Personalized Gesture-to-Sound Mapping for Accessible Digital Musical Instruments: An Interactive Machine Learning Approach

Mikaël Molliex
Department of Music
Université de Montréal
mikael.molliex@umontreal.ca

Étienne Abassi
The Neuro
McGill University
etienne.abassi@mcgill.ca

Abstract

Generative AI is reshaping music-making at scale, often centralizing creative agency in dataset curation, prompt engineering or model features. This paper presents a prototype exploring an alternative trajectory for AI-mediated musicianship: a granular instrument designed to learn personalized gesture-to-sound mapping rather than requiring performers to adapt to a fixed mapping. Built in Max/MSP using FluCoMa, the system follows an interactive machine learning workflow in which preferred sounds can be associated with movements selected according to the performer's capacities and interests. A neural network learns the resulting mapping and drives the synthesizer in real time. The prototype supports multiple input modalities and a two-tier interaction model intended to balance ease of access with expressive agency. We thus present preference-first, example-based mapping as a design proposition for personalized mediation, offering a counterweight to hyper-reproductive AI music.

1. INTRODUCTION

Research on accessible digital musical instruments (ADMIs) has increasingly challenged the assumption that musical accessibility can be reduced to substituting or adapting conventional hardware. Recent literature emphasizes that meaningful musical participation depends on the ability to tailor instruments to a user's specific motor capacities, preferences, and expressive goals (Frid, 2019; Frid & Ilsar, 2021; Larsen et al., 2016; Samuels 2019; Samuels & Schroeder, 2019). At a moment when AI-mediated music systems increasingly centralize creative agency in datasets, prompts, and model features, this approach reflects a deeper rethinking of what accessibility means in musical contexts: not merely the removal of physical barriers, but the creation of conditions in which diverse users can engage expressively and agentically with sound.

A central but underexamined obstacle is the dominance of visuo-centric interaction paradigms. Many digital instruments, from DAWs to modular environments, are built around explicit visual control of parameters via graphical interfaces. While precise, such interfaces impose dual-channel cognitive load: users must simultaneously attend visually to interface states and aurally to sound output, while developing technical knowledge of parameter topographies (Duarte et al., 2023; Frid, 2019). For users with physical, cognitive, or perceptual divergences, these demands raise barriers not only to performance but to playful, exploratory engagement with sound. The question is therefore whether interaction models grounded in listening, gesture, and progressive appropriation can provide a more inclusive entry point.

Interactive machine learning (IML) has emerged as a strategy that addresses both the mapping and the accessibility problem. Systems such as *Wekinator* allow users to construct gesture-to-sound mappings by demonstration rather than by programming (Fiebrink et al., 2009), and subsequent work has shown that such workflows can support accessible music-making in practice, enabling facilitators to build customized interfaces for children with disabilities (Parke-Wolfe et al., 2019). The key insight is that IML repositions the user from passive recipient of a fixed instrument to active co-constructor of the mapping layer, a particularly powerful form of agency in accessible instrument design.

The present project builds on this tradition to propose an accessible granular instrument implemented as a standalone Max/MSP application, using FluCoMa (Tremblay et al., 2019) for machine learning and the Grainflow~ library for granular synthesis. Its central design feature is a deliberate two-tier agency structure: a first tier with predefined gesture-sound associations lowers the access barrier for users unfamiliar with gestural controllers or parameter-based synthesis; a second tier allows users to define their own target sounds, import audio, and retrain the model from personalized gesture vocabularies. This graduated structure responds to a well-documented tension in accessible instrument design, that is reducing complexity to enable access often reduces expressive agency and proposes a continuum rather than a binary as its resolution (Frid & Ilsar, 2021; Samuels & Schroeder, 2019).

Beyond its accessibility framing, the project participates in a broader trajectory of AI-mediated musicianship in which machine learning functions as an adaptive mapping layer rather than as an autonomous generator of musical outputs. Where prompt-based and dataset-driven systems tend to flatten difference between users into a shared statistical space (Caramiaux & Donnarumma, 2021), interactive machine learning here is used to amplify difference: each performer's sonic preferences and bodily capacities can become design

variables rather than limitations. Recent work on AI-ADMIs emphasized user-specific customization, embodied interaction, and systems that adapt to the heterogeneous movement capacities of disabled musicians (Meyer et al., 2026), while warning that personalization alone does not eliminate normative assumptions or establish accessibility in practice. Following this framework, here we asked if an example-based machine learning workflow can support the creation of a personalized, live-playable gesture-to-sound mapping for an accessible granular instrument. This paper presents the design rationale, system architecture, and a preliminary functional evaluation of the prototype.

2. BACKGROUND AND RELATED WORK

2.1 Accessible Digital Musical Instruments

ADMIs are systems that facilitate music-making for users who encounter barriers with traditional instruments or conventional digital interfaces. Frid (2019) argues that accessible instruments should be evaluated not by technical novelty but by the extent to which they support agency, expressivity, and meaningful participation. In a later survey of electronic music-makers, Frid and Ilsar (2021) found that desired interfaces were often imagined as hybrids of gestural controllers, responsive mappings, simple interaction models, and broad synthesis access, with generally positive attitudes toward machine learning as a component of instrument design.

Other work has emphasized participatory design and individualized co-construction. Samuels and Schroeder (2019) argue that inclusive instrument design should begin from the strengths and desires of the user rather than from predefined assumptions about correct performance technique, and that improvisation can provide a strong framework for inclusive practice. A consistent finding across this literature is that customization is essential but difficult to scale: bespoke accessible instruments can be highly effective, yet they often require specialist expertise, extensive iteration, and maintenance infrastructures that are hard to sustain beyond a research project. This creates strong motivation for re-trainable, software-centered approaches that can adapt to users without new hardware built each time.

Duarte and colleagues (2023) extend this analytical lens through disability models, arguing that instrument design tends to encode implicit assumptions about what counts as standard embodiment, and that truly inclusive instruments must reconsider the interaction paradigms on which most digital instruments are built, a conclusion that motivates the present project's departure from visuo-centric design.

2.2 Visuo-Centric Paradigms and Their Limits

A significant portion of digital musical instruments is built around what can be described as visuo-centric interaction paradigms: models in which users navigate graphical representations of synthesis parameters and exercise control through manipulation of visible elements. While this offers legibility and precision for experienced users, it imposes dual-channel cognitive load and demands a level of technical literacy that is not equitably distributed (Duarte et al., 2023; Frid, 2019).

Beyond cognitive demands, visuo-centric paradigms can be particularly limiting for users with motor impairments, visual impairments, or atypical learning profiles. More fundamentally, they presuppose a specific form of engagement with music-making, one centered on control and planning rather than on listening, discovery, and progressive embodiment. The challenge is therefore not to eliminate visual feedback, but to design interaction models in which auditory perception constitutes the primary organizing logic of navigation: where the user moves and listens in a dynamic loop rather than watching a graphical representation and commanding a parameter change. This reorientation toward listening-centered interaction motivates both the choice of granular synthesis as the sound engine and the use of a two-dimensional preset space navigated through gesture and auditory feedback rather than direct parameter display

2.3 Mapping by Demonstration and Interactive Machine Learning

Mapping by demonstration has become one of the most influential paradigms in machine learning for musical interaction. Fiebrink and collaborators introduced *Wekinator* as a "meta-instrument" supporting on-the-fly learning, allowing users to iteratively generate examples, train, evaluate, and retrain within a single environment (Fiebrink et al., 2009; Vigliensoni & Fiebrink, 2024). This work was foundational because it repositioned machine learning from a hidden back-end optimization tool to an interactive component of instrument design itself. Subsequent work emphasized that the value of IML in music does not lie primarily in model accuracy but in how people evaluate, debug, and creatively shape learned systems in practice (Fiebrink, Cook, & Trueman, 2011; Visi & Tanaka, 2021).

Caramiaux and colleagues (2014) extended this line by examining the perceptual and corporeal dimensions of gesture-sound mapping, demonstrating that effective IML-based instruments should support users in discovering correspondences that feel natural rather than merely technically functional. This directly supports a key principle of the present project: that the gesture-sound relationship should emerge from the user's own embodied responses rather than from abstract parameter logic. Tanaka's work (2010) on biosignal-based instruments develops this further, arguing that the most expressive and accessible mappings emerge when the sensing modality aligns with the user's available and preferred movement repertoire, a principle motivating the multi-modal input layer described in Section 3.3.

In their reflection on sustained ML-based instrument practice, Fiebrink and Sonami (2020) argue that ML can help create complex "synthesis terrains" rather than merely precise control functions, while emphasizing the importance of flexibility, retraining, and continuous exploration over time. That perspective is especially relevant for a granular instrument, where the aim is to navigate a timbral field rather than trigger deterministic events. Related work by Scurto, Bevilacqua, and Françoise (2017) describes systems in which gesture-sound relationships emerge from physical demonstration, with an emphasis on improvisation and exploratory practice.

2.4 Sound Control and Accessible Precedents

The closest direct precedent for the present project is Sound Control (Parke-Wolfe et al., 2019), a Max/MSP environment that enables teachers and therapists to build custom musical interfaces for children with disabilities using IML, multiple sensors, and several sound generation modes including granulation. Their study showed that rich customization supported not only performance preparation but also agency, listening, movement, and social participation. The present project builds on that tradition but shifts the focus toward live granular performance, treats user sound preference as a first-class design input, and proposes a two-tier interaction model that explicitly negotiates the tension between accessibility and expressivity.

2.5 AI-ADMIs, Embodied Personalization, and Participatory Evaluation

Meyer et al. (2026) recently reviewed the development of AI-enabled ADMIs after Frid's 2019 review and identified user-specific customization, interactive machine learning, camera-based gesture recognition, and adaptive mapping as important developments in the field. They nevertheless caution that AI-ADMIs can reproduce exclusion through unrepresentative datasets, assumptions about normative bodies and repeatable movements, opaque model behavior, and insufficient involvement of disabled musicians in design and evaluation. Their Embodied AI Design Principles consequently emphasize situated, user-specific training, relational interaction, customizable multisensory feedback, and disability-led participatory design. The present prototype is aligned with this trajectory insofar as its mapping can be trained from individual examples and its input modality can be adapted to different movement repertoires. However, it does not yet satisfy the stronger participatory and longitudinal requirements articulated by Meyer et al. Its contribution is therefore limited to the implementation and critical examination of a candidate workflow. Whether this workflow produces accessible, expressive, or sustainable interaction will need to be determined through evaluation with disabled musicians and relevant practitioners.

3. MATERIALS AND METHODS

3.1 Research Design

This project was conducted as an early-stage design-and-feasibility investigation in accessible digital instrument design. The objective was to implement and functionally test a workflow in which a user could construct a personalized gesture-to-sound mapping for a granular synthesizer using example-based machine learning rather than manual programming. This prototype is thus presented as a candidate design for subsequent participatory refinement and evaluation.

3.2 System Architecture

The system was implemented in Max/MSP and deployed as a standalone application,[1] making it accessible to users without a Max/MSP license or programming background. The core uses FluCoMa (Tremblay et al., 2019) for machine learning and the Grainflow~ library[2] as the granular synthesis engine. While the present implementation uses granular synthesis, the underlying framework is not specific to this choice: any virtual or hardware synthesizer controllable via MIDI could serve as the output stage of the same IML pipeline. This is not incidental, matching the sound engine to a user's musical sensibilities constitutes a further dimension of the same personalization logic that drives the gesture-to-sound mapping workflow.

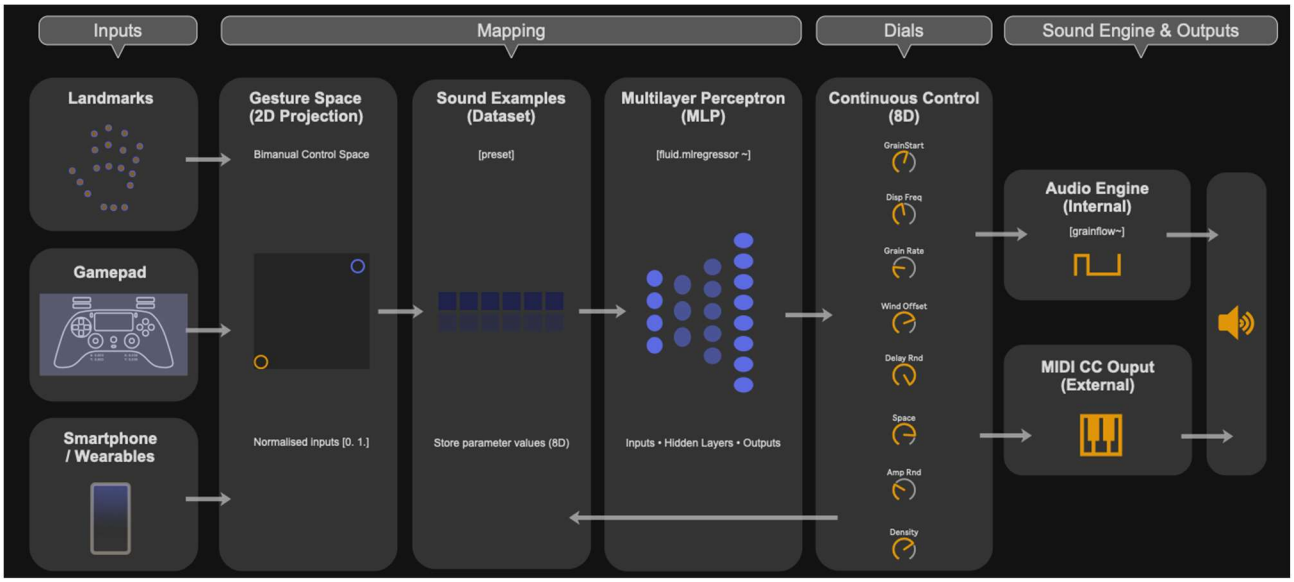

***Figure 1. System architecture**. Input from one of three modalities (webcam hand-tracking, gamepad, or smartphone/wearable) is normalized to a 2D control space and paired with eight-dimensional granular synthesis presets to train an MLP, which at inference maps live gesture to the synthesis parameters. The eight parameters drive an internal granular engine (grainflow~) and can also be routed as MIDI CC to external sound sources.*

[1] Source code repository: https://github.com/mikaelmolliex/drift-map

[2] Grainflow~ project repository: https://github.com/composingcap/grainflow

The architecture (Fig. 1) comprises five functional modules: (1) a standardized input acquisition layer; (2) a two-dimensional sound space and preset navigation system; (3) a dataset collection and model training module; (4) a granular synthesis engine based on Grainflow~; and (5) a live inference module.

The granular engine controls eight synthesis parameters as learning targets: Grain Start (read position within the source audio buffer), Dispersion Frequency (temporal spread of grain onset times around the base rate), Grain Rate (rate at which new grains are triggered), Window Offset (uniform temporal offset shaping overall texture density), Delay Randomization (unipolar random offset from the traversal phasor), Space (stereo or spatial spread of individual grains), Amplitude Randomization (unipolar random amplitude reduction per grain), and Density (probability that a grain will play). These eight parameters were selected to constitute a musically coherent and perceptually distinguishable timbral space: rich enough for exploratory listening but constrained enough to remain learnable from a small number of training examples. This reflects the IML principle that the output parameter space should be curated rather than exhaustive (Fiebrink & Sonami, 2020).

3.3 Input Modalities and Standardized Input Layer

The instrument can be controlled using any device that sends data via Open Sound Control (OSC), USB, or MIDI. Here we primarily used a gamepad, which offers physical tactile reference and ergonomic stability for users who benefit from a grounded interface. The system supports accelerometers (widely present in smartphones and smartwatches, capable of capturing tilt, acceleration, and orientation) or webcam-based hand or body tracking via MediaPipe (Lugaresi et al., 2019), which provides real-time hand landmark estimation for contactless gestural interaction without additional peripherals.

All modalities are normalized through a shared input layer that maps diverse sensor outputs into a standardized two-dimensional XY space. This serves two purposes. First, it makes different bodily interaction styles compatible with the same downstream mapping logic, supporting the principle that accessible interface design should adapt the system to the user rather than the reverse. Second, it concentrates the gestural degrees of freedom into a low-dimensional space that is navigable through listening without requiring visual attention. The sensing modality is therefore a user-centered design decision rather than a fixed default.

The two-dimensional sound space encodes granular synthesis presets at spatial positions, allowing the user to navigate continuously from one sonic configuration to another through gesture. Intermediate positions are computed through interpolation between preset anchors, producing smooth timbral transitions that support listening-centered exploration. Navigation does not require the user to understand the underlying synthesis parameters; instead, it relies on auditory feedback as the primary guide.

3.4 Two-Tier Interaction Model

A central design feature of the system is its two-tier interaction structure, which proposes a graduated approach to user agency in order to balance accessibility with expressive possibility.

In the **first tier (restricted mode)**, a predefined gesture dictionary is used as the basis for initial interaction. A set of gestural configurations has been designed in advance and associated with a curated selection of granular synthesis presets distributed across the 2D sound space. Users begin by selecting their preferred input modality and evaluating a series of sonic configurations. Their preferences are used to populate a personalized sonic map. Users are then invited to explore this space through gestures, without being explicitly told which gesture corresponds to which sound, the relationship is mediated by the trained model. This tier deliberately restricts the range of possible gestural actions to reduce calibration time and lower the entry barrier, accepting a temporary constraint on expressive agency in favor of rapid and accessible onboarding.

In the **second tier (open mode)**, users can define their own target sounds, import external audio, adjust synthesis parameters independently, connect to external systems via MIDI or OSC, and retrain the model from personalized gesture vocabularies. This tier restores full expressive agency for advanced users, performers, or facilitators wishing to adapt the instrument to specific contexts. It is also the tier in which the instrument functions most fully as an IML system in the classical sense: one in which the user actively participates in constructing the mapping through iterative demonstration and refinement.

The articulation between these two tiers enacts a position well-established in the ADMI literature: that accessibility and expressive agency are not inherently opposed, but must be negotiated through design, and that a graduated continuum of engagement offers more genuinely inclusive access than a single fixed interaction model (Frid & Ilsar, 2021; Samuels & Schroeder, 2019).

3.5 User Workflow

The full workflow is structured into four phases:

In the **first phase**, the user is presented with a set of short granular synthesis examples sampling a limited but varied timbral space. For each sound, the user indicates whether it was liked or disliked. This step treats musical taste as a first-class design variable.

In the **second phase**, only the preferred sounds are retained. Each preferred sound is replayed, and the user is asked to associate it with a movement using the chosen sensing modality. In restricted mode, movements are selected from the predefined gesture dictionary; in open mode, they may emerge from the user's own bodily vocabulary and comfort.

In the **third phase**, paired training data are collected. Each example consists of a movement feature vector and a target vector representing the synthesis parameter state of a preferred sound. Additional examples may be added iteratively to refine the mapping, following the standard IML logic of train–test–retrain (Fiebrink et al., 2009; Parke-Wolfe et al., 2019).

In the **fourth phase**, the trained model is deployed for live playing. Incoming movement data are analyzed continuously and passed through the trained MLP, which predicts synthesis parameters used to control the granular engine.

3.6 Machine Learning Model

Figure 2. *Instrument interface to select the multilayer perceptron parameters.*

A multilayer perceptron (MLP) was selected because the mapping from movement to synthesis parameters was expected to be nonlinear and multi-dimensional. Prior accessible music work has shown that neural networks can be effective for regression-based sound control from small to moderate datasets in real-time creative systems (Parke-Wolfe et al., 2019).

The MLP was implemented using FluCoMa within Max/MSP. The network comprised four input neurons (the normalized XY left and right coordinates of the input), three hidden layers of three neurons each, and eight output neurons (one per synthesis parameter). Hidden layers used a tanh activation; the output layer used a linear identity activation, allowing unbounded regression outputs that were subsequently clipped to valid parameter ranges. Training was conducted with a batch of 25 examples, adaptable to user needs, with default values of 2000 maximum iterations, learning rate 0.03, and no held-out validation split. Importantly, all of these parameters are exposed directly in the instrument's interface as editable fields (Fig. 2), allowing users or facilitators to adjust the architecture and training configuration without modifying any code. This degree of surface transparency is consistent with the progressive appropriation logic of the instrument as a whole.

During inference, sensor features were normalized, passed through the trained network, and converted into synthesis parameter predictions. Outputs could be bound to musically valid ranges via a routing matrix using linear, logarithmic, or exponential curves before being sent to the granular engine.

3.7 Evaluation Criteria

As an exploratory prototype, the evaluation in this study was technical and exploratory rather than participant-based. The following criteria served as a basis for functional evaluation and as a guide for future refinement: (1) whether the workflow could construct individualized mappings from selected sounds and gesture examples; (2) whether training and inference could be completed within the integrated Max/MSP environment; (3) whether the trained model supported real-time parameter updates; and (4) whether the learned mappings produced smooth or discontinuous behavior across the control space.

4. RESULTS

4.1 Prototype Implementation

The project resulted in a functional prototype (Fig. 3), capable of receiving movement data from multiple input modalities, presenting curated granular synthesis examples, collecting paired movement–sound data, training an MLP in FluCoMa, and applying the learned mapping to granular synthesis in real time. This prototype was presented informally as an interactive installation during a public science festival for children (Festival Eurêka)[3]. This deployment provided an empirical ecological demonstration that the system could operate in a public-facing setting and support musical interactions outside the development environment. Observations from this event were used to identify practical issues for subsequent iteration.

Figure 3. *Instrument interface to map the presets of the granulator synthesizer to the body gestures.*

4.2 Mapping Construction

The prototype implementation suggests that a user-specific mapping could be generated from two layers of personalization: first by filtering the target sound space according to user preference, and second by associating preferred sounds with self-chosen or predefined movements. This could produce a more personalized interaction than a fixed controller mapping because both the target sonic space and the gestural vocabulary can be partly determined by the user. The preference elicitation phase may then prove effective in reducing the size of the target sonic space to a manageable subset while preserving sufficient timbral variety for expressive navigation.

[3] https://festivaleureka.ca/en/activities/music-under-the-microscope/

4.3 Real-Time Control Behavior

Real-time control behavior was functionally operative during developer testing: incoming movement data produced continuous synthesis-parameter updates without interruptions that prevented interaction. However, the quality and coherence of the learned mapping proved uneven across the 2D sound space. With the present training set and network configuration, regions close to training examples generally exhibited smoother behavior, whereas some interpolated regions produced abrupt parameter changes and perceptible timbral discontinuities. In practice, this manifested as sudden, discontinuous timbral jumps: small gestural displacements produced disproportionately large parameter changes in some regions, resulting in perceptible sonic ruptures rather than smooth transitions. In areas more densely populated by training examples, the mapping was better supporting intentional navigation and incremental timbral adjustment. However, these discontinuities were not without expressive interest: in exploratory contexts they functioned as a form of compositional surprise rather than purely as technical failures. This ambivalence, between the glitch as defect and the glitch as resource, reflects a broader tension in IML-based instrument design between the desire for a well-behaved, controllable system and the recognition that partial opacity and non-linearity can generate forms of uncontrolled musical agency absent from purely linear systems.

4.4 Functional implementation of the Two-Tier Model

The two-tier model proved functional in preliminary testing. Restricted mode provided access to a predefined gesture dictionary and curated sound configurations, allowing the mapping workflow to operate with fewer initial configuration decisions. Open mode exposed sound import, synthesis adjustment, custom gesture collection, retraining, and external MIDI or OSC routing. The implementation therefore establishes that graduated levels of configurability can coexist within the same system.

5. DISCUSSION

5.1 Interactive Machine Learning as an Instrument-Building Tool

The present project demonstrates the technical feasibility of using interactive machine learning as part of an accessible instrument-design workflow centered on personalized mapping rather than automated musical output generation. This distinction matters because the broader discourse around AI in music sometimes overreaches, implying that only creative output itself can be delegated to a model. Here, the ML component is not the artist; it is part of the instrument-building process. This general orientation is not unique to the present work: prior IML research and recent AI-ADMI research have already established machine learning as a tool for adaptive mapping, instrument construction, and embodied interaction (Fiebrink et al., 2009; Meyer et al., 2026). The specific contribution of our prototype lies in combining preference-first sound selection, a two-tier interaction model, and live granular synthesis within a unified workflow.

5.2 Preference Selection as an Accessibility Strategy

One design hypothesis examined through the prototype is that preference selection may form part of an accessibility strategy. The literature on accessible music technology repeatedly foregrounds agency and choice, but many systems still assume that the facilitator or designer chooses the target sound set in advance (Parke-Wolfe et al., 2019). By allowing prospective users to select preferred sounds before gesture examples are collected, the present workflow treats musical preference as a design variable rather than a cosmetic addition. Although this proposition remains unvalidated: preference selection may increase aesthetic ownership, but it could also add decision-making demands or become burdensome for some users. Its value should therefore be evaluated rather than assumed. In any case, that shift changes what accessibility means in practice: accessibility here includes not only being able to produce sound, but being able to shape the sound world one enters, consistent with the framing in Frid and Ilsar (2021), where musicians with disabilities described their ideal interfaces as responding to their own aesthetic intentions rather than to a designer's assumptions.

The use of granular synthesis also deserves attention. Granular sound is well suited to this work because it supports rich, continuous, exploratory timbral interaction rather than only discrete note events. This aligns with prior findings that synthesis modes with broad timbral variability help users notice the effect of their actions and engage in exploratory listening (Parke-Wolfe et al., 2019), and resonates with the broader IML literature, where learned mappings are often valued less for precision than for their ability to open up expressive spaces difficult to program manually (Fiebrink & Sonami, 2020).

5.3 Loss Versus Transformation of Control

The introduction of ML-based mapping does not eliminate parametric control; it transforms it. Where control was previously explicit and linear, it becomes indirect and partially opaque: slight gestural variations can produce significant sonic changes, and the precise causal relationships between movement and sound are not immediately legible. From the user's perspective, this can feel like either a discovery or a frustration. This dual character is structural to any IML-based instrument and has been noted as a fundamental tension between the expressivity gains of learned mappings and their interpretive opacity (Fiebrink & Sonami, 2020; Frid & Ilsar, 2021).

The transition from control to navigation requires a different mode of engagement. For users who engage with the instrument through listening and physical exploration, this may be more natural and less demanding than parametric control. For users who rely on predictability, it may represent a genuine barrier. This suggests that the design of training data, gesture vocabulary, and spatial distribution of presets matter greatly for accessibility: a gesture dictionary whose correspondences are perceptually grounded (Caramiaux et al., 2014) will support progressive appropriation, while an arbitrary or poorly calibrated mapping may frustrate users regardless of how accessible the interface appears on paper.

5.4 Accessibility in Practice: Design Measures and Usage Contexts

The two-tier interaction model is a proxy to address accessibility through several design measures. In restricted mode, the system could reduce cognitive and motor demands by limiting active synthesis parameters, standardizing gestural input across capabilities through the shared input layer, providing auditory rather than visual feedback as the primary navigation guide, and offering preconfigured gesture-sound associations for immediate engagement. In open mode, the system could allow remapping, retraining, and extension, so that the instrument may be adapted to potential users at different stages of engagement or with different access needs.

These design choices suggest several possible usage contexts. Potential contexts include music therapy and special education, where the listening-centered, low-visual-demand interaction model could offer an alternative to keyboard or touchscreen interfaces for users who find those modalities inaccessible. Smooth timbral transitions could also support sustained engagement and sensorimotor feedback loops, though sudden textural shifts may be perceived as aggressive in some contexts and as richness in others. In recreational or community music-making, the two-tier model could support onboarding for users without musical background while remaining expressive for more experienced users. In experimental performance, the emergent and partially non-deterministic character of the learned mapping could be used as a compositional resource. These contexts impose different and sometimes conflicting requirements: therapeutic and educational settings may prioritize predictability, facilitator control, sensory comfort, and repeatability, whereas experimental performance may accommodate or value discontinuity and partial unpredictability. The two-tier model should therefore be understood as a flexible framework for addressing this tension.

6. LIMITATIONS

6.1 Model Opacity and Post-Training Editability

Learned mappings are difficult to refine precisely because mistakes are not locally editable: users cannot adjust a single relationship between a gesture and a sonic outcome without regenerating training data and retraining the model. Frid and Ilsar (2021) report exactly this concern from musicians who found ML-based mapping promising but too opaque to tweak after training. The design of post-training editing tools, which enable users to identify zones of the sound space with unexpected mapping behaviors and to provide targeted corrective examples, represents a significant direction for future development.

6.2 Input Modality and Embodied Grounding

Each supported input modality carries distinct affordances and limitations. Physically grounded controllers such as gamepads provide tactile reference and ergonomic stability but constrain the gestural vocabulary to what the device affords. Body-mounted sensors such as accelerometers offer freedom of movement but no fixed spatial reference. Contactless sensing through computer vision is non-invasive but lacks haptic feedback, making it difficult for users to feel the boundary of the control space or reliably reproduce a gestural configuration. The choice of modality should therefore be a user-centered design decision rather than a system default, with each modality empirically evaluated against representative user groups. Ideally, multisensory feedback could be tailored to each user's sensory profile, leveraging their strengths rather than compensating for the absence of a default modality. In this view, the input layer itself participates in the personalization logic of the instrument.

7. CONCLUSION

This article has presented a prototype for an accessible digital musical instrument for personalized gesture-to-sound mappings for granular synthesis using Max/MSP, FluCoMa, and the Grainflow~ library. The system integrates sound-preference selection, a standardized multi-modal input layer, a listening-centered two-dimensional sound space, a two-tier interaction model, and MLP-based regression to create a live-playable mapping tailored to each performer's movement possibilities and sonic interests.

This prototype aims to make three contributions. First, operationalizing accessibility as a question of adaptable mapping rather than alternative hardware alone, connecting recent ADMI literature with IML practice in a unified design workflow. Second, implementing a graduated two-tier model intended to negotiate initial ease of access and subsequent configurability. Third, exploring a listening-centered interaction logic in which auditory feedback can guide navigation without requiring direct manipulation of synthesis parameters.

More broadly, the approach points toward a use of AI in music that runs counter to the dominant generative paradigm: instead of producing many tracks from one model, it produces one instrument from each user. In a moment defined by hyper-reproduction, this is a deliberate move toward instruments that adapt to people rather than asking people to adapt to the instruments, opening space for musical engagement across a wider range of bodies, capacities, and contexts, with potential implications for artistic performances, music therapy, education and accessibility.

ACKNOWLEDGMENTS

We thank Dominic Thibault and David Piazza for their support and feedback during this work.

AUTHOR DECLARATIONS

The authors declare no financial or non-financial conflicts of interest. A large language model (ChatGPT 5.6) was used during the preparation of this manuscript to improve the clarity and grammatical accuracy of the English written by the non-native English-speaking authors.

REFERENCES

Caramiaux, B., & Donnarumma, M. (2021). Artificial Intelligence in Music and Performance: A Subjective Art-Research Inquiry. In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and*

Developments for Creativity (pp. 75–95). Springer. https://doi.org/10.1007/978-3-030-72116-9_4

Caramiaux, B., Françoise, J., Schnell, N., & Bevilacqua, F. (2014). Mapping through listening. *Computer Music Journal*, *38*(3), 34–48. https://doi.org/10.1162/COMJ_a_00255

Duarte, E. G., Cossette, I., & Wanderley, M. M. (2023). Analysis of Accessible Digital Musical Instruments through the lens of disability models: a case study with instruments targeting d/Deaf people. *Frontiers in Computer Science*, *5*, 1158476. https://doi.org/10.3389/FCOMP.2023.1158476

Fiebrink, R., Cook, P. R., & Trueman, D. (2011). Human model evaluation in interactive supervised learning. *Conference on Human Factors in Computing Systems - Proceedings*, 147–156. https://doi.org/10.1145/1978942.1978965

Fiebrink, R., & Sonami, L. (2020). Reflections on eight years of instrument creation with machine learning. *In Proceedings of the International Conference on New Interfaces for Musical Expression (pp. 237-242).* https://zenodo.org/records/4813334

Fiebrink, R., Trueman, D., & Cook, P. R. (2009). A Meta-Instrument for Interactive, On-the-Fly Machine Learning. *In Proceedings of the International Conference on New Interfaces for Musical Expression*. https://doi.org/10.5281/zenodo.1177513

Frid, E. (2019). Accessible Digital Musical Instruments—A Review of Musical Interfaces in Inclusive Music Practice. *Multimodal Technologies and Interaction 2019, Vol. 3, Page 57*, *3*(3), 57. https://doi.org/10.3390/MTI3030057

Frid, E., & Ilsar, A. (2021). Reimagining (Accessible) Digital Musical Instruments: A Survey on Electronic Music-Making Tools. *International Conference on New Interfaces for Musical Expression*. https://doi.org/10.21428/92FBEB44.C37A2370

Larsen, J. V., Overholt, D., & Moeslund, T. (2016). The Prospects of Musical Instruments For People with Physical Disabilities. *New Interfaces for Musical Expression*. https://doi.org/10.5281/ZENODO.1176056

Lugaresi, C., Tang, J., Nash, H., Mcclanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., Grundmann, M., & Research, G. (2019). *MediaPipe: A Framework for Building Perception Pipelines*. https://arxiv.org/abs/1906.08172v1

Magnusson, T. (2019). Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions. Bloomsbury Academic.

Meyer, G., Schroeder, F., & Rainer, A. (2026). Embodied AI and the Future of Accessible Music-Making: A Review of Developments in AI-ADMIs Post-frid. In *International Conference on ArtsIT, Interactivity and Game Creation* (pp. 93-112). Springer, Cham.

Parke-Wolfe, S. T., Scurto, H., & Fiebrink, R. (2019). *Sound Control: Supporting Custom Musical Interface Design for Children with Disabilities*. https://doi.org/10.5281/ZENODO.3672920

Samuels, K. (2019). The meanings in making: Openness, technology and inclusive music practices for people with disabilities. *Leonardo Music Journal, 25, 25-29.* https://doi.org/10.1162/LMJ_a_00929

Samuels, K., & Schroeder, F. (2019). Performance without Barriers: Improvising with Inclusive and Accessible Digital Musical Instruments. *Contemporary Music Review*, *38*(5), 476–489. https://doi.org/10.1080/07494467.2019.1684061

Scurto, H., Bevilacqua, F., & Françoise, J. (2017). Shaping and Exploring Interactive Motion-Sound Mappings Using Online Clustering Techniques. *In Proceedings of the International Conference on New Interfaces for Musical Expression*. https://doi.org/10.5281/zenodo.1176270

Tanaka, A. (2010). *Mapping Out Instruments, Affordances, and Mobiles*. NIME. https://doi.org/10.5281/ZENODO.1177903

Tremblay, P. A., Green, O., Roma, G., & Harker, A. (2019, September). From collections to corpora: Exploring sounds through fluid decomposition. In *International Computer Music Conference and New York City Electroacoustic Music Festival* (pp. 223-228). International Computer Music Association. https://www.flucoma.org/

Vigliensoni, 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

Visi, F. G., & Tanaka, A. (2021). Interactive Machine Learning of Musical Gesture. In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (pp. 771–798). Springer. https://doi.org/10.1007/978-3-030-72116-9_27