EVxRAVE: Incorporating Neural Synthesis in an Augmented String Instrument Platform

Brian Lindgren
Composition and Computer Technologies (CCT)
Department of Music
College of Arts and Sciences
University of Virginia
wdt7wr@virginia.edu

1 Overview

The EV (Lindgren, 2025a) is an augmented bowed string instrument that incorporates neural synthesis through IRCAM's RAVE variational autoencoder (Caillon and Esling, 2021). The aim is to explore the synthesis potential of the instrument's own electro-acoustic sound, but to interface with that potential in a new way. Rather than synthesis being shaped solely by frequency and amplitude envelopes, as is traditional with most instruments, the EV supports both this conventional approach and the ability to navigate the latent space of a model trained on its recordings. Instrumental features such as fundamental frequency (f_0), amplitude, and spectral centroid are mapped to the model's latent dimensions, extending string techniques into new forms of timbral exploration. This is demonstrated in an accompanying performance video, *two tales from the shadows of the grid* (`https://youtu.be/aVPE8Ac6XU8`).

Contributions. (1) A self-referential RAVE integration trained on the EV's corpus, exposing string-idiomatic descriptors (f_0, amplitude, spectral centroid) as controls for latent navigation; (2) a control strategy that reveals the relationship between performance gesture and latent response by allowing one modulation source to be mapped to one latent dimension at a time (e.g., centroid$\rightarrow z_1$ only).

2 The EV Platform

While keyboard interfaces are common, purpose-built string instruments that extend traditional gestural vocabularies into programmable domains remain rare. The EV asks what sonic possibilities might emerge from an instrument that interfaces with a computer yet retains its acoustic strings as sources of resonance and embodied gesture.

Early iterations explored convolution of the acoustic signal with a synthesizer, later expanded with ambisonic encoding to project the instrument across multichannel arrays Lindgren (2025b). The addition of physical models and reverberation further merged digital and acoustic domains into a unified ecosystem. The current Pure Data implementation Lindgren (2025c) gives each string an independent engine comprising multiple synthesizers, FFT convolution, grain delay, reverb, a physical model, Paulstretch freeze, and ambisonic panning. Up to eight voices per module support dense timbres, while a control matrix flexibly maps seven sources to twenty destinations; in practice, a small subset is reserved for descriptor$\rightarrow$latent routing to maintain clarity. The EV's 3D-printed body houses four infrared optical pickups (one per string) whose signals are digitized by a Bela board and transmitted to a laptop running Pure Data.

Beyond technical design, the EV also asks how older musical paradigms can be transposed into hybrid analog–digital worlds. By extending the bowed string into convolution, ambisonics, and machine learning, it reframes traditional practice within a technologically mediated environment.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Music.

3 Related Work: Instruments Integrating RAVE

Since 2021, RAVE has been adopted in a range of new musical interfaces exploring performance integration and latent-space control. Sophtar (Visi, 2024) is a tabletop, guitar-like instrument with onboard computing and sensor input; its model, trained on vocal chants, takes feedback audio as input while fretboard pressure and thumb position navigate the latent space. Stacco (Privato et al., 2024) investigates intuitive control through magnetic objects that interact with attractors defining a field for latent navigation.

A closely related self-referential approach is Shepardson and Magnusson's "Living Looper" (Shepardson and Magnusson, 2023), which trains RAVE on the performer's own recordings to enable timbral re-articulation. The EV differs in its idiomatic gestural substrate (continuous bowed-string control rather than loop layering), its descriptor-to-latent mappings grounded in acoustic performance features (f_0, amplitude, centroid), and its integration within a broader DSP ecosystem (physical modeling, convolution, ambisonics).

4 Integrating RAVE

The RAVE variational autoencoder was selected for integration into the EV for its low latency, ability to operate at standard audio rates, and efficient CPU performance. Four hours of EV recordings made over the past three years were used for training at 44.1 kHz. A lightweight eight-dimensional architecture, originally designed for the Raspberry Pi, enabled reliable operation alongside other DSP tasks and required roughly 15% less processing power than the newer "v2" model.

Within the EV, the model is embedded in the Pure Data–based interface. All eight latent dimensions are available for direct control, while the first four can be mapped to gestural inputs through the control-voltage matrix. These mappings allow features such as spectral centroid, amplitude, and pitch to shape the latent space in real time, extending the instrument's acoustic expressivity into a neural domain. In comparative listening, the convolution and physical-model engines behave predictably, while RAVE produces continuously evolving, descriptor-driven timbral shifts that reveal its distinctive role within the system. The system remains qualitative and exploratory, with mappings tuned for interpretability rather than generalization.

5 Demo and Usage of RAVE

The accompanying video, *two tales from the shadows of the grid*, shows how the EV's gestural interface and RAVE integration combine manual and automated latent-space control to yield contrasting modes of expression.

In the first movement, two LFOs drive the first two latent dimensions, creating a calm, repeating motif whose intensity is shaped in real time. A melody on the A string modulates a third dimension, while the spectral centroid maps to a fourth, coupling pitched gesture to timbral change. Intermittent use of a hand-drawn curve on the third dimension interrupts the LFO regularity, balancing automation with embodied intervention.

The second movement inverts this logic. The model begins in a quasi-stable state that sustains a bass texture. A melody on the C string induces undulation in this layer, while notes on the other strings return it to stasis, redirecting attention back to the C string. Here, pitch, amplitude, and spectral centroid map to the second through fourth dimensions, and gradual activation of the remaining dimensions thickens the spectrum, driving a climactic, timbrally saturated cadence.

6 Conclusion

The EV shows how AI tools such as RAVE can be embedded into traditional instruments, extending expression beyond pitch and amplitude and making performance descriptors direct agents of timbral transformation. By coupling bowed-string features to specific latent dimensions and demonstrating distinct gesture-to-latent relationships, the work clarifies RAVE's role within a hybrid instrument ecology. While qualitative and self-referential in scope, the approach points toward reproducible design patterns for integrating neural synthesis with embodied acoustic gesture.

References

Caillon, Antoine, and Philippe Esling. "RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis." *CoRR*, abs/2111.05011, 2021. `https://arxiv.org/abs/2111.05011`.

Lindgren, Brian. "The EV: An Iterative Journey in Digital–Acoustic String Instrument Augmentation." *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2025, pp. 379–388. The Australian National University, Ngunnawal and Ngambri Country, Canberra, Australia. `https://doi.org/10.5281/zenodo.15698898`.

Lindgren, Brian. "Exploration of Spatial Composition with a New Electronic Stringed Instrument." *Organised Sound*, vol. 30, no. 2, 2025. `https://doi.org/10.1017/S1355771825100575`.

Lindgren, Brian. "An Augmented String Instrument Architecture Created With Pure Data." *Proceedings of the PdMaxCon25* [University of Illinois Urbana–Champaign], 2025, pp. 41–47. `https://doi.org/10.5281/zenodo.17537234`.

Lindgren, Brian. "two tales from the shadows of the grid." *YouTube*, 2025. `https://youtu.be/aVPE8Ac6XU8`.

Privato, Nicola, Victor Shepardson, Giacomo Lepri, and Thor Magnusson. "Stacco: Exploring the Embodied Perception of Latent Representations in Neural Synthesis." *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2024. `https://zenodo.org/doi/10.5281/zenodo.13904898`.

Shepardson, Victor, and Thor Magnusson. "The Living Looper: Self-Referential Neural Synthesis in Live Performance." *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2023. `https://iil.is/pdf/2023_nime_shepardson_magnusson_living_looper.pdf`.

Visi, Federico. "The Sophtar: a Networkable Feedback String Instrument with Embedded Machine Learning." *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2024. `https://zenodo.org/doi/10.5281/zenodo.13904810`.