Expanding Creative Interaction: Toward a Modular Toolkit for Audio-Visual Composition and Augmented Performance

Anna Savery
University of Technology Sydney
Sydney, Australia
anna.savery@uts.edu.au

Sam Ferguson
University of Technology Sydney
Sydney, Australia
Samuel.Ferguson@uts.edu.au

Andrew Johnston
University of Technology Sydney
Sydney, Australia
Andrew.Johnston@uts.edu.au

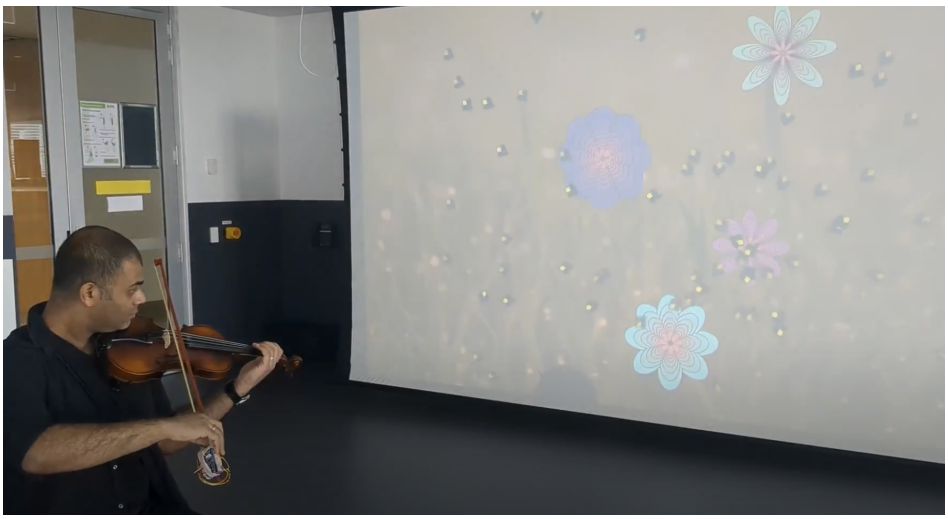


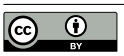
Figure 1: Participant 3, interacting with the system; bees ecosystem visuals responding to bow gestures.

Abstract

We present a modular, sensor-based violin bow interface designed to augment musical performance and audiovisual composition. The system integrates real-time audio processing, visual feedback, and machine learning models for gesture recognition. Through an expert user study with seven violinists, we identified themes of ergonomic adaptation, co-agency, and creative potential. While results highlight the toolkit's promise as both instrument and collaborator, limitations include the absence of non-expert perspectives, training protocols, and audience evaluation. This work contributes to human–agent interaction by positioning augmented instruments as sites of improvisational dialogue and outlines future directions in unifying system components, quantitative evaluation, and broader user studies.

CCS Concepts

• **Computer systems organization** → Real-time system architecture; • **Hardware** → **Emerging tools and methodologies**; • **Human-centered computing** → *User interface toolkits.*



This work is licensed under a Creative Commons Attribution 4.0 International License.
HAI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2178-6/25/11
https://doi.org/10.1145/3765766.3765845

Keywords

music, augmented instruments, multimodal interaction, embodied interaction, machine learning, musical interface design, human-computer interaction

ACM Reference Format:
Anna Savery, Sam Ferguson, and Andrew Johnston. 2025. Expanding Creative Interaction: Toward a Modular Toolkit for Audio-Visual Composition and Augmented Performance. In *13th International Conference on Human-Agent Interaction (HAI '25), November 10–13, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3765766.3765845

1 Introduction

Creative interaction across sound and visuals remains challenging for performers working with data-driven systems. We address this through a minimally invasive, modular bow-mounted interface for violinists that enables new forms of embodied musical control.

2 Related Work

Research on augmented instruments embeds sensing into performance, from Anderson's tape-bow violin [7] and Machover's Hyperviolin [17] to Kimura's wearable systems [1, 10]. Recent work emphasizes modularity and ergonomics, with examples ranging from embedded systems such as TRAVIS II [11] to detachable solutions like the Bela Trill Bar. Overholt's framework [14] provides benchmarks for evaluating these designs in terms of expressive range, gestural intuitiveness, and mapping richness.

In parallel, mapping strategies have evolved from one-to-one gesture–sound links [8] to multi-layer structures [13] and machine-learned models [5]. Gesture classification work on the Hyperbow [18] and Augmented Violin [2] demonstrated the value of performer-informed mappings, while Tanaka [16] emphasized perceptual transparency as key to audience engagement. Building on this, our system combines direct mappings with adaptive machine learning via Wekinator [6], supporting emergent control structures. Within the broader lineage of computational co-performance—exemplified by Lewis's *Voyager* [12] and interactive improvisation systems [4, 9]—we contribute an "ecosystem mode" in which the system responds to performer gestures with adaptive density and memory, foregrounding co-agency and embodied dialogue in live audiovisual contexts.

3 System Overview

A wireless bow interface (pressure sensor, slider, buttons, accelerometer, RGB LED) streams data via Arduino/ZigBee to Max/MSP (audio), Processing (visuals), and Wekinator (gesture recognition) over Open Sound Control (OSC)[15].

Interaction is structured around three modes, together forming what we term the *ecosystem mode*: a performance framework where the system listens, adapts, and develops responses over time. **Mode 1** offers direct control over mapped audio and visual features while recording short audio phrases for later use. **Mode 2** introduces semi-autonomous improvisation, where the system responds unpredictably but responsively to performer gestures. **Mode 3** employs stored material and gesture classifications to generate autonomous responses, framing the system as a creative partner rather than a passive tool. These modes provided a spectrum of control, feedback, and co-agency explored in the study.

While effective for experimentation, the current architecture spans multiple software platforms, creating challenges for stability in performance. A key strand of ongoing work is the consolidation of these modules into a unified environment. All patches and models are openly available in a public GitHub repository.[1]

4 Expert User Study

Seven violinists (P1–P7) with classical/experimental backgrounds completed 60–90 minute sessions (free exploration, structured tasks, improvisation in *Ecosystem Mode*) and a semi-structured interview. Data were transcribed and thematically analysed following Braun and Clarke [3], supported by observations and interaction logs; UTS HREC approval ETH23-8664.[2]

5 Findings

Thematic analysis revealed four main insights. First, participants reported **embodied disruption**, as the additional sensors altered grip and posture, requiring ergonomic adaptation. Second, experiences of **co-agency** were most evident in Modes 2–3, where semi-autonomous and generative behaviours emerged; in contrast, Mode 1 was experienced more as a responsive tool, while later modes prompted descriptions such as "moody" or conversational. Third, **visual feedback** in the form of bee/flower metaphors helped

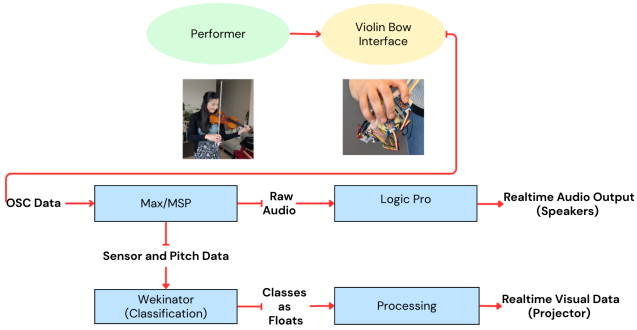


Figure 2: System Overview

structure interaction and enhance timing and narrative flow. Finally, participants highlighted the **educational potential** of the interface, suggesting applications for gesture-based learning, particularly for students less reliant on notation.

These findings highlight both the promise and the challenges of positioning augmented instruments as creative partners. While experts adapted quickly, limitations included initial ergonomic disruption, a lack of training protocols, and no external audience perspective.

6 Discussion and Future Work

Although mid-fidelity, the bow interface was quickly adopted by participants, who highlighted ergonomic disruption, mapping clarity issues, and bow-speed latency. These point to a need for lighter hardware, clearer affordances, and tighter synchronisation across gesture, audio, and visuals. Several performers also sought a stronger system "voice" and more nuanced mappings (e.g., microtonal control, extended techniques).

Co-agency was most productive when framed not as human-like reciprocity but as adapting to computational tendencies emerging from probabilistic mappings and gesture classifications. Performers who adjusted to these patterns described the system as conversational, while those expecting immediate human-like responses found it opaque. Transparent cues, onboarding, and performer-driven calibration can help bridge this gap. Visuals supported interaction and narrative flow, though their audience legibility remains a concern.

Future work will focus on developing a lighter, modular, and fully customisable interface, and on exploring avenues to consolidate the current multi-platform architecture into a unified environment. The insights gained from this study will inform both the physical redesign and the refinement of virtual systems, positioning the augmented bow not only as a technical prototype but as a platform for evolving forms of human–machine creativity.

Acknowledgments

This research is supported by an Australian Government Research Training Program Scholarship

[1] https://github.com/annushka123/expert_user_study_uts
[2] A video highlight reel is available at: https://zenodo.org/records/15575669

References

[1] Jiryis Ballan. 2024. *Capturing the Buzuq's Sonic Landscape: An Expedition Through Synthesis Processes from Soundboard to Digital Music Interface.* Ph. D. Dissertation. University of California, Irvine.

[2] Frédéric Bevilacqua, Nicolas H. Rasamimanana, Emmanuel Fléty, Serge Lemouton, and Florence Baschet. 2006. The augmented violin project: research, composition and performance report. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

[3] Virginia Braun and Victoria Clarke. 2022. *Thematic Analysis: A Practical Guide* (1st ed.). SAGE Publications Ltd, London, UK.

[4] Christopher Dobrian and Daniel Koppelman. 2006. The E in NIME: Musical Expression with New Computer Interfaces. *Proceedings of the International Conference on New Interfaces for Musical Expression* (2006). ISBN: 2844263143.

[5] Rebecca Fiebrink and Perry R Cook. 2011. The Wekinator: A system for real-time, interactive machine learning in music. In *Music Information Retrieval Conference.*

[6] Rebecca Fiebrink and Laetitia Sonami. 2020. Reflections on eight years of instrument creation with machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

[7] Mel Gordon. 1980. Laurie Anderson: Performance Artist. *The Drama Review* 24, 2 (1980). doi:10.2307/1145281

[8] Andy Hunt, Marcelo M Wanderley, and Ross Kirk. 2000. Towards a model for instrumental mapping in expert musical interaction. In *International Computer Music Conference.*

[9] Mari Kimura. 2003. Creative process and performance practice of interactive computer music: A performer's tale. *Organised Sound* 8, 3 (2003). Publisher: Cambridge University Press.

[10] Mari Kimura, Nicolas Rasamimanana, Frédéric Bevilacqua, Bruno Zamborlin, Norbert Schnell, and Emmanuel Fléty. 2012. Extracting human expression for interactive composition with the augmented violin. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

[11] Chantelle Ko and Lora Oehlberg. 2020. Construction and Performance Applications of an Augmented Violin: TRAVIS II. *Computer Music Journal* 44, 2-3 (2020).

[12] George E Lewis. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* (2000), 33–39.

[13] Ali Momeni and Cyrille Henry. 2006. Dynamic independent mapping layers for concurrent control of audio and video synthesis. *Computer Music Journal* 30, 1 (2006).

[14] Dan Overholt. 2009. The musical interface technology design space. *Organised Sound* 14, 2 (2009).

[15] Anna Savery, Sam Ferguson, and Andrew Johnston. 2024. Gesture and Narrative: Blending Human Performance with Visual Storytelling. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

[16] Atau Tanaka. 2010. Mapping Out Instruments, Affordances, and Mobiles. *Proceedings of the International Conference on New Interfaces for Musical Expression* Nime (2010).

[17] Diana Young. 2002. The hyperbow controller: Real-time dynamics measurement of violin performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

[18] Diana Young. 2008. Classification of Common Violin Bowing Techniques Using Gesture Data from a Playable Measurement System.. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*