CamJam: A Modular Collaborative and Accessible Digital Musical Interface

Frej Lorenzen
fslo21@student.aau.dk
Aalborg University
Copenhagen, Denmark

Kevin Hansen
ktha20@student.aau.dk
Aalborg University
Copenhagen, Denmark

Emil Zawistowski
lj10rb@student.aau.dk
Aalborg University
Copenhagen, Denmark

Eirini Liapikou
eliapi24@student.aau.dk
Aalborg University
Copenhagen, Denmark

Abstract

This paper presents CamJam, a modular, collaborative, and accessible digital musical interface designed to support inclusive music-making. The system enables musicians and non-musicians to engage in shared sound creation through a set of embodied, camera-based interaction modules, each contributing a distinct musical role within a synchronized, loop-based structure. CamJam was developed as a no-fail collaborative environment, in which harmonic constraints and temporal synchronization support intuitive participation without requiring prior musical training. The project explores how camera-based sensing can facilitate expressive yet approachable musical control grounded in embodied interaction design. The system has been explored through iterative testing and group-based user studies involving participants with diverse musical backgrounds. Observations from these studies suggest that CamJam supports intuitive engagement and collaborative play across differing levels of musical experience.

Keywords

ADMI, Collaborative Music, Camera-based Interaction, Inclusive Design, OSC

1 Introduction

Digital Musical Instruments (DMIs) have increasingly focused on accessibility and collaboration, yet many systems remain challenging for non-musicians or individuals with disabilities. CamJam addresses this gap by combining camera-based interaction with collaborative design principles to create an accessible entry point for collective music-making.

The system consists of four interaction modules: a gesture-controlled drum machine, an emotion-driven chord sequencer, a plucked string instrument using hand tracking, and a contrast-based synthesizer. Each module employs embodied interaction principles [2], mapping physical actions to sonic parameters through intuitive metaphors that require no specialized musical knowledge.

CamJam builds on prior work in Accessible Digital Musical Instruments (ADMIs) and Collaborative Digital Musical Instruments (CDMIs), particularly systems like LoopBoxes [3] and Sound Forest [7] that emphasize inclusive participation. Our contribution lies in demonstrating how purely camera-based sensing can support collaborative music-making without requiring specialized controllers or tangible interfaces, while maintaining low hardware requirements and portability for installation contexts.

2 Related Work

Research in accessible music technology has explored various approaches to lowering barriers to musical participation. This section examines key precedents in tangible instruments, multisensory installations, and collaborative accessible systems.

2.1 Multisensory Installations

Sound Forest [7] employed five interactive light-emitting strings with multisensory feedback to engage users across abilities. The installation combined haptic platforms, visual cues through mirrors creating forest illusions, and aural feedback. While it demonstrated strong engagement across abilities, its reliance on predefined sample playback limited expressive freedom. The evaluation methodology used picture cards to match participants' associations with composers' keywords, effectively testing alignment but not deeply probing accessibility or usability. Our work extends this by implementing generative synthesis techniques (physical modeling, real-time parameter control) for greater expressive range.

2.2 Modular Collaborative Systems

LoopBoxes [3] demonstrated the effectiveness of modular, "no-fail" loop-based systems for children with special educational needs. The system featured three simple plug-and-play modules supporting solo or collaborative loop-based music making. Teacher surveys confirmed preferences for robust, intuitive units with clear affordances supporting collaborative play. Clear role mappings helped users understand their contribution to the ensemble. We adopt similar principles: modular architecture, loop-based synchronization, and clear roles, but implement them through camera-based rather than button-based interaction.

2.3 Assistive Technology for Motor Impairments

DuoRhythmo [8] extended collaborative accessibility to individuals with ALS through eye-tracking and head-mouse interaction, achieving strong usability scores (SUS: 79.5/100). The system allowed people with ALS to make music collaboratively in real-time, reducing loneliness and stimulating intellectual needs. This work informed our attention to low-latency interaction and the importance of multimodal feedback for accessibility.

NIME '26, June 23–26, 2026, London, UK

2.4 Tangible Interfaces for Visual Impairment

The Tangible Signals project [11] explored tangible user interfaces (TUIs) for visually impaired musicians, testing three physical devices controlling accessible music software. The project surveyed related vision-based interfaces, including EyeHarp, a gaze-controlled instrument developed for ALS patients that allows users to learn, perform, and compose music using only eye movements. Expert feedback from live performances indicated that embodied metaphors can reduce cognitive load significantly. While our system uses camera-based rather than tangible input, we share the goal of reducing cognitive load through embodied metaphors.

2.5 Design Considerations for ADMIs

Research on ADMIs [6] shows that adaptability and customization play crucial roles, alongside including target groups in design processes. Many current ADMIs focus primarily on physical disabilities and children, with multimodal feedback often unexplored. Most are bimodal (auditory-visual), with few emphasizing haptic feedback. Many implementations rely on MIDI, which offers lower data resolution compared to protocols like OSC and HID: we therefore employ OSC throughout CamJam. Studies of ideal interfaces for electronic music-making revealed that musicians desire systems combining existing technologies, providing haptic feedback, being intuitive, and supporting learning.

2.6 Camera-Based and Vision-Based Systems

Vision-based and camera-tracking musical interfaces have a substantial history in NIME and related fields. Paine's taxonomy of real-time interfaces for electronic music performance [9] and the systematic review by Fraisse et al. [4], which analysed 181 peer-reviewed publications on interactive sound installations, together document a wide field in which vision-based sensing figures prominently, spanning optical flow, depth sensing, hand and body pose estimation, gaze tracking, and facial expression analysis. Early examples of camera-based interactive sound in public and collaborative contexts include Franinovic and Visell's Recycled Soundscapes [5], which used camera-based interfaces for environmental sound capture and collaborative collage. Among the most influential accessibility-focused examples is the AUMI (Adaptive Use Musical Instruments) project initiated by composer Pauline Oliveros, which uses webcam-based motion detection to enable people with severe physical disabilities to make music through subtle body movements [1]. More recently, MediaPipe [12] has enabled robust real-time hand and body tracking on commodity hardware, lowering barriers to camera-based DMI development considerably. CamJam builds on this lineage, contributing an integration of multiple heterogeneous camera-based modules (gesture tracking, facial expression recognition, and visual contrast analysis) within a single modular, collaborative system, and evaluating that integration with a diverse participant pool.

3 System Design

3.1 Design Requirements

CamJam was developed according to functional and non-functional requirements derived from an analysis of existing ADMIs. Key functional requirements include entangled modules with interdependent interactions, embodied interactions engaging the full body, melodic and percussive instruments, harmonic constraints based on musical scales, chord progression capability, loop-based sequencing for synchronization, and clear role mapping.

Non-functional requirements address concurrent interaction by multiple users, accessibility for children and users with disabilities, low end-to-end latency (≤15 ms, as an embodied-interaction design requirement), multimodal feedback, portability for installation contexts, a child-safe and robust design, robustness under indoor lighting conditions, a constructed installation housing, aesthetically pleasing material design, and an appropriate attention span and learning curve.

3.2 Hardware and Software

The system employs two MacBook Pros connected via Ableton Link for clock synchronization, with additional displays including an iPad and iPhone via Continuity Camera, and a Logitech C920 webcam. The circular installation design enables collaborative engagement, with cameras and screens oriented outward toward users. All interaction modules run in Max/MSP and Max for Live, communicating via the OSC protocol.

Python scripts using MediaPipe handle computer vision tasks, sending hand landmark coordinates and facial expression data to Max patches. Sound synthesis combines drum samples, Faust-based physical modeling implementing Karplus-Strong string synthesis, and oscillator-based synthesis for the chord sequencer. All instruments operate within a C major scale to ensure harmonic coherence.

The purely software-based approach supports reproducibility across different hardware configurations and facilitates portability in installation contexts. The Continuity Camera feature on Apple Silicon laptops was found to be equal to or better than wired camera setups in terms of latency during informal development testing. The ≤15 ms end-to-end latency figure is a design requirement derived from embodied-interaction literature; formal end-to-end measurement is identified as future work.

3.3 Interaction Modules

Drum Machine. Users control drum pattern complexity and timbral parameters through hand height in three vertical zones corresponding to kick, snare, and hi-hat. Hand position modulates envelope parameters (attack/decay) and filter cutoffs in real-time while a preset 8-step loop plays synchronized to the master clock via Ableton Link. MediaPipe tracks both hands, with the system routing gestures to different drum parameters based on horizontal position across 100 normalized coordinate points. The camera feed is split into three distinct zones with drum images providing visual affordances that guide the user's spatial navigation. The mapping logic is designed to minimize cognitive load, allowing users to focus on rhythmic intensity rather than precise triggering. Drum sounds are synthesized using impulse-triggered envelopes with frequency modulation for the snare, sine waves for the kick, and filtered noise for the hi-hat. All tracking data is transmitted via Open Sound Control (OSC) to Max for Live, ensuring low-latency response between physical movement and sonic output.

Contrast-Based Synthesizer. Inspired by Iannis Xenakis's UPIC system, this module allows users to generate sound through graphical interaction. Users place dark objects (paper, hands) on a white surface captured by an overhead camera. The system utilizes OpenCV to extract a horizontal scanline of 100 pixels from the video feed, mapping these brightness values to oscillator waveform amplitudes in a custom wavetable. This process creates

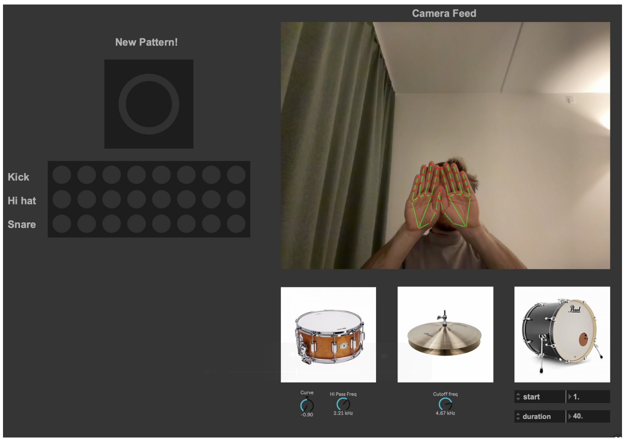

Figure 1: Drum machine interaction: hand height in three vertical zones controls kick, snare, and hi-hat parameters in real-time.

a form of visual wavetable synthesis where the physical shape of the object directly dictates the harmonic content of the timbre. A four-value ADSR envelope derived from another scanline controls filter cutoff modulation, though this was simplified during testing to provide a more immediate "plug-and-play" experience for novices. This tangible interaction provides a bridge between physical manipulation and digital sound generation, grounding abstract synthesis concepts in a visible, graspable medium.

Figure 2: Contrast-based synthesizer: dark paper shapes placed on white background modulate oscillator waveform via brightness scanning.

Emotion Recognition & Chord Sequencer. This module acts as the harmonic anchor for the collaborative ensemble. It utilizes a Mini-Xception CNN model, pretrained on the FER2013 dataset, to perform real-time facial expression recognition. The system maps eight detected emotions (angry, disgust, fear, happy, sad, surprise, neutral, contempt) to specific chord progressions within a C major scale, ensuring that the entire group remains harmonically aligned. To prevent "sonic jitter" caused by rapid facial micro-expressions, a moving average filter of three frames is applied to the classifier's output. The system arpeggiates these chords across four beats synchronized via Ableton Link, creating a direct interdependence between a user's affective state and the collective harmonic movement. A bass synthesizer reinforces the progression by playing the root note via a sawtooth oscillator with low-pass filtering. Visual feedback provides a real-time confidence score for the detected emotion, allowing the user to consciously modulate their expression to influence the group's musical mood. During evaluation, some participants reported discomfort with direct facial expression tracking and limited module responsiveness. Future iterations should consider offering abstract or avatar-based representations as alternative interaction modes to improve inclusivity.

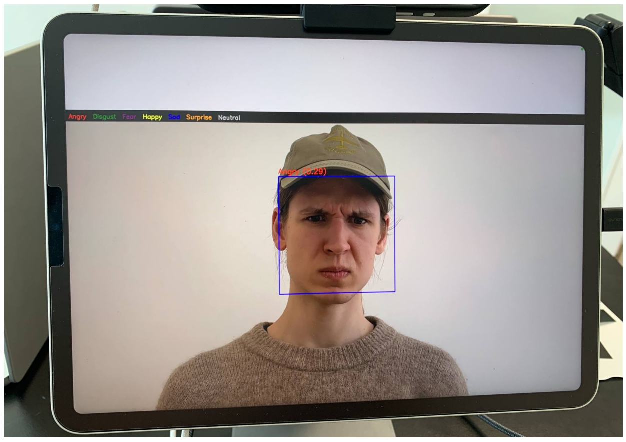

Figure 3: Interface mapping facial expressions to chord progressions with on-screen emotion display.

Plucked String. The Plucked String module employs physical modeling via the Karplus-Strong algorithm, implemented in the Faust functional programming language for high-efficiency signal processing. Vertical hand position maps to pitch (quantized to a C major scale, C4-B4), while plucking gestures are detected as the index finger crosses the frame's horizontal midpoint. Depth (z-axis) information is crucial for accessibility; the system defines a 3D "interaction box," where only gestures within a specific distance from the camera trigger excitations, effectively filtering out background movements. Gesture velocity is calculated in real-time based on the 3D displacement of the hand between frames, which is then mapped to the plucking intensity and the feedback coefficient of the Karplus-Strong model, influencing both amplitude and damping. To ensure harmonic cohesion with other modules, a frequency offset correction (scaled +1 to +3 Hz) is applied to compensate for the inherent slight detuning of the physical model. Users can engage both hands simultaneously, interacting with two independent string instances for polyphonic performance.

3.4 Sonification Strategy

The interactions emulate a traditional band structure: percussion (drum module), rhythm/harmony (chord sequencer with bass), and melodic soloist (string). The synthesizer serves dual roles, with emotion control affecting chord progressions that influence all other instruments through shared harmonic structure. Loop-based sequencing (8 steps for drums, 4 beats for chords) provides temporal constraints supporting "no-fail" participation: users cannot play out of sync. The C major scale constraint ensures harmonic compatibility between melodic improvisations and chord progressions.

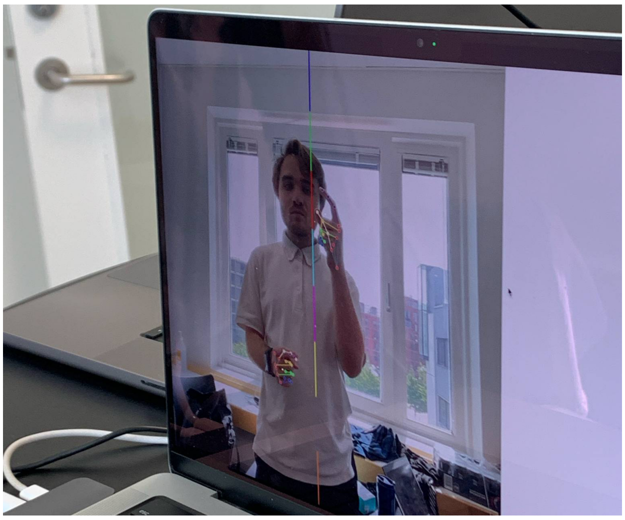

Figure 4: Plucked string interaction: vertical hand position maps to pitch, with plucking gestures detected via index finger motion across the frame midpoint.

Figure 5: CamJam installation showing four interaction stations arranged in a circle, enabling participants to see each other during collaborative performance.

This framework prioritises inclusivity through plug-and-play accessibility for both trained musicians and novices alike. The modular design allows flexible role assignment and rotation, encouraging exploration of different musical functions during collaborative sessions.

4 Evaluation

We conducted a mixed-methods evaluation with 36 participants (ages 18-34) recruited from a university in Northern Europe. Participants were classified as musicians (n=16) if they had attended music university or played an instrument for over one year, or non-musicians (n=20) otherwise. Most musicians met more stringent criteria: playing or singing within the last five years with at least three hours per week of practice during at least one of those years.

Accessibility is assessed at the level of user perception rather than through direct interaction with the primary target groups (children and users with disabilities). This was due to the convenience sampling approach: institutions working with relevant communities were contacted but did not respond to the inquiry prior to the evaluation. All participants provided informed consent for data processing, data were anonymized with unique participant IDs, and demographic data were collected in accordance with GDPR.

A pilot test with four university students from another program identified several issues that were addressed before the main evaluation: the signup and questionnaire were combined, sensitive information was removed, a QR code was created for mobile completion, interactions were tuned for stability, and testing procedures were clarified to reduce participant confusion.

Testing occurred in groups of four in a small room at the university campus. The room was well-lit (important for gesture recognition) with subtle reverberation. Four cameras (two MacBook built-in, one iPhone Continuity Camera, one Logitech C920 on tripod) pointed outward in 90-degree intervals. Participants positioned themselves in a circle, able to see their screens and one another. Four experimenters positioned behind participants individually observed one participant each.

Each group received a brief description and instruction on each module, with emphasis on collaborative music creation and module rotation. The test consisted of eight one-minute blocks where each participant interacted with a single module, rotating clockwise through all four modules twice. If participants needed help, assistance was provided. Following this sequence, groups engaged in three minutes of free improvisation with stronger encouragement for collaboration, module exchange, and short solos.

Participants completed evaluation questionnaires with both closed-ended (11 Likert-scale items on intuitiveness, ease of use, and collaboration) and open-ended questions. Experimenters filled observation questionnaires focused on learning speed, communication patterns, preferences, and frustration signs.

4.1 Data Analysis

Mann-Whitney U tests compared musicians and non-musicians on composite scores for ease of use and collaboration. The Aligned Rank Transform (ART) ANOVA examined intuitiveness ratings across four interaction types, testing main effects of musical background (between-subjects), interaction type (within-subjects), and their interaction. Item-level Mann-Whitney tests with Bonferroni correction explored individual questions. Open-ended responses and observational data underwent qualitative analysis following Popping's framework [10], with both a priori and inductive coding approaches.

5 Results

5.1 Quantitative Results

Mann-Whitney U tests revealed no statistically significant differences between musicians and non-musicians on any measure (all $p > 0.10$). Composite scores for ease of use (median 3.4 for both groups) and collaboration (median 2.5 musicians, 3.0 non-musicians, p=0.77) indicated comparable experiences across musical backgrounds. The highest ratings were for "How easy was it for you to start using the system?" (median 4 for both groups), though this dropped to 3 for "How intuitive was the system overall?" The lowest rating was for "How easy was it to collaborate with others using this system?" (median 2 for musicians, 3 for non-musicians).

Aligned Rank Transform ANOVA showed significant differences between interaction types (F(3,102)=11.8, p<.001) but no interaction effect with musical background (F(3,102)=0.70, p=0.55). Mean intuitiveness ratings varied substantially: String (M=4.28, SD=0.81), Emotion (M=3.33, SD=1.20), Drums (M=3.19, SD=1.14), Paper (M=2.92, SD=1.08). The string interaction's high rating and low standard deviation indicate consistent positive reception, while the emotion module's higher variability reflects technical limitations in facial tracking reliability.

5.2 Qualitative Findings

Open-ended responses identified control, predictability, and precision as both the most valued and most lacking features. System responsiveness issues and complex interaction mechanisms were frequently cited as barriers (15 occurrences across groups). Specifically, "Lack of Control & Predictability" was the most common difficulty factor (10 occurrences), while "Complex Interaction/Understanding Mechanisms" ranked second (9 occurrences). "Technical Precision & Responsiveness" was the most frequently mentioned facilitator for accessibility (9 occurrences).

However, 81% of participants rated the system as suitable for users with disabilities or children, with non-musicians particularly enthusiastic (90% vs. 69% of musicians rated it as highly or moderately suitable). These figures reflect perceived suitability as reported by a convenience sample of university students aged 18–34; none of the participants were recruited as members of the primary target groups (children or users with disabilities), and this question should therefore be read as a proxy perception measure rather than direct validation. Non-musicians valued "Intuitive/Natural Interaction" and "System Assistance/Pre-existing Structure," while musicians more frequently cited "System Limitations/Missing Features" as hindrances.

Collaboration posed challenges: "Lack of Understanding/Focus on Individual Instrument" was mentioned by 13 participants as hindering intuitive collaboration. Similarly, "Instrument-Specific Difficulties/Lack of Understanding" was the most cited challenging factor for multi-user interaction (16 occurrences). Time constraints were mentioned only twice, suggesting the 11-minute testing duration was adequate for basic exploration.

The string interaction received strong preference among non-musicians (8/20) for its "ease of use/intuition/simplicity," while musicians favored drums (6/16), citing "control/predictability/precision". This suggests that resemblance to familiar instruments (guitar-like string plucking) may enhance intuitiveness for novices, while musicians appreciate more complex parameter control.

5.3 Observational Results

Direct observation revealed diverse engagement patterns across 34 participants (two observation sheets were entirely blank and excluded from this analysis). Most participants (52.9%) needed one round of interaction plus guidance beyond the introduction before understanding modules, while 41.2% learned quickly with minimal guidance. Only 5.9% failed to learn interactions throughout the test. Regarding collaboration, 44.1% showed no participant interaction, 32.4% showed some interaction, and only 23.5% demonstrated active collaboration. Preference patterns were apparent for 41.2% of participants, while 58.8% showed no clear preference. Most participants (55.9%) showed no apparent struggle, though 38.2% exhibited some frustration and 5.9% significant frustration.

Specific observations highlighted facial tracking as frequently problematic, with users reporting discomfort or limited responsiveness. The drum and paper modules often required additional explanation due to subtle feedback. Where collaboration occurred, it was often unstructured and lacked coordination, indicating the system does not inherently facilitate cooperative use despite its modular structure.

6 Discussion

The lack of significant differences between musicians and non-musicians validates the embodied interaction approach and confirms that camera-based sensing can support accessible music-making. The string interaction's high ratings (M=4.28) despite being the only module resembling a traditional instrument suggests that metaphorical clarity: plucking produces discrete notes, matters more than physical tangibility.

However, the moderate overall ease-of-use scores (3.4/5) and low collaboration ratings (2.5-3.0/5) indicate room for improvement. Participants valued control and predictability above all else, yet these were precisely what the system lacked. The emotion recognition module's variability (SD=1.20) reflected technical limitations in facial tracking, with some participants reporting the system failed to recognize their expressions or feeling uncomfortable with direct facial monitoring.

The 81% suitability rating for users with disabilities and children is an encouraging outcome; however, it reflects perceived suitability as reported by a convenience sample of university students aged 18-34, none of whom were recruited as members of those target groups. Accessibility is therefore assessed at the perception level in this study, not through direct involvement of the intended communities. Future work should engage disabled musicians and children directly in co-design and evaluation: this was intended for the present project, but the institutions contacted did not respond to the inquiry.

CamJam's most significant shortcoming was fostering genuine collaboration. While the entangled modules created technical interdependence: chord changes affect all players, drum patterns provide rhythmic foundation, this did not translate to social engagement. Observational data showed that 44.1% of participants did not interact with each other during the session. The open-ended responses confirmed this: "Lack of Understanding/Focus on Individual Instrument" was cited by 36% of participants as hindering collaboration.

Several factors likely contributed to weak collaborative engagement. First, the short testing duration (11 minutes structured interaction, 3 minutes improvisation) may have been insufficient for developing collaborative musical relationships. Second, the instruction emphasized individual module understanding rather than ensemble coordination strategies. Third, the circular physical arrangement enabled visual contact but provided no explicit collaborative affordances: no visual indicators showing how one player's actions enabled or constrained another's, no role-rotation prompts, no shared visual representation of the ensemble sound.

The string interaction's success (highest ratings, lowest variability) offers design insights. Its intuitive mapping: vertical position to pitch, crossing midpoint to trigger, created clear cause-effect relationships. The resemblance to guitar-playing provided a familiar metaphor for both musicians and non-musicians. The discrete note events (versus continuous parameter control in other modules) made musical contributions more legible. Physical modeling synthesis provided natural-sounding timbral evolution, reinforcing the instrument metaphor.

Conversely, the paper synthesizer's lower ratings (M=2.92) reflected several issues. The mapping from visual contrast to waveform shape was conceptually interesting but perceptually opaque: users could not predict sonic outcomes from physical actions. The overhead camera angle made it difficult to see what they were doing. The continuous parameter space (versus discrete pitch zones in the string) required more precise control. These findings suggest that for accessible DMIs, discrete mappings with clear perceptual correspondence outperform continuous abstract mappings.

The moderate-to-low ratings across most measures align with findings from Sound Forest and LoopBoxes evaluations, which similarly identified tensions between simplicity and expressivity. Sound Forest's predefined samples ensured aesthetic coherence but limited user agency. LoopBoxes' button-based sequencers provided precise control but required understanding of rhythmic patterns. CamJam attempted to balance these through gesture-based control of generative synthesis, but achieved only moderate success.

7 Conclusions and Future Work

CamJam demonstrates that camera-based sensing can support accessible, embodied musical interaction without requiring specialized controllers or musical training. The system achieved comparable usability across musical backgrounds while maintaining a low barrier to entry. However, true collaborative music-making requires more than concurrent access: it demands design mechanisms that explicitly scaffold listening, turn-taking, and coordinated expression.

Future work will focus on three areas: (1) enhancing real-time responsiveness through optimized computer vision pipelines and automatic sensor calibration, including formal end-to-end latency measurement, (2) implementing explicit collaboration cues such as role-rotation prompts, visual indicators of interdependence, and pleasant chimes reinforcing successful coordination, and (3) conducting longitudinal studies in ecologically valid contexts (festivals, exhibitions) with diverse populations including children and individuals with disabilities.

Technical improvements should address better-trained emotion recognition models capturing diverse expressions and ethnicities, alternative emotion representations (avatars or abstract modes) for users uncomfortable with facial tracking, and enhanced visual feedback (animated drum hits, string motion, waveform evolution). The modular architecture enables rapid iteration, and the purely software-based approach ensures reproducibility across different hardware configurations.

Evaluation methodologies should be expanded to include standardized measures (System Usability Scale, enjoyment scales), task-based assessments (playing specific melodies), and comparison with non-DMI baselines to better understand the system's actual advantages for non-musicians. Testing should also examine whether the string interaction's success reflects resemblance to familiar instruments or simply clearer mappings, informing future design philosophy.

8 Ethical Standards

This study followed institutional ethical guidelines for student research projects at Aalborg University. Formal ethics board review was not required under Danish university regulations for this type of non-medical, non-deceptive study with adult participants. All participants provided informed consent prior to participation, explicitly covering data collection, anonymization, and use of the data for academic reporting. Collected data were anonymized, with each participant assigned a unique ID; no personally identifiable information was retained after anonymization. Demographic data, including information about musical skills, disabilities, and health conditions, were collected and stored in accordance with GDPR. Sensitive fields (e.g. health conditions) were removed from the questionnaire following the pilot study to minimize unnecessary data collection. Participants were free to withdraw at any time without consequence.

Acknowledgments

We thank all participants who volunteered their time for this study. This work was supervised by Daniel Overholt.

References

[1] Thomas Ciufo, Abby L. Dvorak, Kip Haaheim, Jennifer Hurst, Grace S. Leu, Leaf Miller, Diane Mitchell, Sherrie Tucker, and Ellen Waterman. 2024. *Improvising Across Abilities: Pauline Oliveros and the Adaptive Use Musical Instrument.* University of Michigan Press / OAPEN. https://doi.org/10.3998/mpub.12711014

[2] Paul Dourish. 2004. *Where the Action Is: The Foundations of Embodied Interaction.* MIT Press, Cambridge, MA.

[3] Andreas Förster et al. 2023. LoopBoxes – Evaluation of a Collaborative Accessible Digital Musical Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).*

[4] Valérian Fraisse, Marcelo M. Wanderley, and Catherine Guastavino. 2021. Comprehensive Framework for Describing Interactive Sound Installations: Highlighting Trends through a Systematic Review. *Multimodal Technologies and Interaction* 5, 4 (2021), 19. https://doi.org/10.3390/mti5040019 Visited on 03/03/2025.

[5] Karmen Franinovic and Yon Visell. 2004. Recycled soundscapes. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques.* ACM, Cambridge MA USA, 317–317. https://doi.org/10.1145/1013115.1013167 Visited on 05/25/2025.

[6] Emma Frid. 2019. Accessible Digital Musical Instruments – A Review of Musical Interfaces in Inclusive Music Practice. *Multimodal Technologies and Interaction* 3, 3 (2019), 57. https://doi.org/10.3390/mti3030057

[7] Emma Frid, Alon Ilsar, Hedvig Lindetorp, Ingrid Kjellmo, Ludvig Elblaus, and Roberto Bresin. 2019. Sound Forest: Evaluation of an Accessible Multisensory Music Installation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* Glasgow, Scotland, UK, 1–12. https://doi.org/10.1145/3290605.3300907

[8] Balazs Andras Ivanyi et al. 2023. DuoRhythmo: Design and Remote User Experience Evaluation (UXE) of a Collaborative Accessible Digital Musical Interface (CADMI) for People with ALS (PALS). In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* Hamburg, Germany, 1–13. https://doi.org/10.1145/3544548.3581285

[9] Dr. Garth Paine. 2010. Towards a Taxonomy of Realtime Interfaces for Electronic Music Performance.

[10] Roel Popping. 2015. Analyzing Open-ended Questions by Means of Text Analysis Procedures. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 128, 1 (2015), 23–39. https://doi.org/10.1177/0759106315597389

[11] Jens Vetter, Martin Kaltenbrunner, and Erich Schmid. 2024. Sonic Interactions – Towards Accessible Digital Music-Making. In *Computers Helping People with Special Needs (Lecture Notes in Computer Science, Vol. 14750),* Klaus Miesenberger, Petr Peňáz, and Makoto Kobayashi (Eds.). Springer Nature Switzerland, Cham, 489–496. https://doi.org/10.1007/978-3-031-62846-7_58

[12] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. https://doi.org/10.48550/arXiv.2006.10214 arXiv:2006.10214 [cs].