M.E.S.H. – Navigating the Latent Space Through the Body-Instrument; Reclaiming One's Digital Traces in the Age of Large AI models

Matteo Bellefleur-Martinez

Department of Music
University of Montreal

mateo.bellefleur-martinez@umontreal.ca

Evan Moore

Department of Music
University of Montreal

evan.moore@umontreal.ca

Stéphane Drouin

Department of Music
University of Montreal

stephane.drouin@umontreal.ca

Abstract

This paper introduces M.E.S.H., a gestural musical instrument that reclaims human agency and physical engagement in AI-assisted music creation. While large generative text-to-audio models often restrict individual artistic expression, M.E.S.H. utilizes neural networks in Max MSP as a co-creative mediator between bodily intent and sound. By placing the performer's body at the center of the workflow, our research-creation project wishes to establish a novel relationship between biotechnology and algorithmic generation, one that actively confronts the ethical implications of biological data use within the corporate technology ecosystem. We evaluate the system across creative and ethical axes, using a live performance as one case study to demonstrate the dynamic negotiation of agency between musician and AI. Ultimately, we are sharing this framework as open software so that other artists can reappropriate their biometric data to create music that is highly personalized and physically engaged with the assistance of a small AI system.

1 INTRODUCTION

The current environment of hyper-reproduction of media, in which generative systems are exponentially multiplying and transforming existing material (Coelho, 2025), fundamentally alters the roles of musicians and performers. Large text-to-audio models enable many users without a background in music theory to generate complete pieces of music using just their personal computers, requiring no audio equipment. However, these generative tools (Suno, Udio, Lyria, etc.) present significant structural limitations for musicians. Trained on massive datasets, these systems lack the specificity needed to support the development of a sustainable and personalized musical practice. Rather than encouraging individual expression, these vast datasets reflect and amplify overarching trends (Vigliensoni & Fiebrink, 2024), which is problematic in a field such as music that traditionally values originality and diversity. Also, a fundamental limitation of these generative systems is their reliance on simple text prompts, which circumvents the physical body and expressive connection between movement and sound. This paradigm of sound creation leaves little to no room for nuanced musical interpretation.

Therefore, designing a system that uses AI for dynamic, creative interaction requires a critical shift in perspective. AI must be understood not as a reproductive tool, but as a mediating environment that allows for the reframing and reinvention of ideas. Interactive machine learning and training models on small, customized datasets enable the development of creative relationships between movement and sound (Vigliensoni & Fiebrink, 2024). In this respect, M.E.S.H. builds on a longstanding tradition of embodied human-computer interaction and interactive machine learning for musical performance. Rebecca Fiebrink's Wekinator, a "meta-instrument" that lets performers train gesture-to-sound mappings on the fly rather than through batch processing (Fiebrink, Trueman & Cook, 2009), is a direct precursor to our approach, as are body-centric controllers such as the MiMU Gloves (Mitchell, Madgwick & Heap, 2012) and Kinect-based instruments. M.E.S.H. situates itself within this lineage while distinguishing itself on two points: it requires no dedicated sensor or wearable, relying instead on a standard camera, and it foregrounds the ethical reappropriation of biometric data as an integral part of the instrumental gesture. Currently, the use of these architectures imposes a significant barrier to entry, as they are too complicated to use for musicians and performers without niche AI expertise (Meyer et al., 2025).

Concretely, M.E.S.H. is a gestural instrument programmed in Max MSP: a local Python script uses MediaPipe to track the performer's face, hands, and body, and small neural networks (built with the FluCoMa toolkit) map this biometric data, in real time, onto three audio engines: internal microphone feedback, audio synthesis, and sample granulation. The M.E.S.H. project centers on the development of an accessible instrument designed to facilitate a co-creative workflow, one that adapts to the physical and aesthetic singularity of each artist. The methodology of this project involves evaluating the interface's potential for musicking (Small, 1999) through applied artistic practice, while working within a common interaction pipeline based on facial recognition and machine learning. This pipeline is applied across three specific audio engines: internal microphone feedback, audio synthesis, and sample granulation from personalized sound banks. This research-creation methodology yields experimental insights into agency through the synthesis of observations and critiques of the system. Furthermore, this cross-experimentation process provides a framework to analyze how the instrument facilitates the singular application of a small-scale AI system for personalized artistic needs across diverse musical contexts.

Inspired by recent developments in latent movement-to-audio mapping (Meyer et al., 2025; Nabi et al., 2024), M.E.S.H. is conceptualized as one of the several stepping stones toward a fully co-creative workflow with AI, one that centers the body and gesture at the center of music creation. The progress of our research is evaluated in the article along two complementary axes. On a creative level, this research examines the limitations of the digital instrument, programmed in Max MSP with topographic detection handled in Python. This research examines the expressive capacity of neural learning, specifically whether it can meaningfully link movement and sound by soliciting the artist's proprioception. We will evaluate the limits of user customization within the tool and the effectiveness of the workflow, particularly considering the algorithm's non-linear behaviours and emergent properties (Coelho, 2025), which place the musician in a constant negotiation of agency with the AI.

On an ethical level, the evaluation centers on the reappropriation of biometric data. The project aims to redirect facial recognition technology away from its surveillance-driven applications, towards reclaiming the body and face as engines of artistic identity and creative action. Finally, we will validate our commitment to technological anti-elitism and accessibility by ensuring that utilizing and adapting the instrument to specific musical needs can be accomplished with minimal equipment and a negligible environmental footprint.

2 THEORETICAL FRAMEWORK

The emergence of New Interfaces for Musical Expression (NIME) currently prompts a semantic revision of the notion of the "instrument." As highlighted by Madeleine Le Bouteiller, a doctor of musicology at the University of Strasbourg: "The dynamics surrounding the creation and use of digital instruments contribute to forging an 'open' identity for them" (Le Bouteiller, 2020).

In other words, these systems establish new parameters of expressivity and musical instrumentality, thereby creating tensions between traditional definitions and contemporary situations. Consequently, we will evaluate the position of M.E.S.H. through the notion of "agency negotiation," which is inherent to the principles of neural network learning and latent space navigation.

A consensus or common sense has formed around a nominal definition of the musical instrument. According to this definition, an instrument must: produce sound, be controllable by gestures, and respond faithfully to the musician's intentions. Thus, the primary questions we ask are the following: Should we categorize M.E.S.H. as an interface, a device, or a musical instrument? How can we formulate a definition that encompasses these contemporary situations and, within this framework, responds to the intrinsic reality of our object?

2.1 Identity and Creative Agency

First, it is worth recalling that agency refers to the "capacity of an individual to act intentionally, to influence their environment, as well as the feeling of being the initiator of their own actions." In the context of M.E.S.H., this agency is redistributed and reinvested both physically and perceptually. The user learns to inhabit their own body and their own possibilities for facial expressivity as a functional and musical interface.

By "agency negotiation," we therefore mean this process in which a dialogue exists between, on the one hand, the performer's gestural actions and, on the other, the results generated by the algorithm that interprets (encodes) and proposes a sonic environment. Here, the user constantly seeks their own identity and expressivity within a space of multiple sonic potentials. It is within this friction that a new bodily listening develops by observing the sonic consequences of their micro-movements, the musician sharpens their proprioceptive capacity, transforming an unconscious facial gesture into a deliberate musical act.

This "delegation" of agency raises several fundamental questions regarding creative processes such as this one: First, what are the creative reasons justifying the delegation of a portion of one's agency to a third-party agent? And fundamentally, does the musician's ability to rediscover their expressivity and identity within the sonic space and to manipulate sound through this constant negotiation truly constitute musical play?

To answer these questions, we argue that this agency negotiation is primarily rooted in an ethical approach, carrying a discourse that is both political and artistic. M.E.S.H. relies on the subversion of a technology: facial recognition and biometric analysis, which are historically and socially connoted by logics of surveillance, profiling, or digital identity. By reinvesting this principle within an intimate and creative framework, the performer is invited to reappropriate their own biometric data. The face is then no longer passive data to be extracted or analyzed for control purposes; it becomes once again living matter, a vector of expressivity. This approach is thus akin to a form of bio-art, where technology is used to restore the body's musical and poetic sovereignty.

2.2 A "Model" and Metaphorical Instrument

Our hypothesis is as follows: M.E.S.H. is a model that allows the performer to maintain what Claude Cadoz terms an "instrumental-type relationship" with the algorithm (Cadoz, 1999). In other words, a semantic revision would still be necessary to formally qualify the

system as a "musical instrument." Nevertheless, we hope to demonstrate its expressive and instrumental potential, but also to explain that this agency negotiation constitutes the core of the performer's musical performance, leading them to develop a certain form of virtuosity linked to their proprioceptive capacity.

For Atau Tanaka, a musical instrument is an "expressive tool in the hands of the performer, an open system, a useful metaphor, a model that serves the creativity of musicians; it is an operating model and not a strict definition" (Tanaka, 2000). It is therefore within this open and functional conception that M.E.S.H. is situated. The system forms a complete entity dedicated to live performance by transforming micro-movements into musical gestures through the algorithm's interpretation.

Finally, if M.E.S.H. reflects the open instrumental model defined by Atau Tanaka, its nature also invites us to think of it as a "meta-instrument". M.E.S.H. is an architecture offering a multiplicity of sonic configurations (synthesis, microphone feedback, and sampling), which relies on the ethical reappropriation of digital data through a negotiation and a co-creative act between human and algorithm.

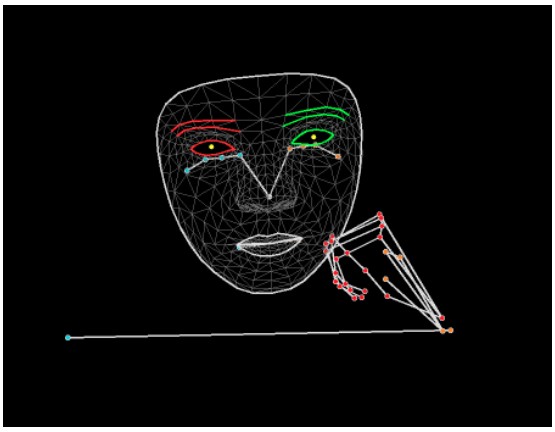

Figure 1 Real-time modeling of facial topography (MediaPipe, Python)

3 INSTRUMENT DESCRIPTION

3.1 Technical Architecture

Advocating an ethical commitment to accessibility, the M.E.S.H. ecosystem relies on free and open-source software tools (apart from Max MSP). The central Python script was developed through vibe-coding with the Tiptree Systems LLM; a dialogue-driven programming approach with an AI language model.

Concretely, the Python script analyzes a camera's video feed and deploys three modules from MediaPipe (Google), an open-source machine learning framework for biometric analysis.

The system captures overall body positioning, with heightened precision for the hands and face. These coordinates are condensed into a continuous stream of 1,119 values, which are then routed to the Max MSP environment via the OSC communication protocol. This large volume of raw data is analyzed and coupled with artificial neural networks using the free FluCoMa toolkit (Tremblay et al., 2022). This final processing stage is what enables the performer to physically interact with the three instrumental facets offered by M.E.S.H.

3.2 Axes of Control

The performer continuously controls the three instrumental bodies of M.E.S.H. by manipulating their facial features and body. Each of these instruments deploys its own control axes and embodies a fundamental ethical objective at the heart of our research-creation process.

The first instrument establishes a direct coupling between real-time biometric data and an audio synthesis engine, directly translating physicality into sonic material. This axis highlights the radical personalization of timbre: it is the performer's unique topography that sculpts the sound. This represents a true reappropriation of biometric data, diverting it from its usual applications to serve intimate artistic expression. This facet of the instrument illustrates the power of neural networks in mapping vast amounts of data to sonic parameters.

The second instrument deploys a three-voice polyphony system by exploiting the computer's microphone and audio output system (internal or external), forming a reactive auditory ecosystem based on acoustic feedback. It carries an anti-elitist message by emphasizing accessibility, demonstrating how modest equipment, once placed at the heart of a feedback loop, can successfully generate complex and personalized sonic colors in conjunction with AI.

Finally, the third model offers the performer the ability to organically navigate within personalized audio sample banks. By integrating the musician's sonic memory into the very architecture of the system, this axis transcends the simple playback of audio files. Its goal is to harness the potential of individual expression, allowing the artist to fuse their biometric footprint (the gesture) with their own cultural and musical background (the sample). These three models thus explore the possibilities for developing and personalizing individual artistic expression, which we sought to deepen within a digital lutherie built with AI assistance.

3.3 Sonic Personalization

Traditionally, the ergonomics of a musical instrument impose its morphology on the musician, requiring them to discipline their body to conform to it. M.E.S.H. attempts to reverse this paradigm: through the MediaPipe mask, the instrument becomes the face and body of the performer. The 478 landmarks of the virtual mask map onto the unique topography of their features. This setup actively engages the musician's proprioception, utilizing their own muscular resistance as a familiar and instantaneous sensory mediator in their interaction with the artificial intelligence instrumental model.

M.E.S.H. also offers the performer the freedom to import and interact with algorithmic models shaped by other users. This open architecture transforms the instrument into a space for identity exchange; not only can the performer interact with the loaded algorithmic model through their own topography, but they can also explore their relationship with the instrument through the sonic and identity-based personalization of the user who previously programmed the model. At this point, the topography of a first person can merge with that of a second through their relationship with the digital instrument.

4 NEURAL NETWORK TRAINING AND EXPERIMENTAL RESULTS

4.1 Neural Network Programming

Following the experiments conducted to embody a musical work with M.E.S.H., this section presents the results of our research. We examine the integration of AI within M.E.S.H., its deployment for musical performance, and the relational dynamics between the user and the system.

In the traditional creative process, multiple technical steps typically separate artistic intention from its realization, compromising the musician's stimulating and creative spontaneity. For instance, an electronic musician usually has to manually assign external MIDI controllers and adjust specific response curves to link the intensity of a gesture to the evolution of timbre. Through M.E.S.H., AI enhances and accelerates this process by directly interpreting the musician's intention. The distance between creative intent and sonic result is thus reduced to a single complex equation, mediated by the AI. Nevertheless, the user retains decision-making power over the system's architecture thanks to the accessibility of the training and learning parameters. The performer's movement, however mundane, becomes the "painter of the equation," thereby framing the AI as a co-creative partner.

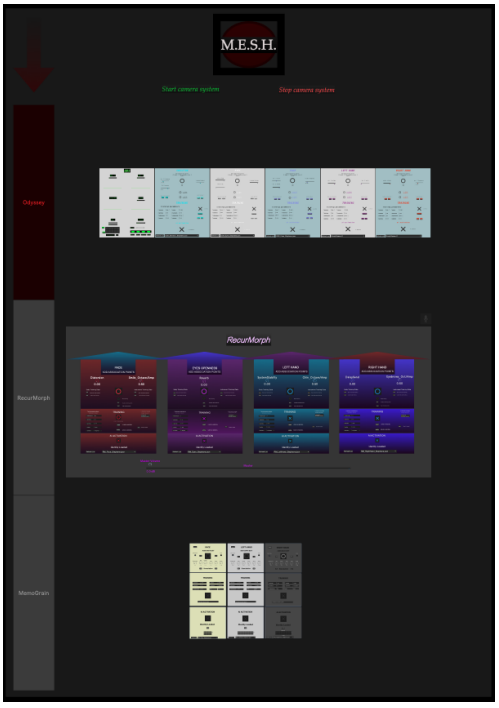

Figure 2 The M.E.S.H. interface in Max MSP

Through the neural network linked to the M.E.S.H.'s interface, the user can create their own instrumental model during a training phase, establishing an active feedback loop between them and the neural network. Concretely, the operator sculpts their desired sonic profile by manipulating virtual knobs, then directly associates this sonic state with a specific body posture that will become its trigger. Once these anchors are established and training is initiated, the algorithm generates and offers bespoke expressivity, intimately tied to their biometrics and gestures. Not only does the operator reappropriate their biometric data, but their interaction with the artificial intelligence also shortens the temporal gap between creative desire and its realization. Thus, through an interface like M.E.S.H., the user is no longer solely a musician; they become the programmer of their own digital lutherie, embedded in a co-creative flow whose speed of execution is immediately apparent.

Following this brief association phase, the musician meets their new instrument by navigating within the latent space of an initial model. They discover how their micro-movements trace unique trajectories in this multidimensional space, offering a musical performance that reflects not only their topography but also the choices made during training.

To tame the complex topography of this latent space, the user then rediscovers the patience inherent in mastering any instrument, transposed here to the heart of an artificial neural ecosystem. Far from generative instantaneity, this artisanal and somatic optimization process requires time, deep listening, and a degree of improvisation. The performance phases act as a diagnosis: they invite the musician to add or modify landmarks in their playing to refine their navigation. In this continuous learning loop, each trained model encourages the performer to push the boundaries of their interaction with the machine. To perfect an expressive language, the user is called upon to restart the neural network's training and learning process, thereby indefinitely prolonging the co-creative relationship with the system.

4.2 Productive Tension, Experimental Co-creativity and Music Performance

Following our experimentation with the three facets of M.E.S.H., we observe that navigating the latent space **departs from** the predictability typically associated with traditional parameter-linked knobs. When the algorithm calculates interpolations between the various bodily and sonic anchor points defined by the user, it does not trace a simple straight line; it "invents" non-linear trajectories. It is precisely in these interstices that the algorithm speaks the loudest. By generating unexpected sonic textures during transitions, the AI offers a continuous creative prompt. It no longer merely executes a command; it submits a creative hypothesis to the performer.

Faced with this proposition, the musician can react in real-time during performance: they can choose to integrate this interpretive "anomaly" of their gesture into their expressive vocabulary, linger on it to explore its nuances, or, discard it by altering their posture. This dynamic radically transforms the instrument's performance possibilities. The musician is no longer in a position of absolute mastery, but rather in a stance of active listening and constant negotiation with the instrument presented. The artificial intelligence acts as an improvising playing partner, capable of surprising its performer and thus sparking a creative dialogue. Musical performance therefore emerges from the stimulating friction between the somatic gesture and its algorithmic interpretation.

Following our experiments, we observe that the instrument achieves its full expressive potential when manipulated by it's very own trainer. Because the neural model was fed by the biometric specificities and muscular micro-tensions uniquely guiding its creator's topology, there is an intimate alignment between the software

architecture and the body of the model's trainer. Consequently, the algorithm will tend to be anchored more often in the "creative interpolations" of the latent space rather than in its originally intended programming axes.

Furthermore, various uses of M.E.S.H. across multiple distinct environments highlight that poor performance conditions in front of the system can make an otherwise well-trained gesture difficult for the system to interpret. Indeed, poor lighting or a gesture extending beyond the camera's frame can significantly alter the system's response, inviting the user to constantly monitor their digital representation window to ensure their body is being properly captured by the system. Finally, our personal observations indicate that designing the initial neural learning model cultivates a stimulating, co-creative relationship, but that optimizing the sonic output can prove to be a laborious and less spontaneous endeavor.

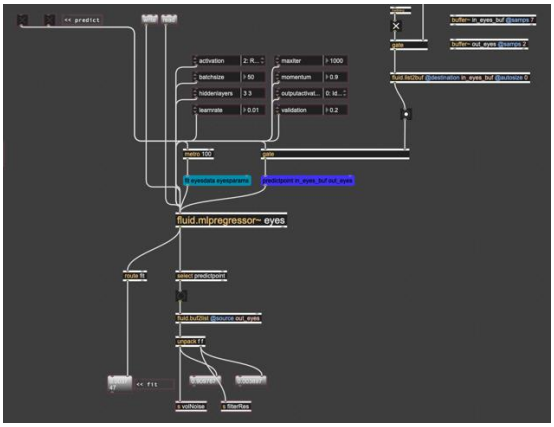

Figure 3 Implementation of regression via FluCoMa in Max MSP

5 TOWARDS AN AI OF REAPPROPRIATION

Central to the M.E.S.H. project is a deliberate resistance to large-scale generative models of hyper-reproduction. The artist uses their own biometric data to convey musical intent; the use of small, personally curated databases allows for the avoidance of bias and homogenization associated with mass-market tools. In contrast to more technical applications of machine learning, where training data are often viewed as objective representations of *truth*, here they are conceptualized as interpretations of a subjective truth (Vigliensoni & Fiebrink, 2024).

5.1 Facial Recognition as a Medium for Reappropriated Expression

Facial recognition technologies in the digital industry have historically been designed for surveillance and profiling purposes. The M.E.S.H. project justifies the use of this technology through an act of ethical subversion. The MediaPipe pose estimation model extracts bodily coordinates to link them to a predetermined sound space. Through this process, the unique topography of the face generates direct mappings to sound, allowing to develop an individualized creative relationship with the system.

Biometric data always remains under the artist's control and is never shared online: the Python script using MediaPipe runs locally on the musician's computer. Facial landmarks are instantly converted into a continuous stream of abstract digital data transmitted locally and in real time via the OSC protocol. Biometric data is therefore never stored or clearly identified; it is immediately converted into an audio control parameter. While the artist retains exclusive control over their data, this reappropriation relies on a detection architecture developed by the tech industry (Google's MediaPipe). This reliance means that the autonomy of M.E.S.H., while robust in terms of data privacy, is not absolute.

5.2 Latent Space Navigation and Artistic Singularity

When developing the initial iterations of M.E.S.H., we were aware of the expressive limitations of neural learning systems; these often create latent spaces that are too abstract for straightforward control. To address this challenge, we adopted an approach rooted in active listening, positioning the musician's proprioception (movement awareness) in the role of the performer. Through this *embodied exploration* (Nabi et al., 2024), the body becomes the mediating interface between creative intent and sound.

By allowing users to input their own biometric data and control settings into the different features of M.E.S.H., we consider this to be a step toward accessible, customized digital lutherie, as it is a way to personalize the instrument and enable a more individualized experience. Nevertheless, while such custom integration is possible, the interaction ultimately serves as a control interface for a pre-existing instrument. Designing an AI-assisted musical instrument that truly empowers any user to reappropriate their biometric data on their own aesthetic terms, allowing to deploy a unique musical identity, remains an ideal that future research must strive towards.

6 CONCLUSIONS

Ultimately, this methodology effectively answered the research question, demonstrating that it is possible to design an instrument that places the bodily gesture at the center of AI-assisted musical creation, while offering a co-creative workflow where the user builds a new expressive practice based on proprioception and active listening. Yet, absolute personalization faces a structural boundary: although the model individualizes the gesture by customizing mappings between the body and the sound engine, the audio output remains confined to the pre-programed facets of the instrument. Moreover, the optimization of both the model and the resulting sound remains a lengthy and demanding process. Developing an accessible instrument that requires no niche AI expertise and empowers each user to achieve a singular music identity remains a major challenge for future research in the field.

From an ethical standpoint, the sharing of the Max MSP and Python source code, along with the exclusive use of equipment that is both cost-effective and low-impact, reinforces our commitment to technological and material anti-elitism. However, while this approach eliminates the need for costly sensors, the project's accessibility is constrained by its reliance on Max MSP. Nevertheless, this dependency represents a deliberate conceptual stance. Rather than providing a closed system, M.E.S.H. seeks to situate itself within the practice of creative coding.

Indeed, the user is granted complete access to the sound engine through a popular music programming software, deliberately encouraging a complete appropriation of the instrument by musicians already engaged in the field of algorithmic sound creation.

SUPPORTING MATERIALS

Video documentation of M.E.S.H. in use is available on the author's YouTube channel: https://www.youtube.com/@M.E.S.H.Instrument
The full source code (Max MSP and Python) is openly available at https://github.com/Spleen98/M.E.S.H.

ACKNOWLEDGMENTS

M.E.S.H. uses MediaPipe for biometric analysis. MediaPipe is developed by Google LLC and is also distributed under the Apache License 2.0.

AUTHOR DECLARATIONS

Some parts of the programming were developed using large language models as coding assistance tools. The project's design, methodology, and artistic framework were developed and directed by the authors.

REFERENCES

Cadoz, C. (1999). Musique, geste, technologie. In H. G. et R. de Vivo (Éd.), *Les nouveaux gestes de la musique* (p. 47-92). Editions Parenthèses. https://hal.science/hal-01083792

Coelho, G. (2025, septembre 10). *AI in Music and Sound : Pedagogical Reflections, Post-Structuralist Approaches, and Creative Outcomes in Seminar Practice*. Zenodo. The AI Music Creativity Conference (AIMC). https://doi.org/10.5281/zenodo.16946639

Fiebrink, R., Trueman, D., & Cook, P. R. (2009). A meta-instrument for interactive, on-the-fly machine learning. *Proceedings of the 9th International Conference on New Interfaces for Musical Expression, NIME '09*, 280-285. https://doi.org/10.5281/zenodo.1177513

Le Bouteiller, M. (2020). *Instruments numériques et performances musicales : Enjeux ontologiques et esthétiques* (Numéro 2020STRAC002) [Theses, Université de Strasbourg]. https://theses.hal.science/tel-02860790

Meyer, J., Bryan-Kinns, N., Fdili Alaoui, S., Grierson, M., & Fiebrink, R. (2025). Interactive Movement-to-Audio with Pre-Trained Neural Networks. *Proceedings of the 2025 Conference on Creativity and Cognition, C&C '25*, 491-493. https://doi.org/10.1145/3698061.3734415

Mitchell, T. J., Madgwick, S., & Heap, I. (2012). Musical interaction with hand posture and orientation: A toolbox of gestural control mechanisms. *Proceedings of the 12th International Conference on New Interfaces for Musical Expression, NIME '12*. https://doi.org/10.5281/zenodo.1180543

Nabi, S., Esling, P., Peeters, G., & Bevilacqua, F. (2024). Embodied exploration of deep latent spaces in interactive dance-music performance. *Proceedings of the 9th International Conference on Movement and Computing, MOCO '24*, 1-9. https://doi.org/10.1145/3658852.3659072

Small, C. (1999). Musicking—The meanings of performing and listening. A lecture. *Music Education Research*, *1*(1), 9-22. https://doi.org/10.1080/1461380990010102

Tanaka, A. (2000). *Musical Performance Practice on Sensor-based Instruments*. https://www.semanticscholar.org/paper/Musical-Performance-Practice-on-Sensor-based-Tanaka/df07df0e783c63f5848230cb7560f41b02b54e78

Viglensoni, G., & Fiebrink, R. (2024). Data- and interaction-driven approaches for sustained musical practices with machine learning. *Journal of New Music Research*, *53*(1-2), 19-32. https://doi.org/10.1080/09298215.2024.2442361