# A Visualization and Control Interface for Rhythmic Relations

Cagri Erdem
University of Oslo
Oslo, Norway
cagrie@ifi.uio.no

Davide Rocchesso
University of Milan
Italy, Italy
davide.rocchesso@unimi.it

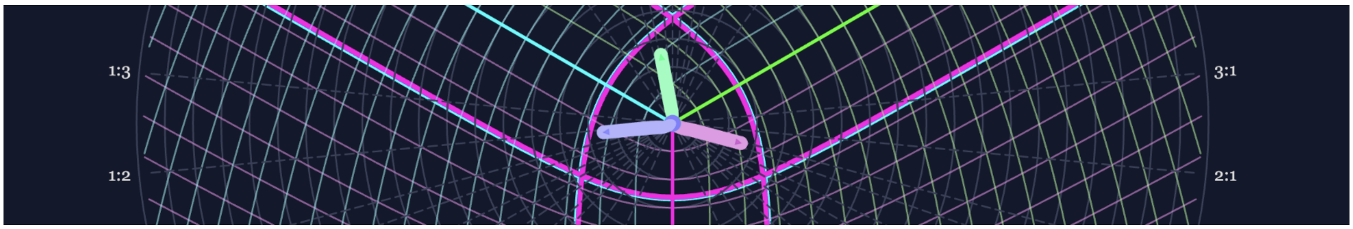


**Figure 1: DOT animation of overlapping relations between duplets, triplets, and quarter notes played by kick, snare, and hi-hat.**

## Abstract

RhyGlyph is a radial visualization system that represents rhythmic interaction as three pairwise spacetime trajectories within a compact three-wedge glyph, supporting both overview and detailed reading. Building on this representation, RhyDiff is a generative framework that learns aligned latent spaces for symbolic drum patterns and the encoded trajectory features through paired VQ-VAEs and FiLM-conditioned diffusion. Users edit relational trajectories and metadata to generate new drum patterns by treating visualization itself as a control surface.

## CCS Concepts

• **Human-centered computing** → **Visualization systems and tools**; • **Applied computing** → *Sound and music computing*.

## Keywords

rhythm visualization, radial glyph, cross-modal, diffusion

**ACM Reference Format:**
Cagri Erdem and Davide Rocchesso. 2026. A Visualization and Control Interface for Rhythmic Relations. In *Proceedings of the 2026 International Conference on Advanced Visual Interfaces (AVI '26), June 08–12, 2026, Venice, Italy*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3811427.3811503

## 1 Introduction

Most rhythm visualizations depict single-stream inter-onset intervals (IOIs). RhyGlyph [1] foregrounds *relations*: a three-voice drum segment (kick, snare, hi-hat) is compressed into a radial glyph with three 120° wedges, one per instrument pair, where overlapping rhythms generate spacetime trajectories [6] as in Figure 1. Radial glyphs are widely used for spatiotemporal dynamics [3]; unlike linear-timeline or cycling-circular [4] depictions, RhyGlyph maps each instrument pair to a trajectory within one tri-axial glyph readable at a few degrees of field of view [11]. We pair RhyGlyph with RhyDiff, a cross-modal generative extension that treats spatially edited Discrete Onset Trajectories (DOT) and controllable metadata as conditioning signals for diffusion-based symbolic drum generation.


*AVI '26, Venice, Italy*

ACM ISBN 979-8-4007-2342-1/26/06
https://doi.org/10.1145/3811427.3811503

## 2 RhyGlyph

Rhythm is treated as motion: overlapping rhythms act as velocity components of a moving point, yielding trajectories whose direction reflects relative pace (shorter IOI ⇒ faster progress along that axis). A tri-axial layout defines three 120° wedges for all pairwise interactions. The prototype accepts MIDI drum patterns, parses onsets, and computes two complementary relational encodings described below.

### 2.1 METER Encoding

METER is a beat-quantized "push–pull" encoding: within each wedge, an imaginary needle moves between two instrument poles according to their relative pace. Onsets are converted to discrete beat time instants $b_k$. For each sounding object $i$, local strength is

$$s_i(b_k) = \frac{1}{\mathrm{IOI}_i(b_k) + \varepsilon},$$

which is set to gradually decay when expected onsets are missed. The relative pull between, e.g., Kick and Snare is

$$m_k = \frac{s_K(b_k) - s_S(b_k)}{s_K(b_k) + s_S(b_k) + \varepsilon} \in [-1, 1],$$

autoregressively smoothed via $\tilde{m}_k = \alpha\, m_k + (1-\alpha)\, \tilde{m}_{k-1}$, and mapped to a wedge angle

$$\theta_k = \theta_{\min} + \frac{\tilde{m}_k + 1}{2}\,(\theta_{\max} - \theta_{\min})\,.$$

Radial growth $r_k$ at time $b_k$ yields the spiral point $(r_k \cos\theta_k,\ r_k \sin\theta_k)$.

## 2.2 DOT Encoding

DOT (*Discrete Onset Trajectories*) is an a-metric encoding driven by IOIs in continuous time. The instantaneous velocity per instrument:

$$v(t) = \frac{S}{\mathrm{IOI}(t) + \varepsilon},$$

where $S$ is a scale constant. Three pairwise trajectories are constructed by discrete integration; e.g., for kick–snare:

$$x(t+\Delta t) = x(t) + v_K(t)\,\Delta t, \quad y(t+\Delta t) = y(t) + v_S(t)\,\Delta t.$$

Within a wedge, the trace leans toward the axis of the denser stream.

## 2.3 Projection onto Glyph-space

The orthogonal-space trajectory $p(t) = (x, y)$ is mapped into wedge geometry via a radial projection preserving magnitude $\rho = \sqrt{x^2 + y^2}$ and remapping angle $\varphi = \mathrm{atan2}(y, x)$ onto the wedge between directions $\alpha_1$ and $\alpha_2$:

$$\theta = \alpha_1 + \tfrac{\varphi}{90^\circ} \cdot \Delta(\alpha_2, \alpha_1).$$

The renderer overlays concentric circles, isotime grids, and dashed ratio reference lines for common integer ratios.

## 2.4 Interface and Control

The interface combines the glyph with controls for algorithm, duration in beats, tempo, zoom, background overlays, color theme, thickness and hue that encode velocity sensitivity, and synchronous beat animation and playback with optional OSC output (Figure 2).

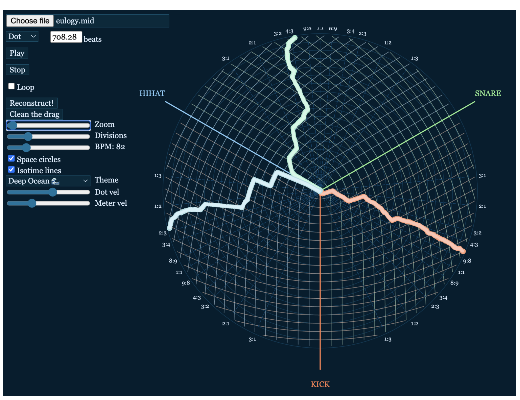


**Figure 2: Prototype interface visualizing DOT encodings of Tool's *Eulogy*.**

In DOT mode, users can select trajectory samples and drag them under wedge constraints (Figure 3) via moving-least-squares deformation [8]; edits propagate across affected dyads and can be reconstructed into onset sequences.

## 3 RhyDiff

RhyDiff extends RhyGlyph into a controllable generative interface: users edit relational cues (trajectories and metadata), and a cross-modal latent diffusion model generates bar-synchronous MIDI drum patterns.

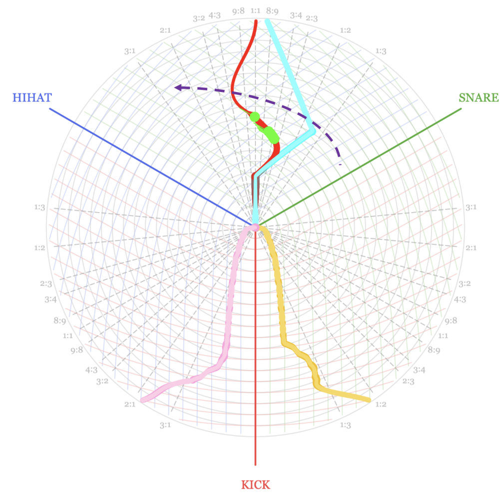


**Figure 3: Constraining algorithms let the user drag one pairwise trajectory, shown in red, while the system recalculates the other two. When a boundary is reached, RhyGlyph colors the affected segment in a distinct shade (neon green in this theme) and shows the direction of the constraint.**

## 3.1 Data Representation

Each bar is represented as paired token streams: (1) a symbolic MIDI stream tokenized in REMI+ vocabulary [7] via MidiTok [12], and (2) a DOT trajectory stream with quantized (`AngleBin`, `RadiusBin`) offsets following the Cursive Transformer scheme [2]. Metadata includes NI-grid polyrhythm cues [9] encoding non-isochronous beat structure as $N(i) = F(i) + S\big(T_{\mathrm{prox}}(i) - F(i)\big)$, where $F$ is the formative grid, $T_{\mathrm{prox}}$ the proximate target beat, and $S \in [0, 1]$ the shift.

## 3.2 Cross-modal Architecture

Paired VQ-VAEs [10] learn a shared bar-level latent space (1024 dims, $16 \times 64$ slices): a symbolic tower (MIDI→latent→MIDI) and a cross-modal tower (DOT→latent→MIDI). Training combines reconstruction and codebook losses:

$$\mathcal{L} = \lambda_{\mathrm{CE}}\mathcal{L}_{\mathrm{CE}} + \mathcal{L}_{\mathrm{code}} + \beta\,\mathcal{L}_{\mathrm{commit}}.$$

A Gaussian diffusion prior [13] operates in the continuous bar-level MIDI latent space. The denoiser is conditioned on DOT tokens and bar metadata (time signature, tempo, NI-grid cues, style/fill tags) via FiLM [5] per-channel scale/shift $(\gamma, \beta)$, trained with standard DDPM noise-prediction objective. Generated latents are decoded to yield bar-level MIDI output.

## 4 Conclusions

We present RhyGlyph, a radial visualization that renders rhythmic relations as trajectories open to direct manipulation, paired with RhyDiff, a trajectory-conditioned diffusion framework for symbolic drum generation. A user study evaluating the visualization is underway; future work will focus on multi-meter variable-length generation and more embodied forms of manipulation through handheld devices.

## References


[1] Cagri Erdem, Carsten Griwodz, and Davide Rocchesso. 2025. RhyGlyph: Radial Glyph Visualization of Rhythm Interactions. In *Proceedings of the 20th International Audio Mostly Conference (AM '25)*. Association for Computing Machinery, New York, NY, USA, 408–416. doi:10.1145/3771594.3771643

[2] Sam Greydanus and Zachary Wimpee. 2025. The Cursive Transformer. doi:10.48550/arXiv.2504.00051 arXiv:2504.00051 [cs].

[3] Marcel Hlawatsch, Philipp Leube, Wolfgang Nowak, and Daniel Weiskopf. 2011. Flow Radar Glyphs—Static Visualization of Unsteady Flow with Uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 1949–1958. doi:10.1109/TVCG.2011.203 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[4] Paul Lascabettes. 2025. Generating non-periodic pitch sequences inspired by the tiling process developed by Tom Johnson. *Journal of Mathematics and Music* 20, 1 (2025), 68–76. doi:10.1080/17459737.2025.2577447 _eprint: https://doi.org/10.1080/17459737.2025.2577447.

[5] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. 2017. FiLM: Visual Reasoning with a General Conditioning Layer. doi:10.48550/arXiv.1709.07871 arXiv:1709.07871 [cs].

[6] Davide Rocchesso, Alessio Bellino, Gabriele Ferrara, and Antonino Perez. 2024. Spacetime trajectories as overlapping rhythms. *International Journal of Human-Computer Studies* 192 (Dec. 2024), 103358. doi:10.1016/j.ijhcs.2024.103358

[7] Dimitri von Rütte, Luca Biggio, Yannic Kilcher, and Thomas Hofmann. 2024. FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control. doi:10.48550/arXiv.2201.10936 arXiv:2201.10936 [cs].

[8] Scott Schaefer, Travis McPhail, and Joe Warren. 2006. Image deformation using moving least squares. *ACM Trans. Graph.* 25, 3 (July 2006), 533–540. doi:10.1145/1141911.1141920

[9] George Sioros and Odysseas Klissouras. 2025. Flow Swing: a system for dynamic control and exploration of non-isochronous timing in musical rhythms. In *Proceedings of the 22nd Sound and Music Computing Conference (SMC2025), Graz, July 2025*. Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Graz, Austria, 152–159. doi:10.5281/zenodo.15843562 Conference Name: Sound and Music Computing 2025 (SMC 2025).

[10] Thomas Theodoridis, Theocharis Chatzis, Vassilios Solachidis, Kosmas Dimitropoulos, and Petros Daras. 2020. Cross-modal Variational Alignment of Latent Spaces. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 4127–4136. doi:10.1109/CVPRW50498.2020.00488

[11] Colin Ware. 2021. Static and Moving Patterns. In *Information Visualization*. Elsevier, NY, USA, 183–243. doi:10.1016/B978-0-12-812875-6.00006-2

[12] Michał Wiszenko, Kacper Stefański, Piotr Malesa, Łukasz Pokorzyński, and Mateusz Modrzejewski. 2024. MidiTok Visualizer: a tool for visualization and analysis of tokenized MIDI symbolic music. doi:10.48550/arXiv.2410.20518 arXiv:2410.20518 [cs].

[13] Jincheng Zhang, György Fazekas, and Charalampos Saitis. 2024. Composer Style-specific Symbolic Music Generation Using Vector Quantized Discrete Diffusion Models. doi:10.48550/arXiv.2310.14044 arXiv:2310.14044 [cs].