MetaBow: Gesture Mapping in Immersive Sonic Environments

Davor Vincze
Hong Kong Baptist University
vincze@hkbu.edu.hk

Roberto Alonso Trillo
Hong Kong Baptist University
robertoalonso@hkbu.edu.hk

Peter Nelson
Hong Kong Baptist University
peteracnelson@hkbu.edu.hk

ABSTRACT

This paper presents the MetaBow, an augmented violin bow designed to control digital sound processing through real-time motion tracking. We discuss the challenges of mapping Inertial Measurement Unit (IMU) data to audio parameters in immersive multi-speaker environments and propose hybrid strategies using both direct mapping and machine learning models. We reflect on design choices, trade-offs, and performer experience, drawing from technical development and performance contexts. Three condensed case studies illustrate the system's versatility in spatial and interactive musical performance.

Author Keywords

Augmented string bow, machine learning and gesture, mapping strategies in immersive environments

1. INTRODUCTION

The MetaBow (Figure1) builds on the tradition of hybrid instruments that integrate acoustic performance with digital interactivity [18]. It is a sensor-augmented bow frog containing a 9DOF IMU, a MEMS microphone, and capacitive sensors [1]. Its form factor replicates traditional bows, minimizing performer adaptation. However, real-time IMU tracking introduces challenges such as jitter, drift, and non-intuitive data behavior. Our goal was to create a responsive, expressive interface for immersive environments, combining data processing with gesture recognition tools such as FluCoMa [10, 17]. We developed custom software to facilitate communication between the bow and Max8, translating raw and fused sensor data into control parameters usable in real-time performance contexts. These included acceleration, velocity, quaternion [7] orientation, and derived high-level features skewness, tilt, and roll (see MetaBoard in Figure 2).

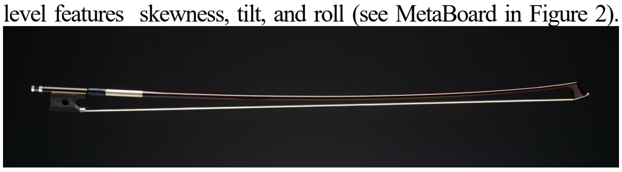

Figure 1. The MetaBow.

2. RELATED WORK

Gesture-based instruments using IMUs and machine learning (ML) have been widely explored in NIME and related contexts [9]. Notable examples include the K-Bow [15] and Wekinator [6], both of which employ classification techniques for mapping performance gestures to musical outputs. Our approach builds on this foundation while specifically targeting multi-speaker immersive environments. Other relevant research includes augmented instruments that integrate ML for expressive interaction [8], sensor fusion in IMU-driven systems [11], and gesture-based mapping paradigms using tools like PCA and UMAP [14]. The MetaBow project differs in its combination of low-latency performance, multi-modal mapping techniques, and spatial sound control across varying speaker configurations.

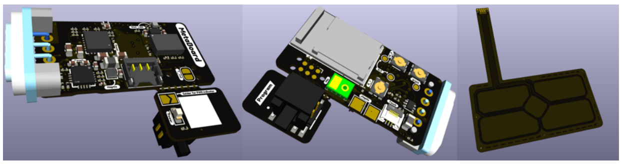

Figure 2. MetaBoard's Design.

3. TECHNICAL FRAMEWORK AND DESIGN RATIONALE

The MetaBow Suite (Figure 3) routes Bluetooth data to Max/MSP via Open Sound Control (OSC). It collects and processes IMU data using both native and custom fusion algorithms [2]. Data smoothing techniques averaging 50–100 milliseconds were employed to minimize jitter, while filters and update caps mitigated instability from quaternion drift (see data extraction modules in Figure 4). Through experimentation, we found gyroscopic and frame data to be most useful for movement detection, while quaternion data worked best for spatial mapping within limited arcs. To maintain longitudinal stability, we calibrated the system per session and avoided full 360° mappings, focusing instead on predictable ranges.

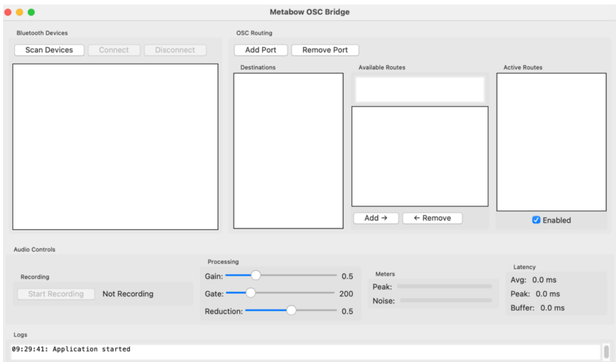

Figure 3. The MetaBow Suite.

We chose IMUs over camera-based systems to maintain portability and reduce setup time, especially in variable performance contexts. While this introduced limitations in tracking precision and required recalibration, the system's flexibility and affordability made it ideal for asynchronous, distributed creative workflows. Our architecture prioritized minimal performer disruption, ensuring the bow remained as close as possible in weight and feel to traditional models. This minimized sensorimotor adaptation and allowed the violinist to focus on musical goals rather than system navigation.

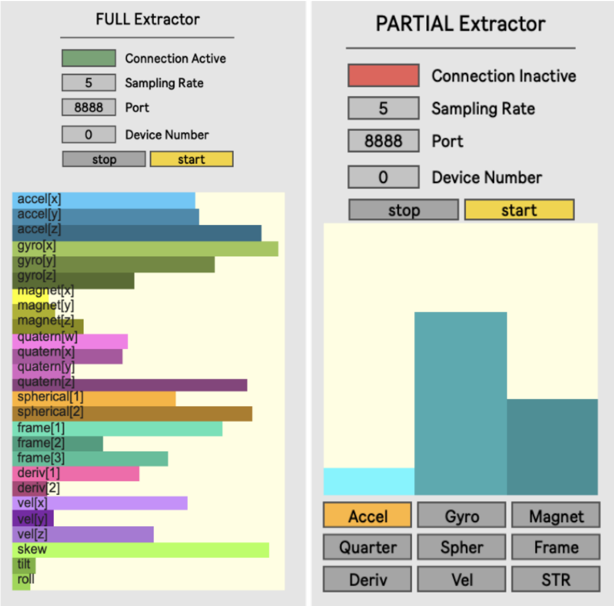

Figure 4. Custom Data Extraction Module.

4. MAPPING STRATEGIES AND PERFORMER EXPERIENCE

We adopted a hybrid approach to gesture mapping:

- **Direct Mapping of raw sensor data** (e.g., accelerometer or gyroscope) to control sound parameters like filter cutoff, delay feedback, or spatial rotation.
- **Threshold Triggers** for discrete events such as turning effects on/off, avoiding activation by unintentional micro-movements.
- **ML Classification** using FluCoMa's [fluid.mlpclassifier~] to map states like bow tilt or motion onset into musical modes (e.g., triggering transitions or switching between spatial presets).
- **Regression models** for tracking spatial angles, albeit limited to under 180° due to wrap-around and calibration issues.
- **Latent Space Navigation** via PCA and UMAP to explore sound banks and textures, offering gestural memory and fluid timbral shifts.

We prioritized intuitive mappings and artistic responsiveness. For instance, overly complex mappings reduced expressivity, especially when multiple parameters were tied to single gestures. Performer feedback led us to simplify mappings, reserving complexity for rehearsed or semi-automated sections. The result was a more confident, expressive performance that encouraged gestural exploration without overwhelming the player.

5. SPATIAL AUDIO INTEGRATION

Spatialization was handled through IRCAM's Spat~ library in Max/MSP [5], with setups in both fixed (Visualization Research Center – VRC) and portable (OTTOsonics) speaker configurations (Figure 5). Spatial strategies included:

- **Vector-Based Amplitude Panning** (VBAP) for real-time gesture-based sound movement [13].
- **Automated diffusion** tied to performance dramaturgy.
- **Speaker-specific zones** assigned to performers or effects for clarity.

Performance-specific adjustments addressed venue constraints. In the VRC [4], we adapted to missing mid-level speakers by emphasizing horizontal gestures. In OTTOsonics [12], we compensated for low-end deficiencies using subwoofers and EQ to distribute energy effectively across the frequency spectrum. These spatial approaches balanced physical gestures with immersive responsiveness, reinforcing the embodied aspect of sound diffusion. Beyond these mappings, most spatialization strategies were automated, aligning with overall dramaturgy rather than real-time interface control, drawing inspiration from strategies developed by artists such as Natasha Barrett [3] or Enrique Tomás [16].

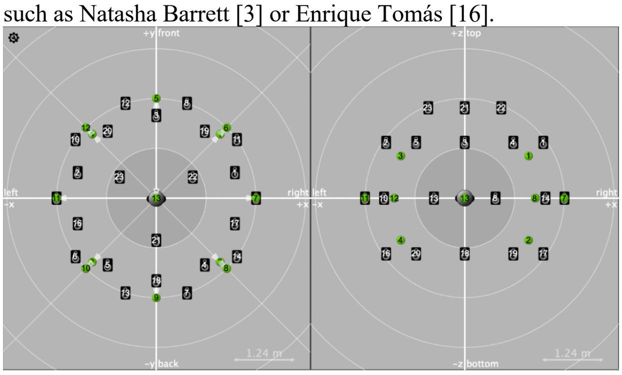

Figure 5: An Examples of a Spat environment.

6. CASE STUDIES

6.1 9 Shards

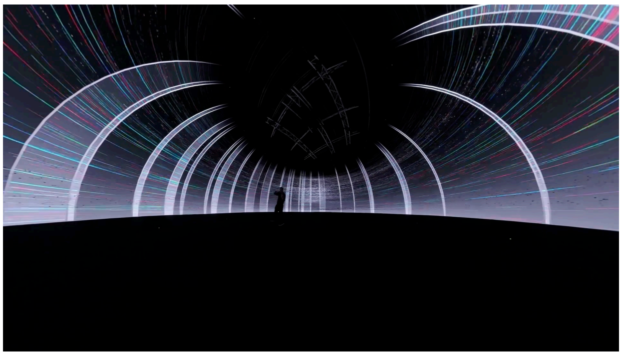

Figure 6: Roberto Alonso performing with MetaBow in VRC.

In this immersive audiovisual work, MetaBow controlled low-frequency gestures and PCA-based sample browsing (Figure 6). Initial gestures triggered subwoofer impacts using gyroscopic thresholds. Tilt-based navigation explored sound clusters rendered spatially across the dome. Mid-performance, gestural control was relinquished to semi-automated systems, with MetaBow modulating DSP effects subtly. The final section reintroduced gesture-based panning, mapping bow direction to speaker positions. Performer feedback indicated that restrained mappings enhanced expressivity and immersion.

6.2 Remote Gestures

This distributed performance leveraged OSC and video streaming to synchronize violinists and a robotic system in separate locations (see Figure 7). In the first section, MetaBow motion data controlled both sound and synchronized real-time visuals, connecting remote and local players. The second section introduced a robotic arm bowing physical instruments, spatialized using contact microphones and MetaBow gestures. In the final segment, the bow-controlled robot 's calligraphic gestures, which were visualized and used as scores for a percussionist. These embodied visual-music relationships illustrated the MetaBow's versatility in collaborative and interdisciplinary settings.

6.3 Exo Signals

In a 23-speaker ambisonic space, MetaBow controlled DSP depth and spatial position, while the audience used mobile apps to interact. The first section linked bow velocity to reverb density and filter modulation. In the second section, low-motion states reduced activity while increased gesturing triggered complex spatial textures. Audience participants could activate mobile-triggered sounds, layered

Figure 7: Performers are controlling the movement of the robotic arm via MetaBow.

into the performance. In the final section, grains of short, pre-recorded violin phrases were positioned in three spatial axes via MetaBow, while audience members modulated timbral ranges via mobile sliders. This performance showcased MetaBow's effectiveness in participatory and layered sound environments (see Figure 8).

Figure 8: Roberto Alonso controlling the sound source position via MetaBow.

7. REFLECTIONS AND LIMITATIONS

While the MetaBow system achieved responsive, intuitive mapping in performance, we acknowledge limitations. IMU data remains vulnerable to drift, particularly in long, continuous mappings. Orientation inconsistencies across sessions required recalibration and constrained reproducibility. Our reliance on Bluetooth imposed range limitations and necessitated a clear line of sight. These trade-offs, however, were offset by system simplicity, ease of deployment, and strong performer adaptability. Future versions may incorporate real-time recalibration routines, hybrid sensor networks, or alternative wireless protocols.

Beyond technical limitations, a central design challenge remains balancing flexibility and artistic focus. We found that meaningful control often emerged from constraint—limiting mappings to a few reliable gestures encouraged performer fluency. For example, in 9 Shards, the mapping between bow position and latent sound space enabled the performer to navigate a diverse timbral palette and intuitively return to preferred textures, while in Remote Gestures, controlling both the sonic output of the violin as well as the calligraphic strokes of a robot-driven brush forced the player to consider multiple and divergent outputs from a single musical gesture. The use of the Metabow to control spatialised speaker arrays in Exo Signals led the player to a similar splitting of intention, where a musical gesture could be considered as having multiple outcomes for the performance. We found the best results when starting with highly restrained mappings, and that richer and more complex mappings were best introduced gradually, tied to visual or structural cues in the composition. This iterative feedback between system and artistic process underpinned the MetaBow's development and will continue to shape future directions.

8. CONCLUSION AND FUTURE WORK

The MetaBow system represents a step toward integrating gestural nuance and spatial responsiveness in live performance. Our approach, combining direct mapping and machine learning, enabled a rich range of interactions while retaining a focus on artistic clarity. Through three contrasting case studies, we demonstrated its flexibility in immersive audio, distributed performance, and audience interaction. We aim to expand this framework by exploring cross-performer mappings, where one musician's gestures influence another's spatial or timbral parameters. More robust sensor fusion, including magnetometer-free orientation tracking, could improve stability. We are also exploring visual augmentation to reinforce gestural intent, and extending the MetaBow interface to other instruments or wearable configurations. Our long-term goal is to contribute an adaptable, open-source toolkit that supports creative practitioners working at the intersection of embodied interaction and spatial sound.

9. ACKNOWLEDGMENTS

This work was supported by the ITF Future Cinema Systems grant and the Initiation Grant for Faculty Niche Research Areas at HKBU. We are grateful to our collaborators, performers, and students who contributed to the testing, development, and performances documented here.

10. ETHICAL STATEMENT

As an artistic research project, ethical considerations centered on performer agency, transparency in collaborative contexts, and the responsible representation of human-machine interaction. We were mindful to design systems that enhance, rather than constrain, expressive performance, and we ensured that all collaborators were fully informed, rewarded and credited for their contributions.

11. REFERENCES

[1] Alonso Trillo, R., Nelson, P. A., & Michailidis, T. (2024). Rethinking Instrumental Interface Design: The MetaBow. *Computer Music Journal*, *47*(2), 5-20.

[2] Alonso Trillo & R., Nelson, P. A (2025). *MetaBoard* [GitHub repository]. GitHub. Retrieved [05/02/2025], from https://github.com/MetaBow/*MetaBoard*

[3] Barrett, N. (2012). The perception, evaluation and creative application of high order ambisonics in contemporary music practice. *Ircam Musical Research Residency report*.

[4] Bethge, P. H., & Shaw, J. (2025). Ask Me Anything: Information Retrieval for Interactive Narrative Generation. ISEA 2025 Proceedings.

[5] Carpentier, T. (2021). Spat~: a comprehensive toolbox for sound spatialization in Max. Ideas Sónicas, 13(24), 12–23.

[6] Fiebrink, R., & Cook, P. R. (2010, August). The Wekinator: a system for real-time, interactive machine learning in music. In *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)(Utrecht)* (Vol. 3, pp. 2-1).

[7] Gogberashvili, M. (2014). Split quaternions and particles in (2+1)-space. *The European Physical Journal C*, *74*, 1-9.

[8] Jensenius, A. R. (2022). Sound Actions: Conceptualizing Musical Instruments. MIT Press.

[9] Jourdan, T., & Caramiaux, B. (2023, May). Machine learning for musical expression: A systematic literature review. In *New Interfaces for Musical Expression (NIME)*.

[10] Moore, T., Bradbury, J., Tremblay, P. A., & Green, O. (2021). Making Machine Learning Musical: Reflections on a Year of Teaching FluCoMa. *Journal SEAMUS*, *32*(1/2), 24–38.

[11] Nirmal, K., Sreejith, A. G., Mathew, J., Sarpotdar, M., Suresh, A., Prakash, A., ... & Murthy, J. (2016, July). Noise modeling and analysis of an IMU-based attitude sensor: improvement of performance by filtering and sensor fusion. In *Advances in optical and mechanical technologies for telescopes and instrumentation II* (Vol. 9912, pp. 2138-2147). SPIE.

[12] OTTOsonics – UBIMUS. (n.d.). Retrieved February 5, 2025, from https://ubimus2024.fah.usj.edu.mo/presentation/about-ottosonics/

[13] Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the audio engineering society*, *45*(6), 456-466.

[14] Roma, G., Xambó, A., Green, O., & Tremblay, P. A. (2021). A General Framework for Visualization of Sound Collections in Musical Interfaces. *Applied Sciences*, *11*(24), 11926.

[15] Schedel, M., & Fiebrink, R. (2011, July). A demonstration of bow articulation recognition with Wekinator and K-Bow. In *ICMC*.

[16] Tomás, E., Goeschke, F., & Kaltenbrunner, M. (2024). Exploring Design Patterns for Spatial Instruments: User-Driven Strategies, Spatialized Synthesis and Loudspeaker Topologies. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 127-132).

[17] Tremblay, P. A., Roma, G., & Green, O. (2021). Enabling programmatic data mining as musicking: the fluid corpus manipulation toolkit. *Computer Music Journal*, *45*(2), 9-23.

[18] Verrando, G. (2012). *La nuova liuteria: orchestrazione, grammatica, estetica*. Suvini Zerboni.