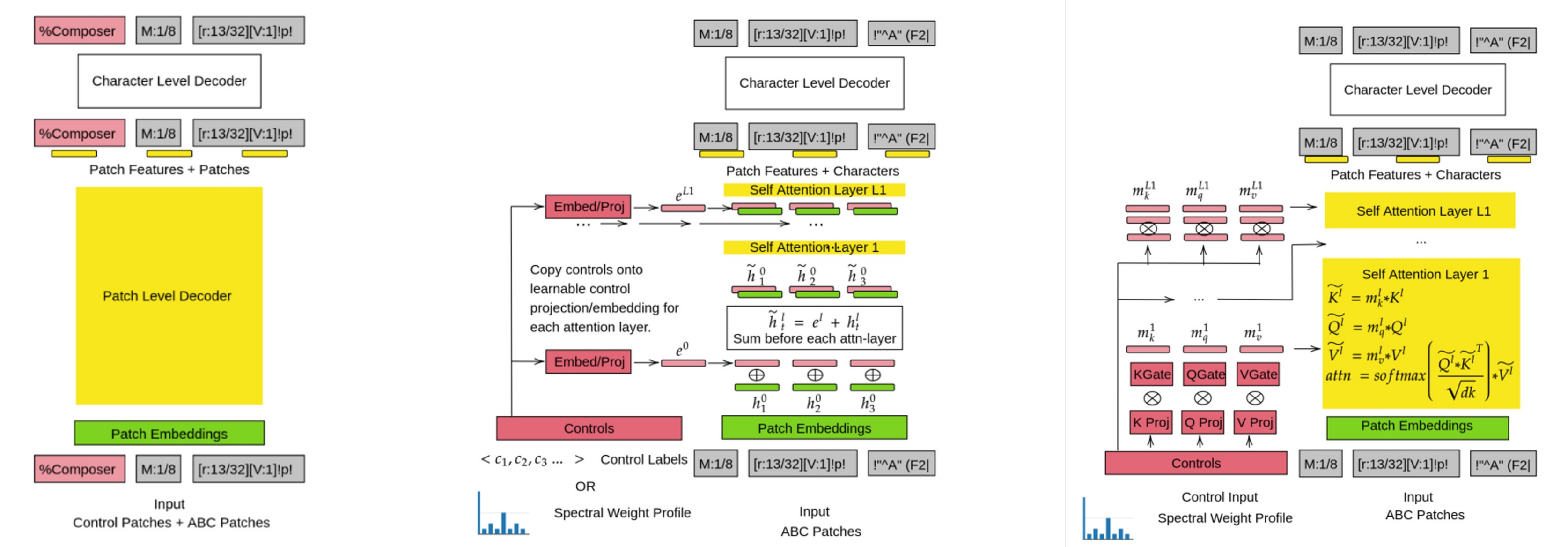
Setup: Fine-tune model on 1/3 datasets, conditioned with ¼ control representations, integrated using 1/3 conditioning mechanisms.

Base Model: NotaGen – GPT2 based music generator pre-trained on 1.6 million songs in ABC notation.

Data: OpenLieder – 1300 19th century art songs.
RAG-Collection: 1100 Ragtime pieces, 1.400 right hand sections

Control Representations:
Note Density Labels – Good baseline for establishing control
Syncopation Labels – Most information dense, simple representation based on Bemman and Christensen (2024)
Spectral Weights – Expressive, complex, time-insensitive Inner Metric Analysis (IMA) based measure
Metric Weights – Expressive, complex, time-sensitive IMA based measure

Conditioning Mechanisms

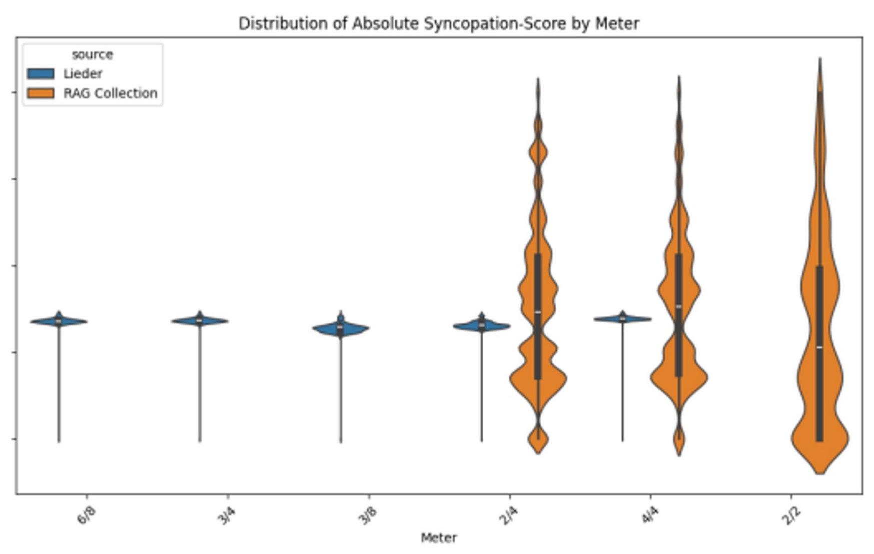


In-Text Conditioning | In-Attention Conditioning | Attention Modulation

Observations:
Very small range of syncopation in LiederDataset, no control possible.

Collapsing voices for rhythmic analysis obscures rhythmic structure/syncopation. Syncopation metric interacts with note-density.

Nuanced rhythmic structure (syncopation) harder to capture than coarse structure (note-density)



Results – Ablation Study

Time varying control over note density possible with note density labels+in-attention conditioning

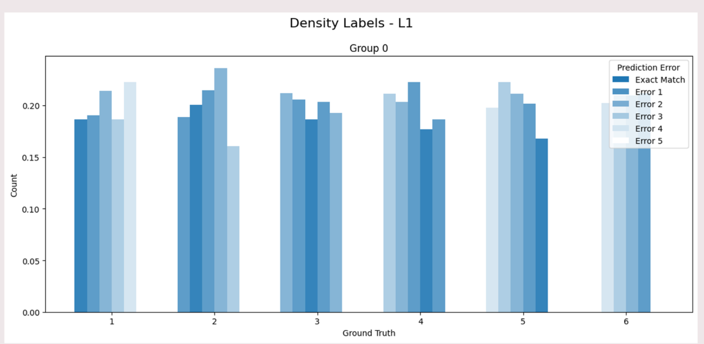
Experiment	Correlation ND	Correlation SY	KLD	ACS	BAE
ND-Labels	0.62	0.52	2.7	0.38	0.12
IMA-S	0.36	0.23	4.4	0.67	0.08
Baseline	0.17	0.41	3.7	0.70	0.06

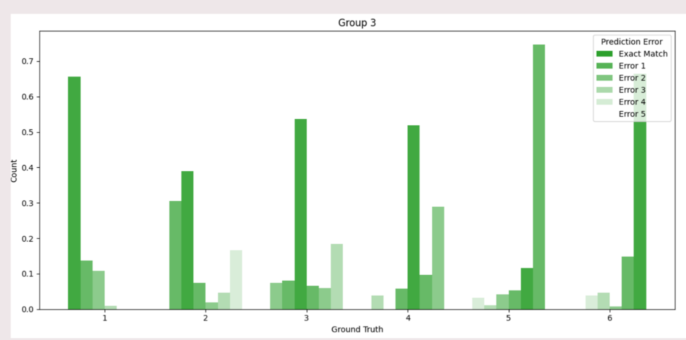
Control over note-density possible with spectral weights and attention modulation.

Experiment	Correlation ND	Correlation SY	KLD	ACS	BAE
IMA-S	0.44	0.42	2	0.68	0.2
IMA-M	0.24	0.41	3.5	0.71	0.18
Baseline	0.17	0.41	3.7	0.70	0.06

Time varying control over syncopation possible with syncopation labels, in-attention conditioning + voice masking

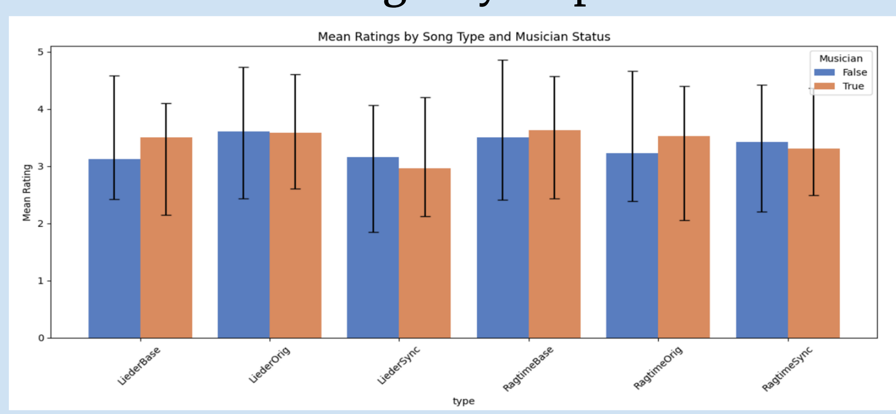
Experiment	Correlation ND	Correlation Syncopation	KLD	ACS	BAE
SY-Labels (RH)	0.32	0.32	2.6	0.73	x
SY-Labels (Masking)	0.71	0.25	2.8	0.77	0.35
Baseline	0.57	0.07	3.7	0.84	0.01

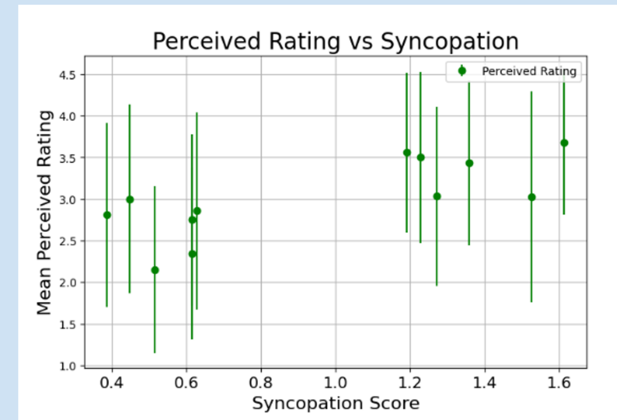

Uncontrolled labels (uniform distribution)


Controlled Labels (strong peaks visible)

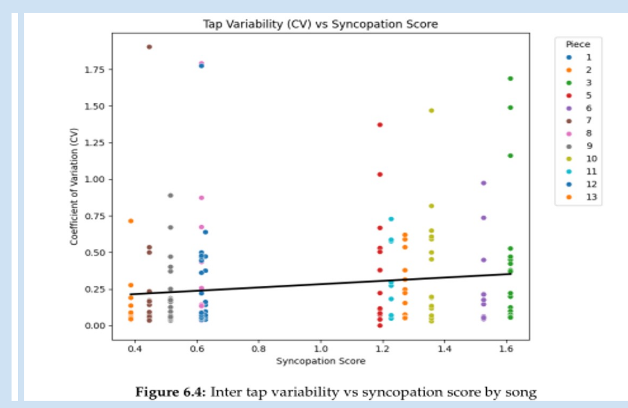
Results – Listener Study (n=40)

Listeners rating of rhythmic complexity and tapping regularity significantly correlates with target syncopation levels.





Listeners are moderately accurate (68%) at detecting boundaries between sections with distinct note-density/syncopation levels. Listeners report similar ratings of enjoyment for conditionally generated, baseline and human-composed music.



Listening Examples

Conclusion: Controlling syncopation requires voice-specific adaptations. Attention Modulation helps with sparse controls

Control fine-tuning with little data enabled through sequence aligned conditioning mechanism, in-attention conditioning/attention modulation.

Future Work: polyphonic syncopation metric, rhythmic complexity metric beyond syncopation, disentangling IMA-profiles into rhythmical attributes.